



HAL
open science

Caractérisation intégrée de la matière organique dissoute : recherche d'empreintes physico-chimiques pour tracer les sources de pollutions anthropiques

Mohamed-Amine Boukra

► To cite this version:

Mohamed-Amine Boukra. Caractérisation intégrée de la matière organique dissoute : recherche d'empreintes physico-chimiques pour tracer les sources de pollutions anthropiques. Chimie. Université Claude Bernard - Lyon I, 2023. Français. NNT : 2023LYO10094 . tel-04584694

HAL Id: tel-04584694

<https://theses.hal.science/tel-04584694v1>

Submitted on 23 May 2024

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 de DOCTORAT DE L'UNIVERSITE CLAUDE BERNARD LYON 1

**Ecole Doctorale N°206
Chimie, procédés, environnement**

Discipline : Chimie de l'Environnement

Soutenue publiquement le 16/06/2023, par :

Mohamed-Amine BOUKRA

Caractérisation intégrée de la matière organique dissoute : recherche d'empreintes physico-chimiques pour tracer les sources de pollutions anthropiques

Devant le jury composé de :

JAFFREZIC	Anne	Professeure, Institut Agro Rennes-Angers	Rapporteuse
VARRAULT	Gilles	Professeur, Université de Paris-Est Créteil	Rapporteur
DEMESMAY	Claire	Professeure, Université de Lyon 1	Présidente, Examinatrice
DERRIEN	Morgane	Professeure assistante, Université d'O'Higgins	Examinatrice
GROSSI	Vincent	Directeur de recherche CNRS	Examineur
MOUNIER	Stéphane	Maître de conférences, Université de Toulon	Examineur
MIEGE	Cécile	Directrice de recherche INRAE	Directrice de thèse
MASSON	Matthieu	Ingénieur de recherche, INRAE	Co-directeur de thèse

Travail réalisé au sein du laboratoire de chimie des milieux aquatiques (LAMA) de
l'UR RiverLy, INRAE Lyon-Villeurbanne, 5 rue de la Doua, 69265 Villeurbanne

Remerciements

Sans la contribution précieuse de nombreuses personnes, qu'elles soient proches ou éloignées, cette thèse n'aurait pas atteint les mêmes résultats. Leur soutien inestimable a été essentiel pour la réalisation de ce travail.

Tout d'abord, je tiens à exprimer ma profonde gratitude envers mes encadrants de thèse, Cécile Miège et Matthieu Masson, pour m'avoir fait confiance dans l'élaboration de ce travail. Vos précieux conseils m'ont grandement aidé tout au long de cette aventure durant laquelle vous m'avez tant apporté sur le plan professionnel mais aussi personnel. Vous avez su créer un cadre de travail très agréable grâce à votre bonne humeur quotidienne, et c'était un réel plaisir pour moi de travailler avec vous.

Je remercie également les membres de mon comité de suivi de thèse, Edith Parlanti, Laurent Jeanneau, Laurent Mazéas et Pascal Molle, pour leur expertise et leurs remarques constructives qui ont fortement contribué à l'amélioration de mon travail. J'aimerais également remercier Anne Jaffrezic et Gilles Varault d'avoir accepté de rapporter ce travail de thèse, Morgane Derrien, Claire Demesmay, Vincent Grossi et Stéphane Mounier d'avoir été examinateurs. Merci pour vos remarques et conseils pour améliorer mon travail.

La thèse est avant tout un travail d'équipe et celui-ci n'aurait pas pu être mené à bien sans la précieuse contribution de l'ensemble des collègues de l'équipe LAMA. Je remercie tout particulièrement Marina Coquery, responsable de l'équipe LAMA, pour son accueil ainsi que ses nombreux conseils tout au long de ces trois années de thèse. Je remercie également l'ensemble des collaborateurs des trois laboratoires de l'équipe LAMA, à savoir le laboratoire des micropolluants organiques, le laboratoire des métaux traces et le laboratoire des éléments majeurs. Je souhaite ainsi vous exprimer ma sincère gratitude pour votre contribution à ce travail. Je vous remercie également pour les moments de rires, de discussions "philosophiques" inspirantes ou simplement de mots fléchés au coin café.

Mes remerciements vont également à tous les collègues de l'unité de recherche RiverLy et Reversaal ainsi qu'à Edith Parlanti et Mahaut Sourzac du LPTC de l'UMR EPOC de l'université de Bordeaux. Votre contribution a été essentielle pour la collecte de données et la réalisation de cette thèse.

Je remercie également Kevin Rocco, Claire Damour, Céline Guillemain, Céline Begorre, Pierre Boutet, Hassan Ba-Haddou, Alexandra Gruat, Tom Ducrocq et Anice Yari pour tous les moments passés ensemble en randonnée, aux restaurants, autour de jeux de société, ou encore en séances d'initiation au crossfit!

Un remerciement particulier à toi, ma soul mate Emma, pour ton amour et ton soutien inébranlable. Tu as toujours cru en moi et tu as toujours été là pour m'encourager et me motiver pour continuer à avancer. Aujourd'hui nous sommes tous les deux docteurs (à 10 jours d'intervalle!) et je sais que quel que soit l'avenir qui se dessine, je serai chanceux de t'avoir à mes côtés.

Enfin, à vous mes très chers parents Hamed et Djamila, ma grand-mère M'a Fatiha, ma sœur Yasmine et mes frères Yacine, Adim et Rayen, je voudrais vous dire merci pour votre amour, votre soutien et vos encouragements constants tout au long de ma vie et de mes études. Je vous aime.

À toi, papa H'bib, mon cher grand-père qui étais un amoureux de la langue française et un homme d'une culture extraordinaire, j'aurais aimé que tu puisses relire ma thèse et que l'on puisse échanger sur mon sujet. Ta sagesse bienveillante et ton soutien demeureront à jamais gravés dans mon cœur.

Amine

Avant-propos

Les travaux de cette thèse ont été réalisés à l'aide du soutien financier de l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), de l'Agence Nationale de Recherche (ANR) dans le cadre du projet CHYPSTER (ANR-21-CE34-0013-01), de la Zone Atelier Bassin du Rhône (ZABR) et de l'Agence de l'Eau Rhône Méditerranée Corse (AERMC) dans le cadre du projet IDESOC. La thèse intitulée « Caractérisation intégrée de la matière organique dissoute : recherche d'empreintes physico-chimiques pour tracer les sources de pollution anthropiques » a été effectuée au sein du Laboratoire de chimie des Milieux Aquatiques (LAMA) de l'unité de recherche UR RiverLy. La thèse a été codirigée par :

- Cécile Miège, directrice de recherche au laboratoire LAMA, de l'UR RiverLy du centre INRAE Lyon-Villeurbanne.
- Matthieu Masson, ingénieur de recherche au laboratoire LAMA, de l'UR RiverLy du centre INRAE Lyon-Villeurbanne.

L'ensemble des analyses a été réalisé au laboratoire LAMA à l'exception de la technique d'excitation-émission de fluorescence qui a été réalisée au sein de l'équipe Physico et Toxico Chimie de l'environnement (LPTC) de l'UMR EPOC 5805 CNRS de l'université de Bordeaux, sous la supervision de Madame Edith Parlanti, chargée de recherche au LPTC.



Résumé

Les activités anthropiques à l'échelle d'un bassin-versant peuvent être très nombreuses (e.g. pratiques agricoles, industries, transport et tourisme). Ces activités engendrent des pressions qui se manifestent par l'apport de polluants vers les cours d'eau (e.g. éléments majeurs, micropolluants organiques, métaux traces). L'identification de ces sources de polluants dans les cours d'eau, et plus particulièrement les sources diffuses, reste un réel défi à ce jour. La matière organique dans sa fraction dissoute (MOD) constitue un mélange complexe de molécules dont la composition et les propriétés physico-chimiques dépendent de son origine qu'elle soit naturelle ou anthropique, terrigène ou autochtone, diffuse ou ponctuelle. Du fait de son ubiquité, la MOD présente dans les cours d'eau pourrait ainsi être utilisée comme traceur des activités anthropiques et donc des sources de polluants à l'échelle d'un bassin versant.

Dans ce contexte, l'objectif de cette thèse est de mettre en évidence des marqueurs basés sur des propriétés physico-chimiques de la MOD (e.g. distribution en poids moléculaire, aromaticité, famille de composés...) qui soient caractéristiques de sources naturelles (e.g. apports terrigènes, production autochtone) et anthropiques (e.g. rejets de stations de traitement des eaux usées STEU, ruissellement urbain, activités d'agriculture et d'élevage) à l'échelle de bassins versants. Pour cela, une base de données originale a été construite avec plus de 130 échantillons d'eau représentatifs de ces différentes sources et analysés avec un large panel de techniques analytiques (dosage du carbone organique dissous, analyse par spectroscopie UV-Visible et excitation-émission de fluorescence, analyse par chromatographie d'exclusion stérique couplée à la détection UV et fluorescence - HPSEC/UV-fluorescence -, et analyse moléculaire par chromatographie liquide couplée à la spectrométrie de masse haute résolution - LC-HRMS -).

L'exploitation de la base de données a tout d'abord permis d'améliorer la caractérisation de la MOD avec l'identification de nouveaux indicateurs optiques (HPSEC/UV-fluorescence) et moléculaires (LC-HRMS). Un protocole de traitement des données complexes issues de l'analyse LC-HRMS non-ciblée a été développé et validé à l'aide de contrôle qualité afin d'extraire les composés moléculaires les plus pertinents pour différencier les sources de MOD. Ces travaux ont également servi à valider une méthodologie d'échantillonnage spécifique aux sources diffuses terrigènes en lien avec l'occupation des sols. Finalement, l'approche intégrée multi-analytique et multi-sources développée dans le cadre de la thèse a contribué à la construction d'empreintes spécifiques des sources de MOD sélectionnées a priori (naturelles, anthropiques, diffuses, ponctuelles). Les résultats de ces travaux permettent clairement d'envisager la MOD comme traceur de l'origine des masses d'eaux et des pollutions associées dans les cours d'eau.

Abstract

Anthropogenic activities at the scale of a watershed can be very numerous (e.g. agricultural practices, industries, transport, and tourism). These activities generate pressures that result in the input of pollutants to the watercourse (e.g. major elements, organic micropollutants, trace metals). The identification of these sources of pollutants in watercourses remains a challenge to date. Organic matter in its dissolved fraction (DOM) constitutes a complex mixture of molecules whose composition and physicochemical properties depend on its origin, whether natural or anthropogenic, terrigenous or autochthonous, diffuse or point source. Because of its ubiquity, the DOM present in the rivers could be used as a tracer of the anthropic activities and thus of the sources of pollutants at the scale of a catchment.

In this context, the objective of this thesis is to identify markers based on physico-chemical properties of DOM (that are characteristic of natural (e.g. terrigenous inputs, autochthonous production) and anthropogenic (e.g. wastewater treatment plant discharges, urban runoff, agriculture and livestock activities) sources at the watershed scale. For this purpose, an original database was built with more than 130 water samples representative of these different sources and analyzed with a wide range of analytical techniques (dissolved organic carbon determination, analysis by UV-Visible spectroscopy and fluorescence excitation-emission, analysis by steric exclusion chromatography coupled with UV and fluorescence detection - HPSEC/UV-fluorescence - , and molecular analysis by liquid chromatography coupled with high resolution mass spectrometry - LC-HRMS -).

The exploitation of the database allowed to improve the characterization of DOM with the identification of new optical (HPSEC/UV-fluorescence) and molecular (LC-HRMS) indicators. A protocol for processing complex data from non-targeted LC-HRMS analysis was developed and validated using quality control in order to extract the most relevant molecular compounds to differentiate DOM sources. This work also validated a sampling methodology specific to land-based diffuse sources in relation to land use to build their footprints at the watershed scale. Finally, the integrated multi-analytical and multi-source approach developed in the framework of the thesis allowed the construction of specific fingerprints of the selected sources of DOM (natural, anthropogenic, diffuse, point). The results of this work clearly show that DOM can be considered as a tracer of the origin of the water masses and the associated pollutions in the rivers.

Table des matières

Valorisations scientifiques	15
Liste des figures	17
Liste des tables	27
Liste des abréviations	29
Introduction générale	33
Chapitre 1. État de l'art	41
1.1 Généralités sur la matière organique dissoute (MOD)	41
1.2 Les sources de matière organique dissoute	45
1.2.1 Sources autochtones	45
1.2.2 Sources allochtones.....	47
1.2.3 Caractérisation de la MOD dans les différentes sources	55
1.3 Les techniques de caractérisation de la matière organique dissoute.....	57
1.3.1 Dosage du carbone organique dissous (DOC).....	57
1.3.2 Analyse des isotopes stables.....	57
1.3.3 Techniques de caractérisation optique.....	58
1.3.4 Résonance magnétique nucléaire (NMR)	67
1.3.5 Chromatographie d'exclusion stérique (SEC).....	68
1.3.6 Techniques de spectrométrie de masse	70
1.3.7 Caractérisation de la MOD à travers une approche intégrée	75
1.4 Références	78
Chapitre 2. Objectifs et questionnement scientifique.....	93
Chapitre 3. Matériels et méthodes.....	99
3.1 Stratégie d'échantillonnage.....	99
3.1.1 Présentation des sites d'étude.....	99
3.1.2 Choix des sous-bassins homogènes et des sites de prélèvement des eaux associés.....	103
3.1.3 Protocole de collecte des échantillons d'eau	107
3.1.4 Méthode d'échantillonnage des sols et d'extraction de la MOD	108
3.1.5 Prélèvement des effluents de stations de traitements d'eaux usées (STEU) ..	112
3.1.6 Prélèvement des eaux de ruissellement urbain.....	114

3.2	Stratégie analytique.....	115
3.2.1	Analyse du carbone organique dissous (DOC)	115
3.2.2	Spectroscopie UV-Visible	115
3.2.3	Spectroscopie d'excitation-émission de fluorescence.....	117
3.2.4	Chromatographie d'exclusion stérique HPSEC/UV-fluorescence	120
3.2.5	Chromatographie liquide couplée à la spectrométrie de masse haute résolution	121
3.3	Analyses statistiques des résultats.....	124
3.4	Références	126
Chapitre 4. Mise en place de nouveaux indicateurs HPSEC/UV-fluorescence et LC-HRMS de caractérisation de la MOD.....		131
4.1	Proposition de nouveaux indicateurs HPSEC/UV-fluorescence	132
4.1.1	Indicateurs classiques de distribution de poids moléculaire : Mw, Mw et ρ ..	132
4.1.2	Nouveaux indicateurs basés sur la décomposition des chromatogrammes HPSEC-UV et fluorescence : LMW, MMW et HMW	134
4.2	Proposition de nouveaux indicateurs LC-HRMS.....	145
4.2.1	Acquisition des données LC-HRMS en approche non-ciblée	145
4.2.2	Retraitement des données HRMS pour la génération de la liste finale de <i>features</i> en vue de l'étude des sources de MOD (publication n°2)	153
4.2.3	Exploitation des données LC-HRMS dans un contexte d'étude des sources de MOD dans les cours d'eau récepteurs	192
4.2.4	Conclusion générale du chapitre 4.....	199
Chapitre 5. Etude des sources diffuses de matière organique dissoute		203
5.1	Échantillonnage des sources diffuses terrigènes dans les cours d'eau : influence de l'utilisation des sols et des conditions hydrologiques sur les caractéristiques de la matière organique dissoute (publication n°1, publiée).....	203
5.1.1	Résumé de la publication scientifique	203
5.2	Etude de l'empreinte globale des sources diffuses terrigènes à l'échelle de plusieurs bassins versants.....	252
5.2.1	Teneur en DOC et description globale des indicateurs optiques classiques dans différents types de sources diffuses de MOD	253
5.2.2	Caractérisation intégrée de la MOD issue des sources diffuses : identification d'un effet bassin versant.....	259
5.2.3	Sur quel type de source diffuse l'effet bassin versant est-il visible ?	261
5.2.4	Est-il possible de construire des empreintes d'occupations de sols indépendamment du bassin versant ?.....	268

5.2.5	Conclusion	270
5.2.6	Annexes	271
Chapitre 6. Utilisation de la MOD comme traceur de sources de pollution à l'échelle d'un bassin versant péri-urbain : approche optique et moléculaire non-ciblée (Publication n°3, en préparation).....		281
6.1	Résumé de la publication	281
Chapitre 7. Conclusion et perspectives.....		327
7.1	Nouveaux indicateurs de caractérisation de la MOD issue de différentes sources	328
7.2	Validation d'une méthodologie de prélèvement des sources diffuses.....	332
7.3	Démonstration de la pertinence de la MOD comme traceur de sources de pollution dans les cours d'eau à l'échelle d'un bassin versant : application de l'approche intégrée.	334

Valorisations scientifiques

Articles rédigés dans le cadre de la thèse

Article #1 (publié)

Boukra A, Masson M, Brosse C, Sourzac M, Parlanti E, Miège C. Sampling terrigenous diffuse sources in watercourse: Influence of land use and hydrological conditions on dissolved organic matter characteristics. 2023. STOTEN. <https://doi.org/10.1016/j.scitotenv.2023.162104>

Article #2

Boukra A, Masson M, Rocco K, Margoum C, Miège C. LC-HRMS non-targeted data to characterize dissolved organic matter DOM in watercourses impacted by land uses. En cours de soumission. Journal of environmental sciences JES.

Article #3

Boukra A, Masson M, Miège C, Parlanti E, Coquery M. Using dissolved organic matter as a tracer of natural and anthropogenic inputs at the scale of a peri-urban watershed: optical and non-targeted molecular approach. En préparation.

Autres articles publiés en collaboration

Djaoudi K, Tesan J, **Boukra A**, Guesney L, Portas A, Barry-Martinet R, Angeletti B, Mounier S, Lenoble V, Briand JF. Seawater copper content controls biofilm bioaccumulation and microbial community on microplastics. 2022. STOTEN. <https://doi.org/10.1016/j.scitotenv.2021.152278>.

Diamond J, Bernal S, **Boukra A**, Cohen M, Lewis D, Masson M, Moatar F, Pinay G. Stream network variation in dissolved oxygen: Metabolism proxies and biogeochemical controls. 2021. Ecological indicators. <https://doi.org/10.1016/j.ecolind.2021.108233>.

Présentations orales et posters

Boukra A, Masson M, Brosse C, Sourzac M, E Parlanti, Miège. Optical and molecular non-targeted approach to characterize and trace natural and anthropogenic sources of dissolved organic matter. Congrès FROG V : les molécules de nos paysages. Oral. 4-6 juillet 2022. Rennes, France.

Boukra A, Masson M, Brosse C, Sourzac M, E Parlanti, Miège. Multi-analytical approach to investigate sources of dissolved organic matter DOM in a peri-urban watershed. EGU General Assembly 2022. Oral. 23-27 mai. Vienne, Autriche.

Boukra A, Masson M, Brosse C, Sourzac M, E Parlanti, Miège. Approche multi-analytique optique et moléculaire pour construire des empreintes de MOD typiques de sources naturelles et anthropiques. Journées thématiques du Groupe Français – IHSS (GF-IHSS). Oral. 18-19 mai 2022. Nancy, France.

Masson M, **Boukra A**, Lebre S, Rocco K, Merel S, Miège C. Stratégie pour la construction d'une base de données d'empreintes moléculaires organiques dissoutes par chromatographie liquide couplée à un spectromètre de masse haute résolution (LC-HRMS). Réunion des sciences de la terre RST. Poster. 1-5 novembre 2021. Lyon, France.

Masson M, Parlanti E, **Boukra A**, Brosse C, Sourzac M, Arhror M, Richard L, Miège C. Caractéristiques optiques de la matière organique dissoute dans des eaux usées domestiques. Réunion des sciences de la terre RST. Oral. 1-5 novembre 2021. Lyon, France.

Boukra A, Masson M, Arhror M, Richard L, Brosse C, Sourzac M, E Parlanti, Miège. Etude des sources de matière organique dissoute à l'échelle d'un bassin versant péri-urbain. Réunion des sciences de la terre RST. Poster. 1-5 novembre 2021. Lyon, France

Boukra A, Masson M, Arhror M, Richard L, Brosse C, Sourzac M, E Parlanti, Miège. Towards a better identification of dissolved organic matter diffuse sources: comparative study between soil leachate and adjacent watercourse to search for specific markers. OZCAR-TERENO conference. Poster. 5-7 octobre 2021. Strasbourg, France.

Liste des figures

Figure 1-1. Continuum de la taille de la matière organique dans les eaux naturelles. Adapté de Stumm 1977	42
Figure 1-2. Composition et propriétés physicochimiques de la DOM. D’après Xiao et al., 2020	43
Figure 1-3. Interactions de la matière organique dissoute avec les différents compartiments écologiques.....	44
Figure 1-4. Illustration des principales sources de MOD à l'échelle d'un bassin versant.....	45
Figure 1-5. Interactions de la matière organique dissoute avec les microorganismes et les producteurs primaires à l'échelle du milieu aquatique. D’après Zhang et al., 2018	47
Figure 1-6. Schéma des principaux phénomènes intervenant dans le transfert des sources diffuses terrigènes. Image : Pr. André Musy. Cours d’hydrographie générale.....	49
Figure 1-7. Caractéristiques de la MOD en fonction des écoulements dans les rivières. Adapté de Lupon et al., 2020.....	50
Figure 1-8. : Exemples de procédés de traitement des eaux usées (adapté de Simate, 2015)	52
Figure 1-9. Principales composantes de la MOD en sortie de STEU.....	54
Figure 1-10. Principe de mesure d'un spectre d'absorption (Luciani, 2008).....	59
Figure 1-11. Diagramme de Jablonski, adapté à partir de Savoca et al., 2016.....	62
Figure 1-12. Exemple d'une matrice d'excitation émission de fluorescence avec la localisation des principaux fluorophores de la MOD selon Coble et al., 1996	63
Figure 1-13. Techniques RMN utilisées pour caractériser la MOD.....	67
Figure 1-14. Représentation schématique de la séparation par chromatographie d'exclusion stérique. Les molécules de petite taille diffusent à travers les particules poreuses alors que les molécules de taille plus importante éluent sans rétention. Adapté de (Mori and Barth, 1999)	69
Figure 1-15. Diagramme de van Krevelen réalisés à partir des compositions élémentaires de deux échantillons caractéristiques d’une rivière et d’un lac. D’après Minor et al., 2012	72
Figure 1-16. Schéma de l'analyseur de masse Orbitrap montrant une trajectoire en spirale stable d'un ion. D’après Hecht et al., 2019	73
Figure 1-17. Schéma d'un instrument QTOF (waters)	74

Figure 1-18. Principales techniques utilisées pour caractériser les différentes fractions de la MOD	76
Figure 3-1. Carte du bassin versant de l’Ardières-Morcille indiquant la localisation des sites d’échantillonnage et les limites de l’occupation des sols (<i>Corine Land Cover CLC 2018</i>)......	100
Figure 3-2. Carte du bassin versant de l’Yzeron indiquant la localisation des sites d’échantillonnage et les limites de l’occupation des sols (<i>Corine Land Cover CLC 2018</i>). Les triangles blancs représentent les points ruissellement urbains prélevés deux fois pour la variabilité temporelle (voir partie 1.5).....	101
Figure 3-3. Carte du bassin versant du Gier-Grand Malval indiquant la localisation des sites d’échantillonnage et les limites de l’occupation des sols (<i>Corine Land Cover CLC 2018</i>)......	102
Figure 3-4. Principales étapes traitement sur le logiciel QGIS pour la sélection d’un sous- bassin versant et le calcul du pourcentage de contribution des usages et occupations des sols à l’exutoire.	103
Figure 3-5. Principales étapes traitement sur le logiciel QGIS pour la sélection d’une sous bassin versant et le calcul du pourcentage de contribution des usages et occupations de sols à l’exutoire.....	103
Figure 3-6. Exemple de l’identification d’un sous-bassin versant agricole (1.8 km ²) sur le bassin de l’Ardières avec les pourcentages de contribution des usages de sols associés. La station de prélèvement de ce sous-bassin homogène est définie comme le point aval du cours d’eau drainant ce sous-bassin.	104
Figure 3-7. Exemple de l’identification dun sous bassin versant agricole sur le bassin de l’Ardières avec les pourcentages de contribution des usages de sols associés. La station de prélèvement de ce sous bassin homogène est défini comme le point aval du cours d’eau drainant ce sous-bassin.	104
Figure 3-8. Conditionnement des échantillons d’eau avant analyse.....	108
Figure 3-9. Illustration d’un échantillonnage de sol dans un sous-bassin versant à proximité en amont d’un point d’échantillonnage dans le cours	109
Figure 3-10. Schéma descriptif des étapes d’extraction et de dégradation de la MOD extraite des sols.	112
Figure 3-11. Exemple de spectres UV-Visible d’eaux de STEU brute et traitée ainsi que d’une eau naturelle prélevée en aval d’un sous-bassin versant forestier.	116

Figure 3-12. Exemples de matrices d'excitation-émission de fluorescence d'eaux de STEU brute et traitée ainsi que d'une eau naturelle prélevée en aval d'un sous-bassin versant forestier.	119
Figure 3-13. Courbe d'étalonnage (logarithme du poids moléculaire vs volume d'élution) des standards PSS utilisés pour évaluer le poids moléculaire (MW) moyen des échantillons de MOD	120
Figure 3-14. Illustration de la répartition des features (signaux caractérisés par un rapport masse sur charge m/z en Da et un temps de rétention RT en min) générés dans une séquence d'analyse en LC-HRMS non-ciblée.....	124
Figure 4-1. Exemple de 2 chromatogrammes HPSEC-UV à 254 nm : haut : une eau de STEU brute et traitée, bas : une d'une eau naturelle prélevée en aval d'un bassin versant forestier.	134
Figure 4-2. Illustration d'une décomposition gaussienne d'un chromatogramme HPSEC....	135
Figure 4-3. Test de comparaison entre injection de PSS séparément et en mélange avec application du modèle de décomposition gaussienne	138
Figure 4-4. Analyse en composantes principales des échantillons de DOM en fonction A. des nouveaux indicateurs (LMW, MMW et HMW) et B. des indicateurs classiques de distribution de poids moléculaire (M_n , M_w et ρ).....	141
Figure 4-5. Matrice de corrélation des indicateurs UV-Visible, EEM de fluorescence et HPSEC/UV-fluorescence. Les cases blanches indiquent une corrélation Pearson non-significative ($p > 0.05$, $n = 136$ échantillons de DOM).....	142
Figure 4-6. Exemple de chromatogrammes TIC (<i>Total Ion Chromatograms</i>) générés par LC-HRMS d'eaux de STEU brute et traitée ainsi que d'une eau naturelle prélevée en aval d'un bassin versant forestier. Le TIC représente la somme des signaux générés par tous les ions qui sont élués de la colonne chromatographique au cours de l'analyse.....	145
Figure 4-7. Histogrammes empilés pour 3 séquences d'analyse avec le nombre de <i>features</i> détectés par XCMS (WorkflowForMetabolomics) et MZmine et les <i>features</i> en commun aux deux logiciels.	147
Figure 4-8. Etude de la variation de RT et m/z des traceurs d'injection diuron d6 et tébuconazole D6 dans 5 séquences d'analyses de janvier à avril 2021.....	148
Figure 4-9. Comparaison des traceurs d'injections retrouvés après traitement par MZmine et XCMS.	149

Figure 4-10. Analyses en composantes principales des injections d'échantillons de DOM et contrôles qualités (blancs d'injection : n=17, QCpool : n=25 et QC mix : n= 3) en fonction des intensités des <i>features</i> générés par MZmine et XCMS.	151
Figure 4-11. Résumé graphique de la publication scientifique: <i>Processing LC-HRMS non-targeted data to characterize dissolved organic matter in watercourses impacted by land uses.</i> En cours de soumission dans : <i>Journal of Environmental Sciences</i>	154
Figure 4-12. Map of Ardieres-Morcille catchment with lands-use boundaries (Corine Land Cover CLC 2018) showing the location of the water-sampling sites. FW: forest water, AW: agricultural water, WR: water reservoir. The percentage of land use contribution for each point was calculated based on the upstream land use area.	161
Figure 4-13. LC-HRMS data processing steps.	165
Figure 4-14. Violin plots of A. Relative standard deviation (RSD) of adduct intensities in QCP before (left) and after (right) intensity sum. B. RSD of isotope intensities in QCP before (left) and after (right) isotopes intensity sum.....	167
Figure 4-15. First factorial plane of the principal component analysis (PCA) performed on the compounds with intensities lower than the threshold based on injection blanks (n=5532).170	
Figure 4-16. Data processing steps of removal of compounds with highly variable intensities in QCP (n=24) and in each triplicate injection (n=28 triplicates). The red cross means that the compounds were deleted at that stage.	171
Figure 4-17. First factorial plan of the principal component analysis (PCA) of quality controls and DOM samples. A. Data processing with a maximum threshold at 50% for the relative standard deviation (RSD) on compound intensities in QCP and randomized triplicates samples (number of molecular compounds, n =1478). B. Data processing with a maximum threshold at 30% for the RSD on compound intensities in QCP and randomized triplicates samples (n = 457).	173
Figure 4-18. Ascendant hierarchical classification (HAC) of quality controls and DOM samples. A. Data processing with a maximum threshold at 50% for the relative standard deviation (RSD) on compound intensities in QCP and randomized triplicates (number of molecular compounds; n =1478). B. Data processing with a maximum threshold at 30% for the RSD on compound intensities in QCP and randomized triplicates (n = 457).	175
Figure 4-19. First factorial plan of principal component analysis (PCA) performed on compounds intensities (n=463 molecular compounds). A. The data correspond to DOM	

samples collected during high and low water periods. B. The data correspond to DOM samples collected only during high water periods..... 177

Figure 4-20. Exemples de la répartition en RT et m/z de composés moléculaires dans des échantillons de DOM issus de sources différentes 192

Figure 4-21. A. Illustration des indicateurs globaux LC-HRMS basés sur la répartition des composés moléculaires en fonction des temps de rétention, des masses et des intensités des composés. B. Exemple d'application sur des échantillons de DOM issues de sources différentes (e.g. eau naturelle en sortie de bassin forestier, ruissellement urbain, entrées et sorties de STEU) 194

Figure 4-22. Matrice de corrélation entre les indicateurs globaux LC-HRMS (en rouge), les indicateurs UV-Visible (en bleu) et les indicateurs EEM de fluorescence (en vert) 196

Figure 4-23. Diagramme illustrant le nombre d'intersections et de non-intersections entre les composés moléculaires détectés dans différents types de sources de DOM. Les points reliés par un trait représentent les intersections, i.e. les composés détectés dans deux ou plusieurs sources. Les points non reliés indiquent le nombre de composés spécifiques à une seule source donnée..... 197

Figure 5-1. Résumé graphique de la Partie 5.1 sur la méthodologie de prélèvement des sources diffuses (Boukra et al., 2023). 205

Figure 5-2. Map of Ardieres-Morcille catchment showing the location of the water-sampling sites (FP: forest sampling points, VP: vineyard sampling points) and the land-use boundaries (Corine Land Cover CLC 2018). The soil samples were collected upstream the water sampling points..... 210

Figure 5-3. A: Dissolved organic carbon (DOC) concentrations measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods. B: Percentage of DOC for soil extracts during the 28 days incubation experiment relative to the DOC concentration measured before incubation..... 219

Figure 5-4. UV-Visible indicators measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods. A: SUVA₂₅₄, B: E2.E3 (abs250/abs365),

C: Slope S1 (275_295 nm), D: Slope ratio SR (S275_295nm/S350_400nm), E: E2.E4 (abs254/abs436) and F: E3.E4 (abs300/abs400).....	221
Figure 5-5. EEM fluorescence indicators measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods. A: humification index HIX, B: biological fluorescence index BIX, C: fluorescence index FI and D: fluorescence index YFI.	223
Figure 5-6. PARAFAC components C1 (<240(320)-420 nm), C2 (<240(390)-510 nm), C3 (365-443 nm) and C4 (280-330 nm) measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after at 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods.....	224
Figure 5-7. Low Molecular Weight (LMW; blue bars), Medium Molecular Weight (MMW; orange bars) and High Molecular Weight (HMW; purple bars) percentage of contribution to: A: Total Sec Absorbance TSA at PDA 254 nm, B: Total Sec Fluorescence TSF at Excitation Emission wavelengths 345-443 nm, C: TSF at fluorescence Excitation Emission wavelengths 280-330 nm, D: TSF at fluorescence Excitation Emission wavelengths 278-354 nm.. Samples plotted are soil extracts before incubation (B.I) and then at 3, 7, 14 and 28 days of incubation as well as for water samples collected from streams adjacent to the forest and vineyard soils during high and low water periods.	228
Figure 5-8. First factorial plane of the principal component analysis (PCA) performed on data from forest and vineyard soil extracts before incubation (B.I), after 28 days of incubation and their adjacent water stream samples collected during high and low water periods. The variables considered in the PCA are the UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence indicators.	231
Figure 5-9. A. Linear discriminant analysis (LDA1) calculated with 26 indicators from UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence to discriminate forest and vineyard DOM taken during high water from low water DOM. B. Linear discriminant analysis (LDA2) calculated with the same indicators to discriminate forest and vineyard DOM taken during low water periods from High water DOM.	233

Figure 5-10. Boxplots des concentrations de carbone organique dissous (DOC) en mg.L ⁻¹ dans les échantillons représentatifs des sources diffuses.....	254
Figure 5-11. Valeurs médianes des indicateurs UV-Visible dans les échantillons de sources A. agricoles en hautes eaux (HW) et basses eaux (LW), B. forestier/prairies en HW et LW et C. retenues d'eau en HW et LW. E2.E3 : inversement corrélé à la taille des molécules, E2.E4 : inversement corrélé à la production autochtone, E3.E4 : inversement corrélé à l'humification, SUVA : aromaticité de la DOM (L.mg ⁻¹ cm ⁻¹), SR : inversement corrélé au poids moléculaire.	256
Figure 5-12. Composantes PARAFAC validées dans les échantillons de sources diffuses. C1 (Ex240-350 Em440), C2 (Ex260-Em505), C3 (Ex280-Em330), C4 (Ex325-Em395 nm), C5_EN (Ex<240-Em390)	257
Figure 5-13. Valeurs médianes des indicateurs de fluorescence dans les échantillons de sources A. agricoles en hautes eaux (HW) et basses eaux (LW), B. forestiers/prairies, fractions HW et LW et C. retenues d'eau, fractions HW et LW. BIX : corrélé à l'activité microbienne, FI et YFI : indicateurs d'apports terrigènes, HIX : indicateur d'humification, C1 : composé terrigène dégradé (R.U), C2 : composé type humique, C3 : composé type protéique, C4_EN : composé type humique issue de l'activité microbienne, C5_EN : composé terrigène aromatique.....	258
Figure 5-14. Analyse en composantes principales des échantillons de DOM représentatifs des sources diffuses prélevés dans les bassins versants de l'Ardières-Morcille, de l'Yzeron et du Gier-Grand Malval.	260
Figure 5-15. Analyse en composantes principales des échantillons de DOM représentatifs des sources diffuses agricoles, prélevés dans les bassins versants de l'Ardières-Morcille, de l'Yzeron et du Gier-Grand Malval en périodes de hautes eaux (HW) et de basses eaux (LW)	262
Figure 5-16. Pourcentage de contribution des petites molécules (< 1 k Da) à la fluorescence totale aux longueurs d'onde d'excitation-émission 275-354 nm (à gauche) et 280-330 nm (à droite), dans les sources des 3 bassins versants (jaune : Yzeron, bleu : Gier, rouge : Ardières-Morcille).	263
Figure 5-17. Analyse en composantes principales des échantillons de DOM représentatifs des sources diffuses forestières/pâturages, prélevés dans les bassins versants de l'Ardières-	

Morcille, de l'Yzeron et du Gier-Grand Malval en périodes de hautes eaux (HW) et de basses eaux (LW).....	265
Figure 5-18. Analyse en composantes principales des échantillons de DOM représentatifs des retenues d'eau, prélevés dans les bassins versants de l'Ardières-Morcille, de l'Yzeron et du Gier-Grand Malval en périodes de hautes eaux (HW) et de basses eaux (LW)	266
Figure 5-19. Analyse linéaire discriminante des échantillons de DOM représentatifs des sources diffuses terrigènes prélevés dans les BV de l'Ardières-Morcille, de l'Yzeron et du Gier.	269
Figure 6-1. Résumé graphique de la publication scientifique: <i>Using dissolved organic matter as a tracer of land-use-inputs at the scale of a peri-urban watershed: optical and non-targeted molecular combined approaches.</i>	282
Figure 6-2. Sampling strategy of DOM sources with A/ the location of diffuse source related to land-use and urban runoff sources (Corine Land Cover CLC 2018) and B/ description of wastewater sewage (inflows and outflows) sources	288
Figure 6-3. Classification of the collected DOM sources according to the concept of source discrimination level (SDL). WWTP: wastewater treatment plants, P.E: population equivalent.	290
Figure 6-4. Splitting the distribution of molecular compounds according to their retention time and mass. e.g. RT0.6_MW0.500 represents the sum of the intensities of the compounds which elute between 0 and 6 min and whose neutral mass is lower than 500 Da. The sum of the intensities is then normalized by the DOC concentration of the sample.	296
Figure 6-5. First factorial plane of the principal component analysis performed on DOM indicators (UV-Visible, EEM fluorescence, HPSEC/UV-fluorescence and LC-HRMS) measured on wastewaters (WWTP's inflows and outflows), urban runoffs and waters influenced by various land uses and hydrological conditions.	304
Figure 6-6. First factorial plane of the principal component analysis performed on DOM indicators measured on diffuse sources influenced by various land uses (forests/pastures, agriculture and water reservoirs) and hydrological conditions (high and low water flows ..	307
Figure 6-7. First factorial plane of the principal component analysis performed on DOM indicators (UV-Visible, EEM fluorescence, HPSEC/UV-fluorescence and LC-HRMS) measured on point high-anthropogenic sources (PRF and AS WWTP's inflows and outflows and urban runoffs).....	310

Figure 6-8. Barplot highlighting the number of intersections and non-intersections (specific or common molecular compounds) between A: DOM sources influenced by various land uses with DOM sources from urban runoff. B: DOM sources influenced by various land uses with DOM sources from planted reed filter WWTP. C: DOM sources influenced by various land uses with DOM sources from activated sludge WWTP..... 312

Figure 6-9. Barplot highlighting the number of intersections and non-intersections (specific or common molecular compounds) between DOM sources from urban runoff, DOM sources from planted reed filter WWTP and DOM sources from activated sludge WWTP 314

Liste des tables

Tableau 1-1. Principaux fluorophores de la MOD	63
Tableau 3-1. Récapitulatif des échantillons d'eau représentatifs de sources diffuses prélevés dans le cadre de la thèse (les coordonnées sont exprimées en degrés décimaux dans le système géodésique mondial WGS84).....	106
Tableau 3-2. Coordonnées géographiques détaillées des zones d'échantillonnage des sols avec les concentrations particulières de carbone organique (<i>Flash Dynamic Combustion</i>) dans chaque sol (les coordonnées sont exprimées en degrés décimaux dans le système géodésique mondial WGS84).....	110
Tableau 3-3. Descriptif des échantillons prélevés dans les stations de traitement des eaux usées.....	113
Tableau 3-4. Descriptif des échantillons représentatifs des ruissellements urbains ((les coordonnées sont exprimées en degrés décimaux dans le système géodésique mondial WGS84).....	114
Tableau 3-5. Indicateurs UV-Visible calculés à partir des spectres d'absorbance	116
Tableau 3-6. Indicateurs calculés à partir des matrices d'excitation-émission de fluorescence	118
Tableau 3-7. Traceurs deutérés d'injection (en bleu) et composés du contrôle qualité des étalons en mélange (QCmix)	122
Tableau 4-1. Temps de rétention (RT) associés aux composantes identifiées par le modèle de décomposition dans le mélange de PSS et les RT des PSS injectés séparément.	139
Tableau 6-1. Dissolved organic carbon, DOM optical and molecular indicators: median, maximum and minimum values	301

Liste des abréviations

OM/MO	Organic Matter / Matière Organique
DOM/MOD	Dissolved Organic Matter/ Matière Organique Dissoute
CDOM	Chromophoric Dissolved Organic Matter
FDOM	Fluorescent Dissolved Organic Matter
DOC	Dissolved Organic Carbon
TOC	Total Organic Carbon
UV	Ultraviolet
SR	Slope Ratio
SUVA	Specific UV Absorbance
EEM	Excitation Emission Matrix
PARAFAC	Parallel factor analysis
HIX	Humification Index
BIX	Biological Index
FI	Fluorescence Index
nm	Nanomètres
NMR	Nuclear Magnetic Resonance
FTIR	Fourier Transform Infrared
HPSEC	High Performance Size Exclusion Chromatography
TSA	Total Sec Absorbance
TSF	Total Sec Fluorescence
LMW	Low Molecular Weight
MMW	Medium Molecular Weight
HMW	High Molecular Weight
Mn	Nombre de poids moléculaire moyen
Mw	Poids moléculaire moyen
P	Polydispersité
RT, TR	Retention Time, Temps de Rétention
LC-HRMS	Liquid Chromatography-High Resolution Mass Spectrometry
QToF	Quadrupole Time of Flight
NTS	Non-Targeted Screening

FTICR-MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
EIC	Extracted Ion Chromatograms
ESI	ElectroSpray Ionisation
QC	Quality Controls
W4M	Workflow for metabolomics
Da	Daltons
CPS	Count Per Second
RSD	Relative Standard Deviation
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
HAC	Hierarchical Ascendant Classification
STEU	Station de Traitement des Eaux Usées
WWTP	Wastewater treatment plant
FPR	Filtres Plantées de Roseaux
RU	Ruissellement Urbain
HW	High Water
LW	Low water
BA	Boues Activée
WR	Water reservoir
PVDF	Polyvinylidene Fluoride
SDL	Source Discrimination Level
MNT	Modèle Numérique de Terrain

Introduction générale

Introduction générale

La matière organique est définie comme le plus grand réservoir actif de carbone sur la planète (Hedges, 1992). Son omniprésence dans les milieux terrestres et aquatiques est principalement due à des apports à la fois naturels, par dégradation des plantes et animaux par les microorganismes (Kracht, 2001), et anthropiques, à travers les différents rejets urbains, agricoles, et industriels (Guo et al., 2007). Historiquement, les premières publications sur la matière organique dissoute (MOD) datent des années 1900, quand August Pütter décrit comme matière organique « non formée » la fraction de la matière organique passant à travers un filtre. Cette définition opérationnelle est encore utilisée à ce jour et les seuils de filtrations ont été définis entre 0,2 et 0,7 μm pour différencier la MOD de la fraction particulaire de la matière organique.

Dans les écosystèmes aquatiques, la MOD correspond à un mélange complexe et hétérogène de molécules organiques, regroupant des composés comme les acides aminés, les protéines, les lipides, les composés de type-humique et les polyphénols. La complexité de la composition de la MOD vient principalement de la grande diversité de composants chimiques présentant des structures et groupements fonctionnels extrêmement variés. L'hétérogénéité de la MOD est également associée à la variabilité des composants en termes de taille. En effet, la MOD est souvent présentée comme un mélange de composés caractérisé par un continuum de taille très large, allant d'une centaine de Daltons (e.g. acides aminés, monosaccharides) à des dizaines de milliers de Daltons (e.g. protéines, composés de type-humique) (Nebbioso and Piccolo, 2013). Cette MOD, complexe et hétérogène, occupe un rôle majeur dans les cycles biogéochimiques, notamment en tant que principal vecteur de carbone terrestre vers le milieu aquatique, en tant que substrat pour la croissance des microorganismes (Bolan et al., 2011), et également en raison de son rôle dans le devenir et la biodisponibilité des polluants organiques et inorganiques (Aiken, 2014).

À l'échelle d'un bassin versant, les apports de MOD dans les rivières peuvent provenir d'un large éventail de sources naturelles et anthropiques (Nebbioso and Piccolo, 2013), qui peuvent être ponctuelles ou diffuses (Williams et al., 2010). Les sources ponctuelles sont celles qui proviennent d'un point de rejet unique, identifiable ou canalisé, comme les rejets des stations de traitement des eaux usées (STEU) ou des industries (Lapworth et al., 2012). Les

sources diffuses sont réparties sur une vaste zone et n'ont pas de source unique facilement localisable : c'est par exemple le cas du ruissellement issu des activités agricoles (Tran et al., 2015). Ces sources peuvent apporter un large éventail de polluants, tels que des pesticides, produits pharmaceutiques et métaux traces, qui peuvent altérer la santé de l'écosystème (Kayode-Afolayan et al., 2022; Leenhardt et al., 2023; Singh et al., 2023). Ainsi, l'identification de ces sources de pollution à l'échelle d'un bassin versant est essentielle pour le suivi de la qualité de l'eau et la prévention de scénarios catastrophiques pour l'environnement.

La complexité de la composition de la MOD ainsi que son rôle essentiel dans les écosystèmes et son lien étroit avec les sources de pollution ont suscité l'intérêt de la communauté scientifique. Ainsi, un grand nombre d'études se sont intéressées à l'étude de la composition de la MOD à travers ses caractéristiques globales (e.g. dosage du carbone), optiques (e.g. UV-Visible, fluorescence d'excitation-émission, fluorescence synchrone) et moléculaires (e.g. spectroscopie de résonance magnétique nucléaire (RMN), chromatographie d'exclusion stérique (HPSEC), spectrométrie de masse à haute résolution (HRMS) couplée ou non à la chromatographie en phase liquide ou gazeuse) (Derrien et al., 2019). Parmi l'ensemble des techniques utilisées pour caractériser la MOD en fonction de ses sources, la spectroscopie UV-Visible et les matrices d'excitation-émission (EEM) de fluorescence se sont avérées pertinentes pour caractériser les fractions chromophores et fluorescentes de la MOD. Par exemple, le développement d'indicateurs optiques, tels que les rapports d'absorbance ou les indicateurs d'aromaticité SUVA et d'humification HIX, a permis de différencier la MOD provenant d'apports terrigènes par rapport aux apports autochtones (Murphy et al., 2008). Ces indicateurs fournissent des informations sur le poids moléculaire, l'aromaticité et la teneur en composés type humique des échantillons de MOD. Les décompositions des matrices de fluorescence ont également permis d'identifier des composés fluorophores spécifiques liés à des sources agricoles ou industrielles (ex : pesticides, hormones) (Hassoun et al., 2017). D'autres indicateurs plus récents, dérivés de la chromatographie d'exclusion stérique, ont permis de différencier les sources diffuses de MOD marine de celles de MOD de rivières, en se basant sur la distribution des poids moléculaires des composés contenus dans un échantillon (Asmala et al., 2021). Enfin, l'utilisation de techniques de spectrométrie de masse à haute résolution (notamment HRMS et spectrométrie de masse à résonance cyclonique ionique) a permis d'explorer l'échelle moléculaire et d'identifier différents groupes de

molécules spécifiques aux sources terrigènes (composés tels que les lignines, les lipides ou certaines protéines) ou aux sources urbaines et industrielles (e.g. Hydrocarbures Aromatiques Polycycliques, phytosanitaires, composés pharmaceutiques) (Choi et al., 2021). Au final, chaque technique fournit une information partielle sur la composition de la MOD, qui peut ainsi être reliée à un type de source de pollution. Malgré cela, l'utilisation de la MOD comme traceur des rejets de pollution via la compréhension des chemins de l'eau n'a pas toujours été réellement envisagée, d'une part parce qu'elle est souvent considérée comme très réactive (elle est en effet soumise à de multiples processus de transformation dont la biodégradation, la photodégradation et l'adsorption), et d'autre part à cause de la diversité et du grand nombre des apports potentiels de MOD dans le cours d'eau qui peuvent influencer sa qualité. Néanmoins, les développements analytiques récents, ainsi qu'une exploitation plus approfondie des données optiques et moléculaires (outils chimiométriques) ont ouvert de nouvelles perspectives qui permettraient l'utilisation de la MOD comme traceur pertinent des sources de pollutions anthropiques dans les cours d'eau.

Dans un contexte d'amélioration des stratégies d'identification de l'origine des masses d'eau et des pollutions associées dans les cours d'eau, l'objectif général de ma thèse est de mettre en évidence des marqueurs ou empreintes physico-chimiques propres à des types de sources de matière organique dissoute. Pour cela, une première étape a été d'identifier et d'échantillonner les sources *a priori* de la MOD à l'échelle d'un bassin versant soumis à un impact anthropique (mise en place d'une base d'échantillons). Dans un second temps, la MOD de chaque type de source et de chaque échantillon a été caractérisée avec une approche dite « intégrée ». Cette « caractérisation intégrée », consiste en une approche multi-analytique et en l'exploitation simultanée de l'ensemble des données générées par ces différentes techniques analytiques, afin d'identifier les caractéristiques et les propriétés physico-chimiques de la MOD étudiée. Dans une dernière étape, des approches statistiques univariées, multivariées et supervisées ont permis de mettre en évidence des empreintes physico-chimiques spécifiques de type de source. Celles-ci correspondent à une série d'indicateurs (ou descripteurs) de la MOD (e.g. aromaticité, humification, taille de molécules) calculés à partir des données générées par chaque technique. Après les études statistiques, les indicateurs retenus comme pertinents pour différencier des sources sont définis comme étant des marqueurs de ces sources.

Ce manuscrit est composé de sept chapitres. Le premier chapitre est consacré à la présentation du contexte de la thèse. Il est divisé en trois parties distinctes. La première partie expose les définitions globales de la MOD, ainsi qu'un état des lieux des connaissances concernant ses interactions à l'échelle écosystémique. La deuxième partie présente les sources potentielles de MOD dans les cours d'eau et identifie les verrous scientifiques liés à l'échantillonnage des sources diffuses. La troisième partie aborde les principales techniques d'analyses utilisées pour caractériser la MOD dans la littérature. Dans cette partie, les verrous méthodologiques liés au traitement et à l'exploitation des données HPSEC/UV-fluorescence et LC-HRMS sont plus particulièrement détaillés.

Le deuxième chapitre est consacré à la présentation de la démarche scientifique de la thèse. Un rappel plus complet de l'objectif général de cette dernière permet de formuler des questionnements scientifiques et des hypothèses de travail et d'établir des objectifs plus spécifiques. La démarche mise en place pour répondre à ces questions et atteindre ces différents objectifs est ensuite détaillée.

Le troisième chapitre concerne la présentation des matériels et méthodes utilisés pour répondre aux différentes questions abordées dans la thèse, à savoir, la méthode de construction de la base de données multi-analytique sur un large nombre d'échantillons issues de différentes sources de MOD. Plus précisément, la présentation des sites d'études, les stratégies d'échantillonnage mises en place, les méthodes analytiques et les approches statistiques pour l'étude des sources de MOD sont présentés dans ce chapitre.

Le quatrième chapitre est consacré à la résolution des verrous liés au traitement et à l'exploitation des données issues de la chromatographie HPSEC/UV-fluorescence et LC-HRMS. Une première partie porte sur la présentation de nouveaux indicateurs issus de la technique HPSEC/UV-fluorescence basés sur des pourcentages de contribution en poids moléculaire. Ces indicateurs sont comparés à des indicateurs classiques de distribution de poids moléculaire puis confrontés à des indicateurs issus d'autres techniques telles que l'UV-Visible et EEM de fluorescence. Une deuxième partie, sous la forme d'une publication scientifique (en cours de soumission), est consacrée au traitement des données issues de la technique LC-HRMS, avec la mise en place de contrôles qualité permettant de fiabiliser la donnée finale. Enfin, une troisième partie est consacrée à la mise en place d'indicateurs issus de la LC-HRMS pour

améliorer la caractérisation des sources de MOD. Ces indicateurs sont basés sur la répartition en masse et polarité des composés moléculaires.

Le cinquième chapitre est consacré à la présentation des résultats sur les sources diffuses, avec une attention particulière apportée à l'impact des usages de sols. La première partie, sous la forme d'un article scientifique (publié dans le journal *Science of the Total Environment*), traite de la problématique liée à l'échantillonnage des sources diffuses. Une étude comparative entre la MOD extraite de sol et la MOD prélevée dans les cours d'eau adjacents a ainsi permis de valider une méthodologie de prélèvement permettant d'obtenir un échantillon représentatif d'une source terrigène. La deuxième partie du chapitre concerne l'application de cette méthodologie et sa généralisation à des échantillons représentatifs de sources diffuses, prélevés dans d'autres bassins versants.

Le sixième chapitre, également sous la forme d'une publication scientifique (version prête à soumettre), présente l'étude finale de l'utilisation de la MOD en tant que traceur de sources de pollution dans les cours d'eau d'un bassin versant péri-urbain. Dans ce chapitre, les indicateurs présentés dans le quatrième chapitre et la méthodologie de prélèvement des sources diffuses présentée dans le cinquième chapitre ont été appliqués pour définir des empreintes de types de sources. De plus, d'autres empreintes de types de sources, notamment des sources terrigènes liées à une occupation de sols, des sources autochtones, des rejets de STEP, des rejets d'eaux domestiques non traitées, et des eaux de ruissellement pluvial urbain, ont été ajoutés à la base d'échantillons pour élargir le panel des sources étudiées.

Enfin, le septième et dernier chapitre présente les conclusions générales de la thèse et des perspectives ouvertes par ces travaux.

Chapitre 1.

Etat de l'art

Chapitre 1. État de l'art

1.1 Généralités sur la matière organique dissoute (MOD)

Le terme de matière organique (MO) est avant tout un terme conceptuel désignant l'ensemble des composés organiques, par opposition aux composés inorganiques, présents dans les milieux terrestre, aquatique, atmosphérique voire extra-terrestre. Dans le milieu aquatique, la MO est souvent définie comme un mélange complexe de molécules organiques ayant des compositions et origines très hétérogènes (Baldock et al., 2004). Cette définition très globale permet de regrouper deux aspects primordiaux dans l'étude de la MO. Le premier concerne l'origine de la MO. En effet, des apports naturels par dégradation des plantes et animaux par les microorganismes, et *a contrario* des apports liés aux activités humaines (rejets urbains, agricoles et industriels), permettent de distinguer une MO naturelle d'une MO anthropique (Guo et al., 2007; Kracht, 2001). L'origine peut également être abordée en fonction des compartiments écologiques, avec la MO d'origine terrestre ou aquatique. Le deuxième aspect concerne la complexité de la composition de la MO. Cette complexité réside dans la présence de composés biochimiques basiques (acides aminés, sucres, vitamines, acides gras), de biopolymères complexes (protéines, polysaccharides, lignines) mais également dans des composés de dégradation d'origine inconnue qui constituent encore à ce jour un réel défi dans la caractérisation de la MO (*e.g.* produits de transformation) (Aiken, 2014). En 2021, Català et al. décrivent la MO comme un vaste univers inexploré de molécules organiques. Les deux aspects, origine et composition, sont dépendants dès lors que l'origine est le premier facteur qui influence la composition de la MO.

D'autres définitions d'ordre pratique ont été attribuées à la MO. Par exemple, dans le milieu aquatique, une filtration de l'eau entre 0.2 et 0.7 μm (Figure 1-1) permet de séparer la matière organique particulaire (MOP) de la matière organique dissoute (MOD) (Sanderman et al., 2009) : la MOP est retenue par le filtre tandis que la MOD passe à travers. Traditionnellement, des filtres en fibre de verre de porosité de 0.7 μm sont utilisés pour la séparation. Cependant, de récentes études ont montrés qu'il n'existait pas de différence significative entre une filtration à 0.7 ou 0.2 μm sur la qualité de la MOD dans les milieux aquatiques (Denis et al., 2017). Actuellement, les filtres de petite taille (0,2 μm) sont souvent utilisés car ils sont

facilement disponibles et éliminent la plupart des bactéries susceptibles de dégrader un échantillon de MOD (Hartnett, 2018).

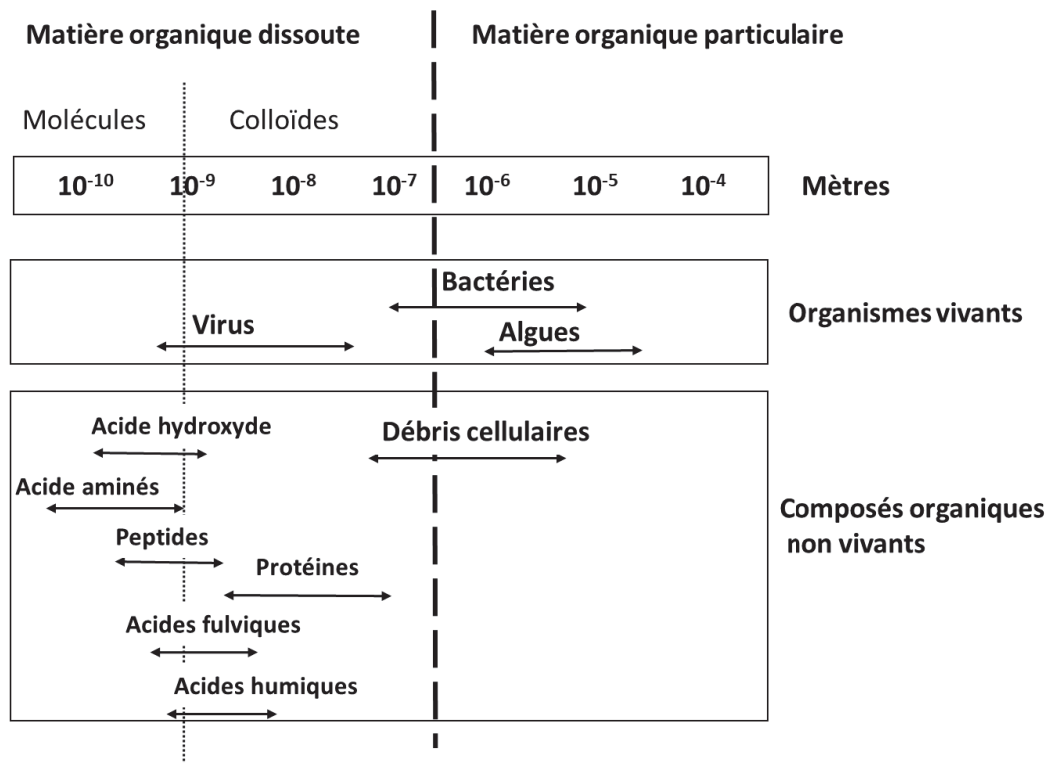


Figure 1-1. Continuum de la taille de la matière organique dans les eaux naturelles. Adapté de Stumm 1977

La MOD est présente dans toutes les masses d'eau, qu'elles soient marines ou d'eau douce. Il s'agit d'un mélange hétérogène de composés organiques provenant principalement d'apports terrigènes, des produits de décomposition des matières végétales, des microorganismes et des algues. La MOD contribue également de manière significative au cycle global du carbone, agissant comme un réservoir qui peut être transporté entre les écosystèmes terrestres et les systèmes aquatiques (Moran et al., 2000; Jiao et al., 2014).

De très nombreuses études se sont focalisées sur l'étude de la composition et des propriétés physicochimiques de la MOD (Figure 1-2). Ces propriétés sont complexes et diverses, et sont principalement définies au moyen de différentes techniques d'analyses. En effet, la MOD peut être décrite sous différents aspects, tels que sa concentration, sa composition élémentaire, sa taille, sa structure et ses groupes fonctionnels, qui fournissent autant d'informations essentielles sur ses propriétés physicochimiques afin d'étudier son rôle dans les processus biogéochimiques (Murphy et al., 2008; Matilainen et al., 2011; Wünsch et al., 2018; Zhang et al., 2021). Par exemple, la concentration de la MOD dans les échantillons est souvent

quantifiée en termes de concentration globale de carbone organique dissous (DOC). La taille des molécules organiques, qui varie entre une dizaine à des centaines de milliers de Daltons, peut être déterminée à l'aide d'une variété de techniques comme l'ultrafiltration ou la chromatographie d'exclusion stérique. La structure de la MOD, constituée d'une large gamme de composés organiques, comme les sucres, les acides aminés, les protéines, les lipides et les substances humique, ainsi que les groupes fonctionnels caractérisés par la présence d'acides carboxyliques et les phénols, peuvent être étudiés par des techniques comme la spectroscopie à résonance magnétique nucléaire (NMR), la spectroscopie infrarouge à transformation de Fourier (FTIR) et la spectrométrie de masse.

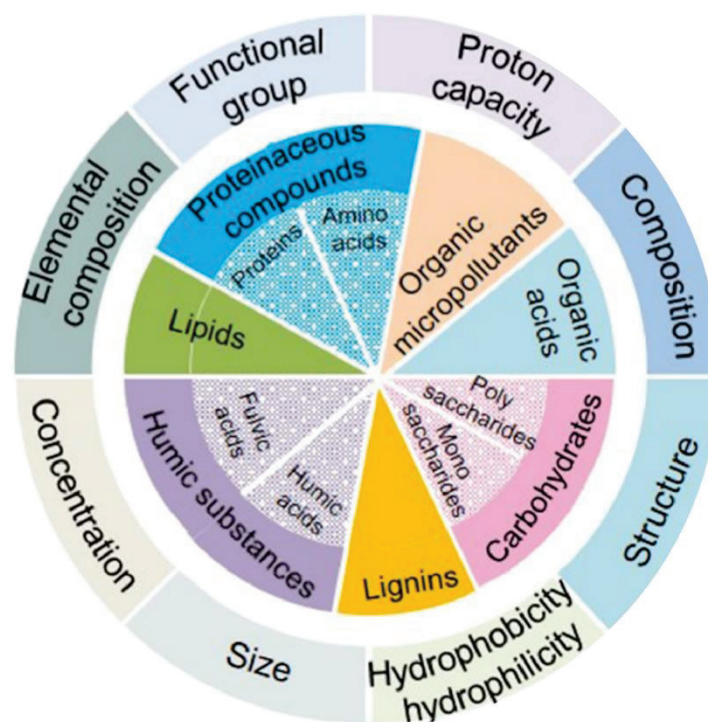


Figure 1-2. Composition et propriétés physicochimiques de la DOM. D'après Xiao et al., 2020

Il est important de noter que la composition et les propriétés de la MOD sont très dynamiques, et peuvent ainsi considérablement varier en fonction de facteurs environnementaux comme la température, le pH, la lumière et la disponibilité des nutriments (Aiken, 2014). La prise en compte des interactions de la MOD avec d'autres composants environnementaux tels les micro-organismes et les polluants est également essentielle pour comprendre les cycles biogéochimiques dans l'environnement (Figure 1-3) (Amin et al., 2012; Hertkorn et al., 2016).

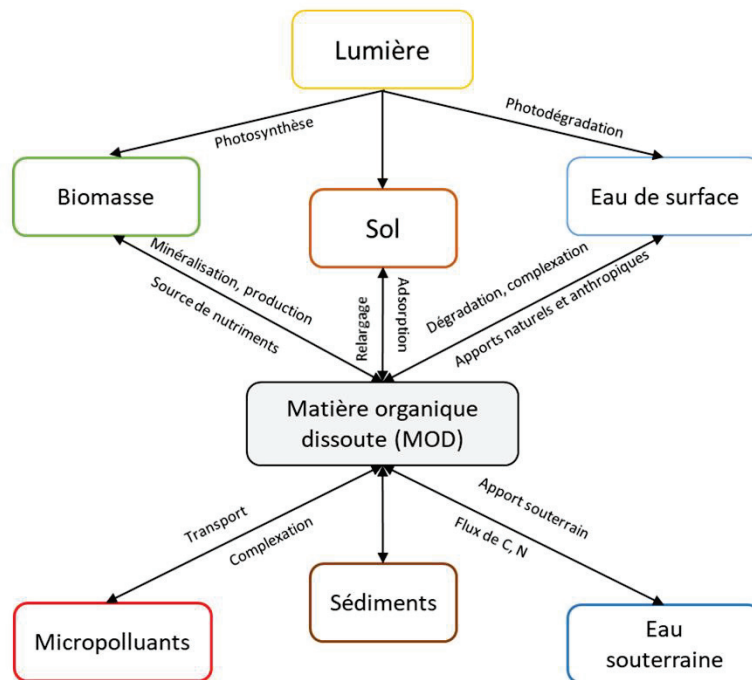


Figure 1-3. Interactions de la matière organique dissoute avec les différents compartiments écologiques

Les sources de MOD peuvent également influencer sa composition et ses propriétés physico-chimiques. En effet, de par son ubiquité, la MOD que l'on retrouve dans les rivières peut provenir non seulement de sources naturelles mais également de sources anthropiques (Figure 1-4). Ces sources peuvent être classées en fonction des apports, avec des sources dites allochtones qui sont à l'origine d'une MOD produite à l'extérieur du milieu aquatique, par opposition aux sources autochtones qui désignent une production *in situ*. Les sources allochtones peuvent être divisées en fonction du type de rejet, avec des sources ponctuelles et des sources diffuses. Les sources ponctuelles sont celles qui proviennent d'une source unique, identifiable ou canalisée, comme les eaux usées des stations d'épuration (STEP) ou des industries (Lapworth et al., 2008). Les sources diffuses, quant à elles, sont réparties sur une vaste zone et n'ont pas de source identifiable, comme le ruissellement des eaux de pluie sur les terres agricoles (Tran et al., 2019). Une source de MOD dans les rivières désigne donc le point ou la zone où les composés de la MOD sont introduits, directement ou indirectement, dans l'écosystème fluvial.

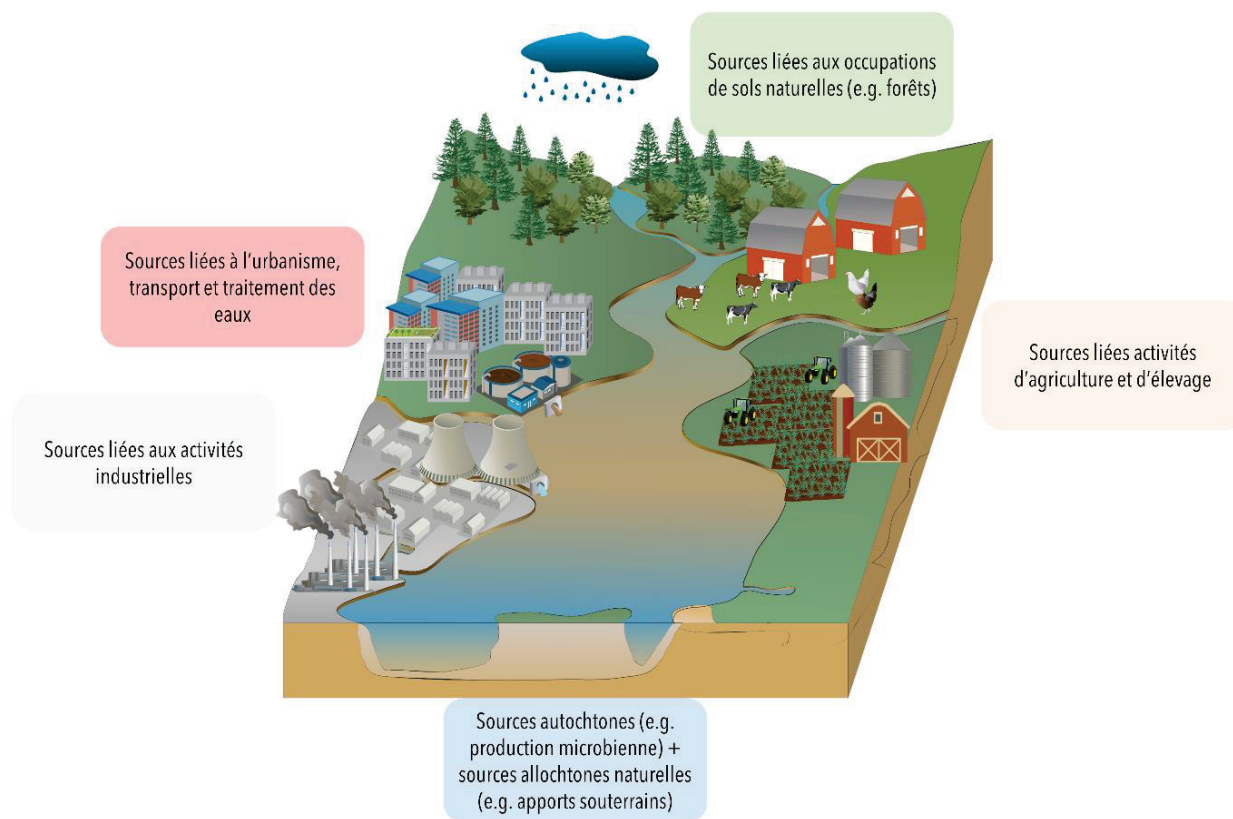


Figure 1-4. Illustration des principales sources de MOD à l'échelle d'un bassin versant

1.2 Les sources de matière organique dissoute

1.2.1 Sources autochtones

La contribution des sources autochtones dans le cycle de la matière organique dissoute peut représenter jusqu'à 20 % de la matière organique totale présente dans le milieu aquatique (Salonen et al., 1992). Cette contribution est cependant très variable selon l'impact climatique, l'hydrologie, les variations saisonnières ainsi que les apports en nutriments (Rosario-Ortiz, 2014).

La production algale est l'un des deux principaux phénomènes à l'origine de la matière organique autochtone (Green et al., 2019). En effet, les algues microscopiques et macroscopiques, à travers la photosynthèse, utilisent l'énergie solaire pour produire des composés à base de carbone. Ce carbone est ensuite libéré à la mort de la cellule (liée à la sénescence du phytoplancton), lors de l'utilisation de celle-ci par d'autres organismes hétérotrophes, ou bien transformé en tant que support pour des activités enzymatiques (Bertilsson and Jones, 2003). Le carbone excrété se retrouve ensuite dans des composés de

faibles poids moléculaire (<600 Da) : ces composés labiles (acides aminés et polymères des membranes cellulaires) et facilement assimilables sont caractéristiques de la MOD autochtone (Wu et al., 2018). D'autre part, la MOD influence également ces producteurs primaires. En effet, des études ont montré que l'augmentation des concentrations de MOD dans l'eau induisait directement l'augmentation de la biomasse du phytoplancton (Kissman et al., 2017). La MOD de faible poids moléculaire (*e.g.* matériaux de type protéique), ainsi que la MOD de poids moléculaire élevé (*e.g.* composés type humique) peuvent ainsi être, directement ou indirectement, par la photodégradation et biodégradation, absorbées par les producteurs primaires (Mayora et al., 2016). Cependant, les groupements aromatiques associés aux composés type humique et fulvique peuvent atténuer les rayonnements UV dans le milieu aquatique et, par conséquent, réduire la croissance de ces mêmes producteurs (Villafañe et al., 2018).

La dégradation microbienne de la MO dans le cours d'eau constitue la deuxième source autochtone. La biodégradation régule à la fois la quantité et la qualité de la MOD dans les cours d'eau (Sankar et al., 2019). En effet, les communautés microbiennes régulent la diversité des molécules de la MOD en dégradant les grosses molécules en petites molécules et en synthétisant des composés réfractaires à partir de substrat labile (Figure 1-5) (Hansen et al., 2016; Jiao et al., 2010). Ces processus conduisent à la modification des caractéristiques moléculaires associées à la MOD, tels que le poids moléculaire, la structure chimique ou encore la biodisponibilité dans différents environnements, tels que le sol, les sédiments et le milieu aquatique (Hu et al., 2022; Lennon and Pfaff, 2005). La composition de la MOD a un impact profond sur la structure des communautés microbiennes et sur leur potentiel métabolique (*e.g.* absorption directe, ectoenzymes). En effet, par exemple, la membrane externe de certaines bactéries permet une absorption des composés labiles inférieurs 1000 Da (monosaccharides). Quant aux composés de poids moléculaires supérieurs (> 1000 Da), ils sont généralement hydrolysés au niveau extracellulaire avant d'être métabolisés en utilisant une variété d'enzymes (Arnosti, 2003).

Des études ont également montré que les composés de type humique de la MOD peuvent inactiver les enzymes extracellulaires en formant des complexes humiques-enzymes, en particulier dans les écosystèmes d'eau douce et estuariens. Cela peut avoir un impact direct sur la structuration des communautés microbiennes présentes au niveau de ces milieux (Li et

al., 2019). En outre, les micro-organismes peuvent modifier les propriétés de la matière organique complexe par le biais de processus de biotransformation, et la composition de la MOD peut considérablement modifier la diversité des microorganismes.

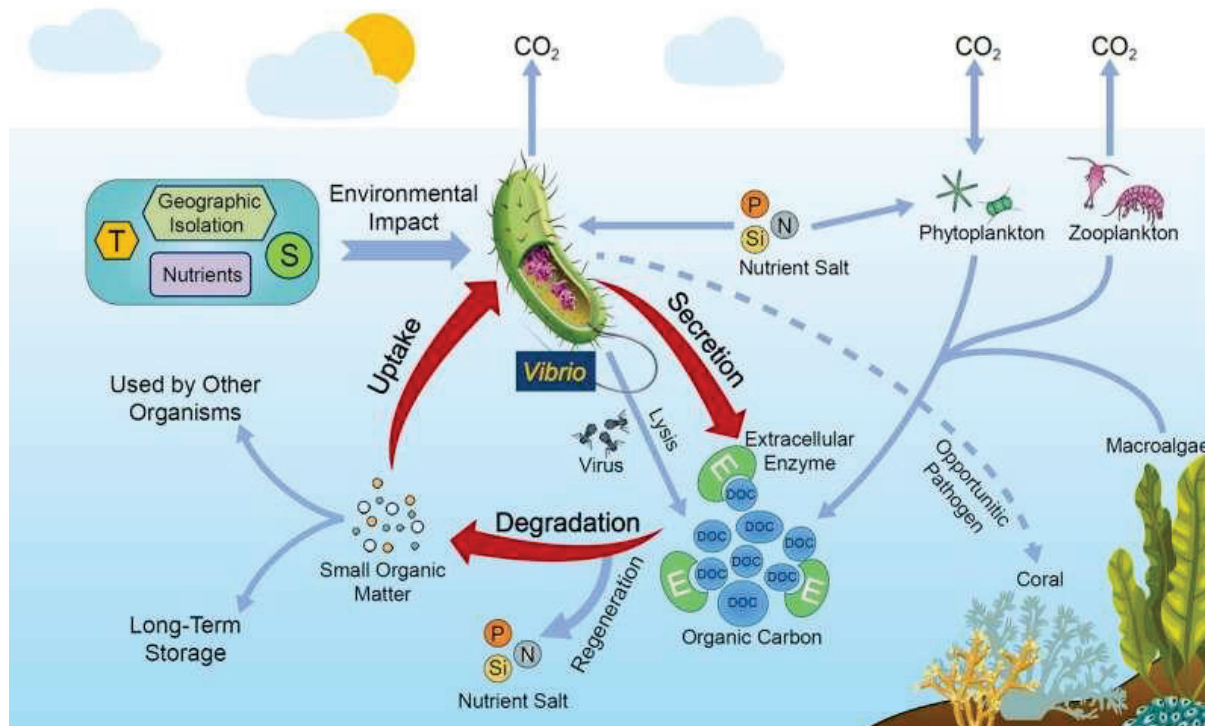


Figure 1-5. Interactions de la matière organique dissoute avec les microorganismes et les producteurs primaires à l'échelle du milieu aquatique. D'après Zhang et al., 2018

1.2.2 Sources allochtones

1.2.2.1 Sources diffuses

Les sources diffuses de MOD correspondent à des apports répartis sur une vaste zone avec une occurrence aléatoire et parfois intermittente. Parmi ces apports, les apports de MOD à partir du milieu terrestre environnant (apports dits terrigènes) représentent la principale contribution des sources diffuses (Aitkenhead-Peterson et al., 2003). Le transfert de la MOD du sol vers le milieu aquatique est régi par les caractéristiques du sol et l'hydrologie du bassin versant. Ainsi, la capacité du sol à constituer un réceptacle de matière organique conditionne le transfert de cette dernière dans le cours d'eau. En effet, un certain nombre d'études ont montré qu'un sol minéral augmentait l'adsorption du carbone, et de ce fait diminuait considérablement le relargage de MOD dans le milieu aquatique (Moore, 1989; Sachse et al., 2005). En plus du type de sol, différents types de ruissellements peuvent se produire et de ce fait contribuer à un apport différent de MOD dans les cours d'eau. En effet, le ruissellement

de surface se produit généralement après des précipitations ou la fonte des neiges, et peut transporter une variété de matières organiques dissoutes et particulaires provenant de la surface du sol (horizon organique) (Figure 1-6). Ces matières organiques, issues de la décomposition des végétaux et des litières, ont été documentées comme riches en composés de faibles poids moléculaires, tels que les acides aminés et les sucres simples (Murphy et al., 2008; Lupon et al., 2020). Le ruissellement de sub-surface ou intermédiaire se produit quant à lui lorsque le sol devient saturé et que l'eau est forcée de s'écouler latéralement à travers les couches du sol. Le ruissellement de sub-surface peut alors transporter des matières organiques dissoutes provenant de couches de sol plus profondes qui ont tendance à contenir une plus grande proportion de composés aromatiques et résistants à la dégradation (Sanderman et al., 2009).

Le sens des écoulements n'est toutefois pas le seul phénomène qui influence les sources diffuses terrigènes. En effet, la MOD provient principalement de l'espace macroporeux des particules, et sa composition chimique peut être également liée à aux usages de sols (Mangal et al., 2020). À l'échelle d'un bassin versant, il est ainsi possible de différencier les sources diffuses terrigènes naturelles et anthropiques. Les sources terrigènes naturelles de MOD comprennent les apports de litière végétale issue de sols non impactés par les activités anthropiques (*e.g.* couverture végétale naturelle de forêts et prairies). En revanche, les sources terrigènes anthropiques de MOD sont associées aux activités humaines, telles que l'agriculture, l'élevage ou l'urbanisation. Les polluants associés à ce type d'activités (*e.g.* pesticides, nutriments, pathogènes) sont transportés par ruissellement à travers la surface du sol et se retrouvent dans le milieu aquatique, influençant ainsi la qualité de la MOD. En effet, des études ont montré qu'une matière organique issue d'un milieu soumis à des rejets urbains possède une composition différente de la MOD naturelle (Park, 2009). Ces différences sont observables au niveau moléculaire et macromoléculaire.

D'autres phénomènes tels que l'humidité des sols et l'évapotranspiration influencent la contribution des sources diffuses de MOD (Ritter and Shirmohammadi, 2000). Des études ont ainsi montré qu'un sol forestier permettait d'intercepter environ 20 % des eaux de pluie (Sheng and Cai, 2019) contrairement à des sols cultivés où le taux d'interception avoisine les 1 % (Ma et al., 2014). L'évapotranspiration, qui désigne la perte d'eau d'un couvert végétal en

fonction de la réserve d'eau du sol (Mjeira, 2015), impacte considérablement la saturation en eau des sols, et ainsi le processus de lixiviation.

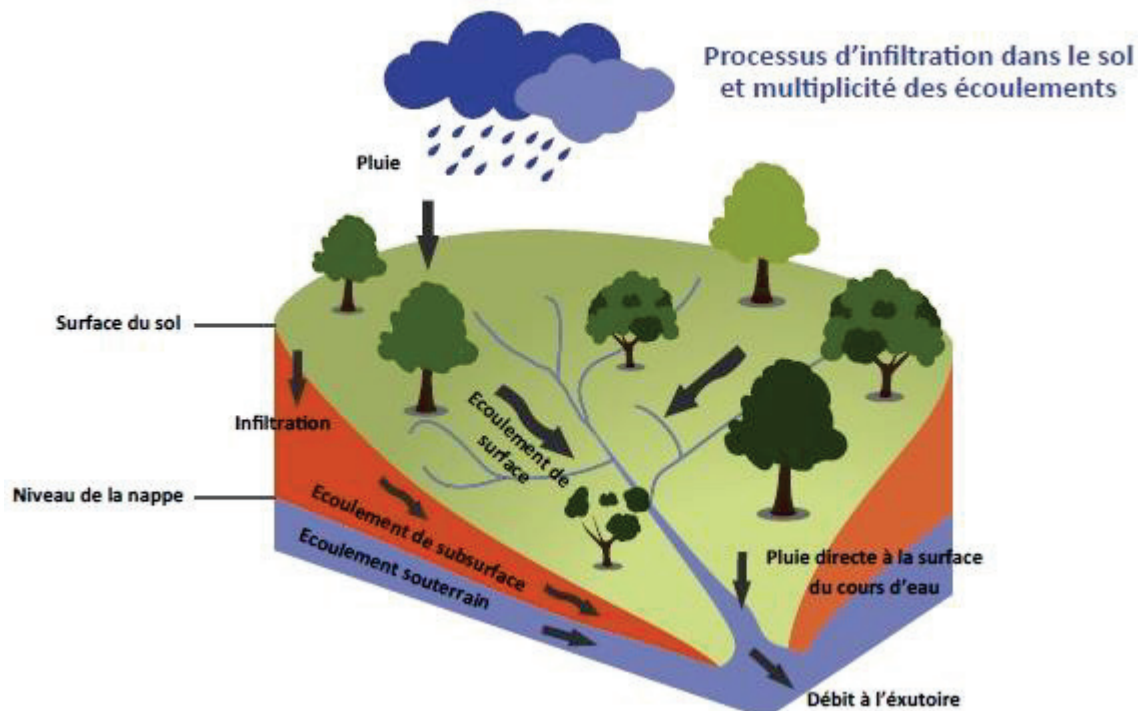


Figure 1-6. Schéma des principaux phénomènes intervenant dans le transfert des sources diffuses terrigènes. Image : Pr. André Musy. Cours d'hydrographie générale.

Le second type de sources diffuses concerne les apports souterrains. Les flux d'eau entre le compartiment souterrain et la rivière ont fait l'objet de nombreuses études (Cartwright et al., 2014; Sophocleous, 2002). En effet, les eaux souterraines représentent la principale source d'eau douce pour les activités agricoles, les usages industriels et l'approvisionnement en eau potable (Rutledge et al., 2021). La contribution en matière organique par le compartiment souterrain est difficilement quantifiable, car celle-ci est variable selon les conditions hydrologiques des bassins versants et le degré de connexion entre nappes et rivière (influence de la géologie) (McDonough et al., 2022). Ainsi, en période de sécheresse, par exemple, l'apport des eaux souterraines peut devenir une source importante, voire unique, pour les cours d'eau, contrairement aux périodes pluviales (Kaule and Gilfedder, 2021). Malgré le manque de connaissances actuelles sur le sujet, plusieurs études s'accordent sur le fait que les apports de carbone et azote depuis l'aquifère seraient inversement proportionnels à leur concentration dans la colonne d'eau, en période de basses eaux. Ainsi, un équilibre dans le flux de ces deux constituants majeurs de la OM (carbone et azote) entre la surface et les eaux

souterraines est établi (Chen et al., 2010; Webb et al., 2019). Des changements dans la qualité de la matière organique ont par ailleurs été observés dans les eaux souterraines (Figure 1-7). La MOD présente dans ces eaux est principalement caractérisée par une augmentation du poids moléculaire moyen et de l'aromaticité et une faible présence de molécules labiles biodisponibles (Rutledge et al., 2021).

D'autres sources naturelles allochtones peuvent également contribuer à l'apport de la MOD dans les milieux aquatiques, tels que les apports directs des eaux de pluie (sans phénomène de lixiviation des sols). Ces sources sont à l'origine d'apports de particules organiques, le plus souvent assimilées à du pollen, des aérosols ainsi que des particules de poussière (Li et al., 2016).

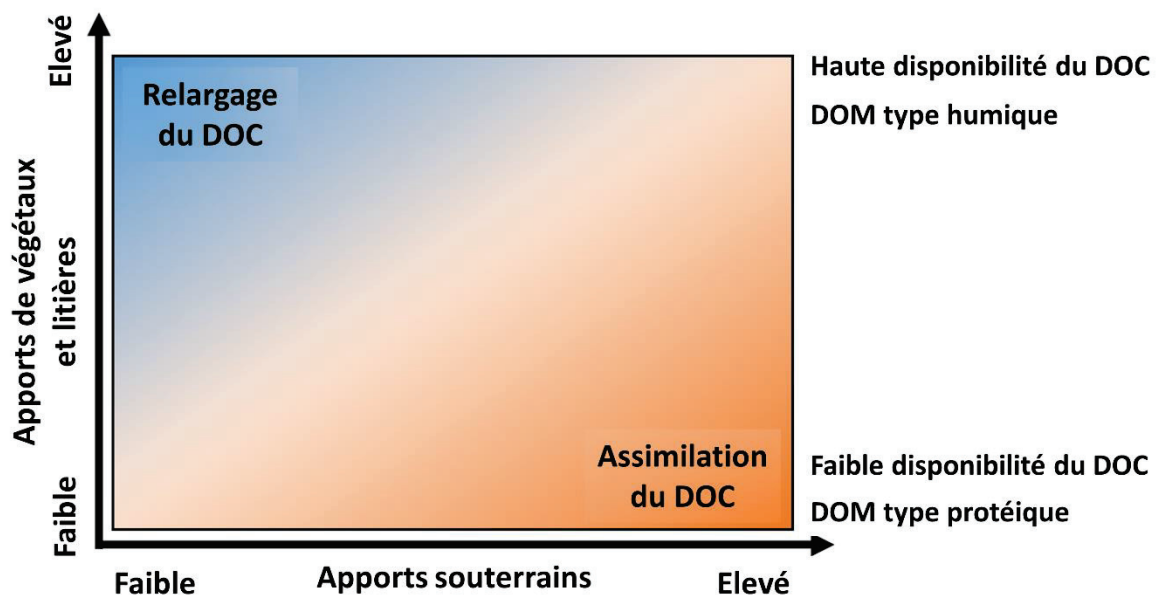


Figure 1-7. Caractéristiques de la MOD en fonction des écoulements dans les rivières. Adapté de Lupon et al., 2020

1.2.2.2 Sources de rejets ponctuels

Les rejets ponctuels (i.e. canalisés) de MOD dans l'environnement sont principalement dus aux activités domestiques et industrielles. En effet, la transformation des terrains naturels en zone commerciales, urbaines et résidentielles augmente considérablement la quantité de pollution rejetée dans les cours d'eau. Les effluents qui drainent ces activités présentent généralement des concentrations plus élevées en carbone et en polluants, entraînent des

effets nocifs sur la santé humaines et provoquent des altérations de la qualité du cours d'eau avec des répercussions sur les activités économiques (Meyer et al., 2005). La MOD intervient dans le transport et devenir des polluants, interagissant avec ces derniers via différents mécanismes de réactivité, de chélation et de sorption. Des études ont ainsi mis en évidence une altération de la biodisponibilité (Halim et al., 2003), de la mobilité (Holm et al., 2003) et du devenir (Evangelou et al., 2007) des polluants organiques et inorganiques par la MOD, et ce dans les écosystèmes marins, estuariens et côtiers. Les groupements fonctionnels tels que les carboxylates, les phénols et les amines permettent quant à eux de complexer les éléments traces métalliques (ETM). Lemaire et al. (2015) ont montré que la MOD autochtone (produite localement par l'activité des microorganismes) favorisait également la complexation des ETM, réduisant ainsi leur biodisponibilité et toxicité, par rapport à une MOD d'origine urbaine.

D'autres aspects de la MOD tels que le degré d'aromaticité et le poids moléculaire ont un impact direct sur la partition (entre dissous et particulaire) et la sorption des polluants organiques dans les sols et les sédiments (Gevao et al., 2000). Les composés organiques tels que les substances de type humique et de type fulvique, qui sont très hydrophobes et aromatiques, sont des sorbants efficaces pour certains polluants organiques tels que les hydrocarbures aromatiques polycycliques (HAP) (Chianese et al., 2020). En effet, les propriétés de la molécule organique considérée et son affinité pour le carbone organique conditionnent son adsorption avec les molécules constitutives de la MOD par diverses liaisons comme les liaisons de Van der Waals (Han et al., 2021).

La MOD est également responsable de la formation de sous-produits de transformation (TPs), notamment des sous-produits de désinfection dans les stations de traitement d'eau potable. La MOD réagit ainsi avec les désinfectants tels que l'ozone et le peroxyde d'hydrogène pour produire des intermédiaires réactifs, tels que les hydroxyles radicaux. Ces intermédiaires réactifs réagissent ensuite avec les composés organiques présents dans l'eau pour produire des TPs (Peleato et al., 2017). La réaction de la MOD avec le chlore induit quant à elle la formation de produits de transformations tels que les trihalométhanes et les acides haloacétiques, qui sont connus pour leur effets néfastes sur la santé humaine (Hood, 2005).

1.2.2.2.1 Rejets de stations de traitement des eaux usées

Les stations de traitement des eaux usées (STEU) domestiques ou industrielles sont essentielles pour canaliser et traiter les fortes concentrations en matière organique et polluants avant leur rejet dans les milieux récepteurs. Pour cela, plusieurs procédés de traitements peuvent être appliqués (Figure 1-8). Les procédés de traitement diffèrent en fonction des types de rejets en amont. Les rejets de collectivités de grande taille (> 5000 habitants) sont traités dans des stations qui recourent à des traitements intensifs (e.g. boues activées plus procédés d'oxydation), tandis que les rejets de petites collectivités (< 5000 habitants) peuvent également être traités non seulement avec ces traitements intensifs mais aussi avec des procédés dits extensifs (e.g. lagunage naturel, filtre plantés de roseaux).

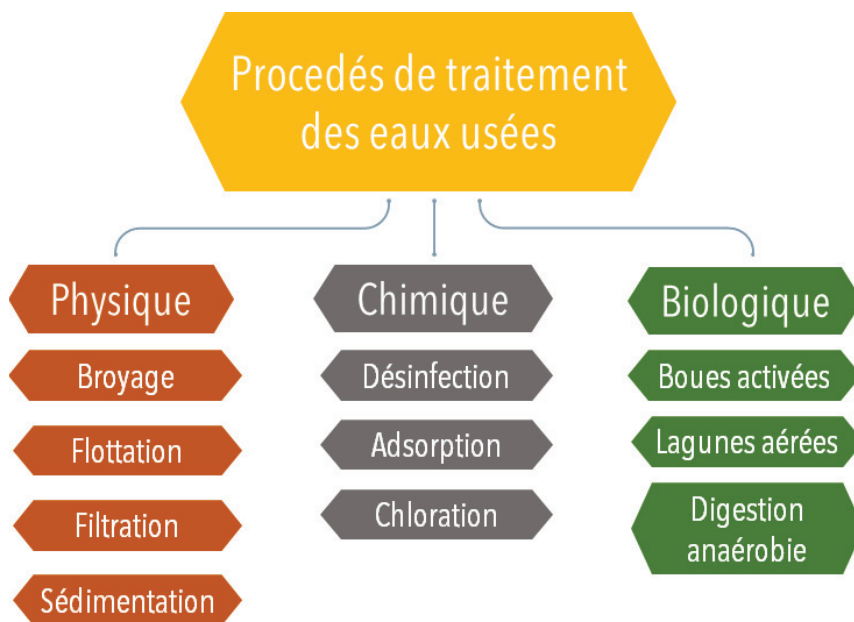


Figure 1-8. : Exemples de procédés de traitement des eaux usées (adapté de Simate, 2015)

i. Traitement par boues activées

Le traitement par boues activées est un processus basé sur la dégradation biologique et l'élimination de la matière organique et des polluants par des micro-organismes. La procédure se déroule dans un certain nombre de bassins de traitement, qui sont aérés afin que les bactéries puissent métaboliser la matière organique et les polluants. Les eaux usées sont mélangées aux boues activées, qui correspondent à des masses contenant des bactéries préalablement sélectionnées pour leur affinité avec les substrats présents dans les eaux usées. Ces micro-organismes décomposent la matière organique présente dans les eaux usées, qui

leur sert de source de nutriments. Une fois que les eaux usées ont été traitées dans les bassins d'aération, elles sont pompées dans une cuve de décantation au fond de laquelle se déposent les boues activées. Les eaux usées clarifiées peuvent alors être rejetées, tandis que les boues décantées sont alors recyclées dans les bassins d'aération (une partie est éliminée en tant que boue excédentaire afin de maintenir une communauté microbienne souhaitée). Ce processus d'épuration est considéré comme un traitement intensif, qui permet de traiter une grande partie de l'assainissement collectif des eaux résiduaires urbaines en France (40 % des STEU en France utilisent ainsi les boues activées) (Hocquet and Golla, 2008).

ii. Traitement par filtres plantés de roseaux

Les traitements à base de filtres plantés de roseaux représentent une filière de traitement extensif basé sur la capacité de remédiation naturelle des eaux usées domestiques, et municipales et des eaux pluviales et agricoles (Boram, 2014). Il s'agit de systèmes constitués d'un certain nombre de bassins peu profonds, ou canaux, remplis d'un substrat (*e.g.* sable, gravier), et plantés avec des roseaux. L'originalité de la filière française est d'accepter des eaux usées brutes directement à la surface des filtres. Les particules sont retenues à la surface du filtre et la matière organique particulaire contenue dans cette couche de dépôt est progressivement minéralisée par les microorganismes. La MOD, qui percole à travers le filtre, subit elle aussi une dégradation aérobie par les microorganismes présents dans le substrat et au niveau des racines des plantes, qui permettent la pénétration de l'air et de l'oxygène dans le filtre par action mécanique. Les plantes peuvent également absorber les nutriments tels que l'azote et le phosphore lorsque les eaux usées traversent le filtre. Enfin, d'autres types de filtres peuvent être ajoutés en aval pour assurer le traitement de l'azote (nitrification en milieu anaérobie) ou du phosphore (*e.g.* filtre d'apatite) (Molle, 2012). Ces filtres sont considérés une option de traitement abordable et durable pour les communautés disposant de ressources limitées.

iii. Caractéristiques de la MOD dans les eaux résiduaires

Les eaux résiduaires sont influencées par les rejets en amont, le type de réseau (unitaire ou séparatif) ainsi que le nombre d'habitants raccordés à celui-ci (Garnier et al., 2006; Goffin et al., 2018). Cela influence par conséquent la composition de la MOD, qui reste difficile à caractériser en raison des fortes variabilités observées dans ces eaux usées. En effet, des

variations qualitatives sont observées durant les heures de la journée (Le et al., 2017) en raison des cycles journaliers des activités domestiques ou industrielles. Néanmoins des études ont rapporté la présence de composés comme des protéines et acides aminés (les composés organiques les plus courants dans les eaux usées), des lipides (tels que les acides gras et les triglycérides également présents dans ces eaux bruts, Gorini et al., 2011) et des substances de type humique et des surfactants organiques.

En sortie de STEU, la MOD des effluents traités est constituée d'un mélange complexe de molécules organiques (e.g composés de type humique, type fulvique, carbohydrates, protéines), de produits de dégradation microbienne, de surfactants, de produits pharmaceutiques et phytosanitaires ainsi que de sous-produits de désinfection (Figure 1-9) (Michael-Kordatou et al., 2015; Verkh et al., 2018). Un grand nombre de ces composés sont réfractaires aux traitements biologiques des STEU et se retrouvent rejetés dans l'environnement.

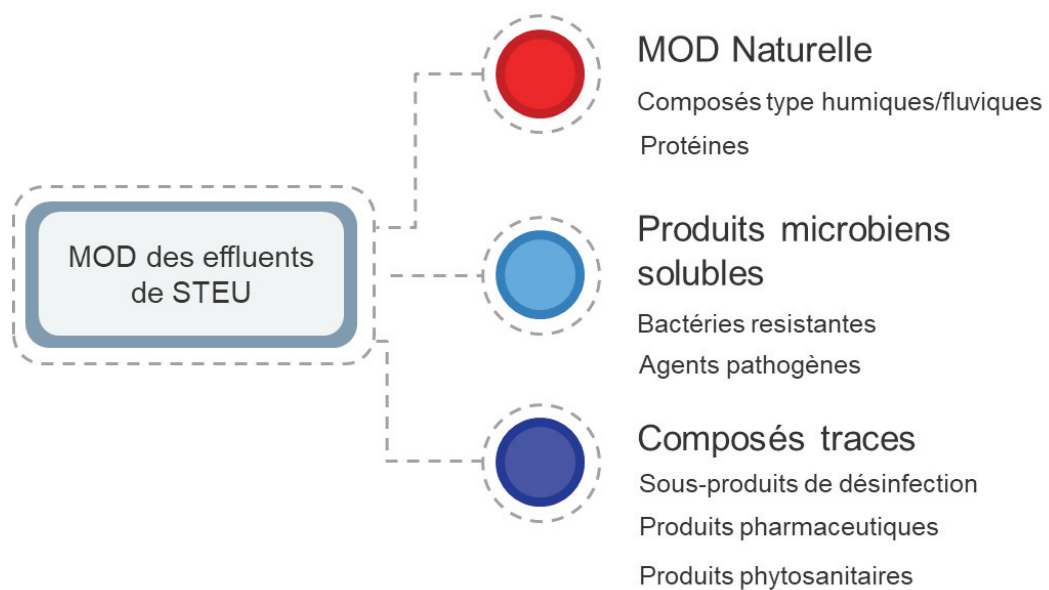


Figure 1-9. Principales composantes de la MOD en sortie de stations de traitement des eaux usées (STEU).

1.2.2.2.2 Ruissellement urbain

Le ruissellement urbain canalisé est une source importante de MOD dans les cours d'eau en milieu urbain. En effet, lorsque l'eau de pluie ruisselle sur des surfaces imperméables telles que les routes, les toits et les trottoirs, elle transporte une matière organique chargée en composés naturels et synthétiques. Ces composés proviennent notamment de la dégradation

des végétaux et de déchets animaux, de dépôts de poussière et des activités anthropiques (*e.g.* transport routier). Des recherches antérieures ont montré la présence importante de métaux (*e.g.* cuivre, plomb, zinc) dans les eaux de ruissellement provenant de zones industrielles et urbaines denses (Brown and Peake, 2006). D'autres polluants organiques, qui comprennent différentes classes de contaminants, tels que les produits ménagers, les pesticides, les produits pharmaceutiques et les produits de soins personnels, ont été associés aux eaux de ruissellement urbain (Gasperi et al., 2022). L'ensemble de ces paramètres influence la qualité de la MOD dans le milieu récepteur. En effet, une précédente étude a montré la présence d'une MOD de composition hétérogène présentant des marqueurs terrigènes (composés de type humique) mais également des marqueurs de production autochtone (composés aliphatiques peu humifiés) (Liao et al., 2021).

Dans le cas de réseaux unitaires, et lors d'événements pluvieux importants, les eaux de rejets domestiques habituellement canalisés se retrouvent dans des déversoirs d'orage qui servent de trop-plein pour alléger le débit d'eau acheminée vers les STEU. Ces déversoirs d'orage constituent ainsi des sources ponctuelles de MOD issue d'un mélange d'eaux usées non traitées et d'eaux issues du ruissellement d'eau de pluie en zone urbaine. A titre d'exemple, à l'échelle du bassin versant péri-urbain de l'Yzeron (bassin de l'ouest lyonnais d'une superficie de 147 km²), les volumes d'eau rejetés par les déversoirs d'orage vers la rivière en 2020 représentent tout de même 3 à 4% du flux annuel de la rivière (Alkhoury, 2022).

1.2.3 Caractérisation de la MOD dans les différentes sources

Les sources diffuses de MOD sont difficiles à localiser et à échantillonner en raison de la complexité des interactions géochimiques, des processus de dégradation et des événements météorologiques incontrôlés à l'origine de ces sources. Dans le cas des contributions terrigènes des sources diffuses de MOD, il est très important de prendre en compte le rôle des usages et des occupations des sols (*e.g.* agriculture, urbanisme) sur la qualité de la MOD dans les cours d'eau. Cependant, la question d'un échantillonnage représentatif d'une source terrigène (impactée par les usages de sols) se pose. En effet, des études ont montré que le type de sol régit la quantité de MOD transférée vers l'environnement aquatique en aval (Bolan et al., 2011). Le temps de séjour de l'eau dans un écosystème peut de plus influencer de manière significative les sources terrigènes de MOD (Du et al., 2021). Enfin, les conditions

météorologiques, comme les précipitations et la température, peuvent influencer les sources de MOD de plusieurs manières. De fortes précipitations peuvent par exemple entraîner une augmentation des apports de matière organique terrestre, tandis que les changements de température peuvent affecter la quantité de MOD produite par les producteurs primaires et les microorganismes (Lu et al., 2021; Xia et al., 2022). A ce jour, la difficulté de distinguer MOD autochtone et MOD allochtone dans l'environnement et les limites des tests de laboratoire sont des défis pour l'étude de la composition et du devenir de la MOD autochtone dans le milieu aquatique (Hoikkala et al., 2015; Xia et al., 2022). Cependant, la plupart des études parviennent à prendre en compte la variabilité des conditions hydrologiques et des températures de l'eau lors de prélèvements d'échantillons représentatifs des sources autochtones (Williams et al., 2010; Fellman et al., 2011). En effet, la MOD autochtone provient de processus biologiques, comme la photosynthèse, la respiration et la décomposition de la matière organique, qui sont eux-mêmes influencés par la température de l'eau, l'intensité lumineuse, la disponibilité des nutriments et le débit de l'eau (Toming et al., 2013). Pendant les périodes de hautes eaux, la contribution en MOD dans les cours d'eau est principalement due à des écoulements de surface et sub-surface. Cela résulte en un apport accru de matière organique terrestre dans la masse d'eau, ce qui peut diluer la concentration de la MOD autochtone et rendre difficile l'obtention d'un échantillon représentatif. D'autre part, pendant les périodes de basses eaux, le débit est réduit et provient principalement des écoulements profonds et des nappes apportant peu de matière organique terrestre. Il peut en résulter des concentrations plus élevées de MOD autochtone, ce qui facilite l'obtention d'un échantillon représentatif.

L'ensemble de ces phénomènes rend difficile l'établissement d'empreintes de MOD issue de sources diffuses dans les rivières. Néanmoins, dans une optique de traçage des sources de pollution, il est primordial d'obtenir des marqueurs propres à ces sources, qui sont très présentes à l'échelle des bassins versants. Certaines études ont ainsi tenté d'obtenir des échantillons représentatifs en collectant des échantillons dans les cours d'eau adjacents à un usage de sol spécifique (McElmurry et al., 2014; Heinz et al., 2015; Reader et al., 2015; Riedel et al., 2016; Liu et al., 2019). Cependant, ces études se limitaient souvent à des zones localisées et ne comportaient qu'un faible nombre d'échantillons et de sources, ce qui rend difficile la généralisation de leurs résultats.

1.3 Les techniques de caractérisation de la matière organique dissoute

Diverses techniques analytiques ont été mises au point pour caractériser la MOD du niveau global au niveau moléculaire. Ces techniques peuvent fournir des informations sur la quantité, la composition, la structure et la réactivité de la MOD.

1.3.1 Dosage du carbone organique dissous (DOC)

Les premières études sur le dosage du carbone organique dissous datent des années 1960 (Menzel and Vaccaro, 1964). Toutefois, ce n'est qu'une décennie plus tard, avec l'invention de l'analyseur TOC (carbone organique total), qu'on a commencé à l'utiliser pour décrire les flux de carbone dans les eaux naturelles (Rosario-Ortiz, 2014). Le carbone organique dissous constitue l'élément majeur dans la composition de la matière organique (58 % de la masse totale de la MO du sol) (Bianchi et al., 2008). Il est mesuré dans les eaux naturelles/usées, après une filtration des échantillons (à partir de 0.7 μm), par une méthode analytique reposant le plus souvent sur le principe de la combustion catalytique. Le CO_2 produit lors de cette phase d'oxydation est alors quantifié à l'aide d'une détection infrarouge non dispersive (NDIR).

Dans les eaux naturelles à caractère oligotrophe, la concentration de DOC dans le milieu varie entre 1 et 30 mg/L dans les fleuves et les rivières (Petitjean et al., 2004). Les plus faibles teneurs sont observées au niveau du milieu marin, avec une concentration médiane de 0.5 mg/L (Kalbitz et al., 2000). Le carbone étant l'élément majoritaire dans la composition de la MOD, son dosage permet de renseigner sur la quantité de MOD présente dans un échantillon. Le calcul du rapport C/N (carbone sur azote) au niveau des sols d'un bassin versant est également un indicateur des sources de matière organique dans les cours d'eau. Cet indice reflète le degré d'évolution de la matière organique : ainsi, un ratio C/N compris entre 1 et 15 indique une minéralisation rapide et un apport continu d'azote (caractéristique des sols agricoles), tandis qu'un ratio C/N élevé (>20) est un indicateur d'un sol dont l'immobilisation microbienne est plus importante que la minéralisation, ce qui correspond généralement à des sols de type forestier (Aitkenhead-Peterson et al., 2003).

1.3.2 Analyse des isotopes stables

Par définition, deux isotopes d'un même élément sont deux atomes présentant le même nombre de protons (Z) et un nombre différent de neutrons (N). Les isotopes du carbone ^{12}C ,

^{13}C et ^{14}C possèdent tous 6 protons, mais leur nombre respectif de neutrons est de 6, 7 et 8. Le carbone ^{12}C et ^{13}C sont des isotopes dits stables (ratio $N/Z < 1.5$) et le ^{14}C est un isotope radioactif. (Unkovich et al., 2013). Dans le milieu naturel, différents processus physiques, chimiques et biologiques peuvent conduire à la formation d'isotopes (on parle alors fractionnement isotopique). Un exemple de fractionnement isotopique du carbone est celui dû à la photosynthèse, au cours de laquelle certaines plantes absorbent de préférence les molécules de CO_2 contenant l'isotope léger du ^{12}C plutôt que celles contenant l'isotope plus lourd du ^{13}C (Lambert, 2013). Ce fractionnement isotopique induit des différences de compositions isotopiques (*e.g.* $\delta^{13}\text{C}$, ratio des isotopes stables $^{13}\text{C}/^{12}\text{C}$) utilisées dans des contextes de traçage de sources de MOD (apports terrigènes, production autochtone) et dans l'étude de processus de transformation de la matière organique (Derrien et al., 2019). En effet, la MOD terrestre est généralement enrichie en ^{13}C et appauvrie en ^{15}N par rapport à la MOD d'origine microbienne, qui est enrichie en ^{15}N en raison de l'absorption préférentielle de ^{14}N par les microorganismes. Les expériences de marquage isotopique peuvent ainsi être utilisées pour suivre le devenir de composés organiques spécifiques dans la MOD et d'identifier les voies et les cinétiques de dégradation de ces composés. D'autres isotopes stables (*e.g.* deutérium, ^{18}O , ^{11}B) ont également été utilisés pour identifier les sources de pollutions dans les eaux de surface et les eaux souterraines (Penna et al., 2013; Carrey et al., 2021).

1.3.3 Techniques de caractérisation optique

Les propriétés optiques de la MOD sont étroitement liées à sa composition et peuvent fournir des informations importantes sur sa structure moléculaire et sa réactivité (Aiken, 2014). Deux techniques sont souvent utilisées pour étudier les propriétés optiques de la MOD : la spectroscopie d'absorbance UV-Visible et la spectroscopie de fluorescence.

1.3.3.1 Spectroscopie UV-visible

En spectroscopie UV-Visible, un échantillon est placé dans une cuvette et exposé à une lumière émise par une source d'une longueur d'onde spécifique obtenue à l'aide d'un monochromateur. La quantité de lumière absorbée par l'échantillon est mesurée en comparant l'intensité de la lumière incidente à l'intensité de la lumière transmise à travers l'échantillon (Figure 1-10). La spectroscopie UV-Visible est une technique analytique souvent utilisée pour caractériser la MOD naturelle et anthropique. L'un des principaux avantages de

cette technique réside dans sa rapidité et sa simplicité d'utilisation. Cette technique est de plus non destructive et nécessite une préparation minimale de l'échantillon. De plus, les spectres obtenus par spectroscopie UV-Visible peuvent être utilisés pour estimer une variété de paramètres importants liés à la MOD.

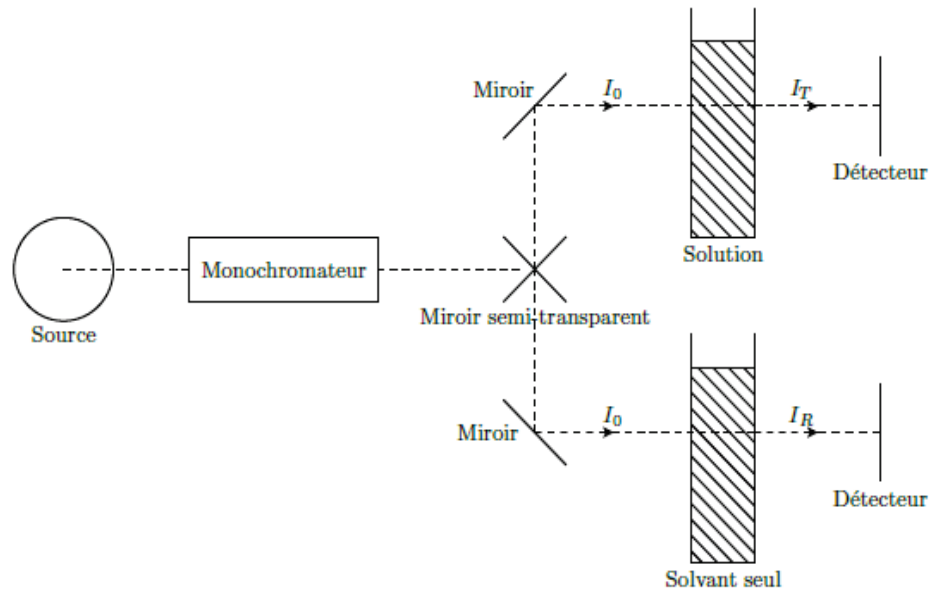


Figure 1-10. Principe de mesure d'un spectre d'absorption (Luciani, 2008)

La spectroscopie UV-Visible permet de mesurer la quantité de lumière absorbée à chaque longueur d'onde dans un domaine compris entre 200 et 800 nm. En ce qui concerne la MOD, les chromophores absorbent essentiellement dans les régions de l'UV et du bleu (Twardowski et al., 2004). L'allure des spectres générés augmente de manière quasi-exponentielle avec la diminution de la longueur d'onde (Helms et al., 2008a). Le spectre d'absorbance brut, noté A (Eq. 1-1) est souvent converti en coefficient d'absorption exprimé en m^{-1} et noté $a(\lambda)$ (Eq 1-2). L'intérêt principal de ce coefficient d'absorption est qu'il normalise l'absorbance par la taille de la cuve et permet une comparaison plus facile des résultats entre différentes études.

$$A = \log \frac{I_0}{I} \quad \text{Eq. 1-1}$$

$$a(\lambda) = \frac{2.303 \cdot A}{L} \quad \text{Eq. 1-2}$$

Où A représente l'absorbance brute, I_0 l'intensité de lumière émise, I l'intensité de lumière transmise, $a(\lambda)$ le coefficient d'absorption (m^{-1}) et L la longueur de la cuve (m).

i. Prédiction des concentrations en DOC

Le coefficient d'absorption à certaines longueurs d'onde (*e.g.* 254, 280, 285 nm) a été utilisé pour prédire les concentrations de DOC dans des échantillons issus de rivières, lacs et estuaires. Plusieurs modèles de prédiction de DOC ont été proposés dans la littérature car il est difficile, à ce jour, d'établir un modèle de prédiction universel. Ainsi, des modèles se basant sur une seule ou plusieurs longueurs d'onde sont rapportés pour prédire le DOC dans différents milieux (Brezonik et al., 2015; Kim et al., 2016; Shen et al., 2016).

ii. Rapports d'absorbance

En plus de l'aspect quantitatif, les rapports d'absorbance, définis comme les rapports des coefficients d'absorption à deux longueurs d'onde différentes, ont été fréquemment utilisés pour caractériser la composition chimique de la MOD chromophorique (Li and Hur, 2017). Par exemple, le rapport E2.E3, qui correspond au ratio d'absorption entre 250 et 365 nm, est souvent utilisé comme indicateur de la taille des molécules et macromolécules organiques (Helms et al., 2008b). Le rapport E2.E4 (ratio 254/436 nm) permet quant à lui d'estimer la contribution relative d'une source autochtone de MOD par rapport à une source terrestre (Jaffé et al., 2004). Il existe également d'autres ratios d'absorption référencés dans la littérature, comme le E3.E4 (ratio 300/400 nm) et le E4.E6 (ratio 465/665 nm) qui reflètent le degré d'humification de la MOD (Claret et al., 2003; Ikeya et al., 2015).

iii. Les pentes spectrales

Le coefficient de pentes spectrales est un paramètre supplémentaire dans la description des échantillons de MOD. Il est généralement calculé à partir d'un modèle d'ajustement exponentiel (Eq 1-3) ou bien linéaire (Eq 1-4) au spectre d'absorption. Les coefficients de pentes spectrales les plus utilisés sont le $S_{275-295}$ et le $S_{350-400}$, qui correspondent respectivement aux gammes de longueurs d'ondes 275-295 nm et 350-400 nm. Le rapport entre les deux coefficients de pentes spectrales ($S_{275-295}$ et $S_{350-400}$), est noté SR et constitue un troisième coefficient.

$$\alpha(\lambda) = a_{\lambda_{\text{ref}}} e^{-s(\lambda - \lambda_{\text{ref}})} \quad \text{Eq. 1-3}$$

$$\text{Ln}(\alpha\lambda) = S \text{Ln}(\lambda) + K \quad \text{Eq. 1-4}$$

Où a_λ est le coefficient d'absorption à la longueur λ , λ_{ref} est la longueur d'onde référence de l'échantillon (généralement 250 nm), S correspond à la pente spectrale et K est un paramètre qui prend en compte tout décalage de la ligne de base.

En 2008, Helms et al. ont mis en évidence que le $S_{275-295}$ et le SR présentait tous deux une corrélation négative avec le poids moléculaire. Par la suite, plusieurs études ont montré que les coefficients de pente spectrale pouvaient constituer de bons outils pour caractériser la composition, la réactivité et les sources de MOD dans les systèmes aquatiques. Des valeurs élevées de coefficient ont ainsi été associées à une MOD de faible poids moléculaire plutôt labile et facilement dégradable par l'activité des microorganismes, contrairement à de faibles valeurs qui indiquent une MOD de poids moléculaire élevé réfractaire et peu dégradable. Cela a permis entre autres de différencier une MOD d'origine terrestre d'une MOD produite à l'échelle des cours d'eau (Galgani et al., 2011; Loiselle et al., 2012; Li and Hur, 2017).

1.3.3.2 Spectroscopie de fluorescence

Certains groupements fonctionnels (*e.g.* groupements aromatiques avec doubles liaisons conjuguées) de la MOD émettent de l'énergie sous forme de fluorescence. Ce phénomène permet à une molécule de passer d'un état instable excité à l'état fondamental S_0 (Figure 1-11).

L'intensité de fluorescence émise lors de ce processus peut être quantifiée à l'aide d'un spectrofluorimètre. Cet appareil permet d'effectuer un balayage d'une gamme de longueurs d'onde d'excitation et d'enregistrer les spectres d'émission correspondants. Cela fournit une carte tridimensionnelle de l'intensité de la fluorescence en fonction des longueurs d'onde d'excitation et d'émission (matrice excitation-émission, EEM). L'intensité de fluorescence mesurée dépend de plusieurs facteurs, comme l'absorbance et le rendement quantique de la molécule cible (Eq. 1-5)

$$I_f = 2,3 \times \phi_f \times I_0 \times \epsilon \times l \times c \quad \text{Eq. 1-5}$$

Où I_f est l'intensité de fluorescence, ϕ_f le rendement quantique de fluorescence, I_0 l'intensité de radiation excitatrice, ϵ le coefficient d'absorption molaire, l la longueur de la cuve et c la concentration molaire.

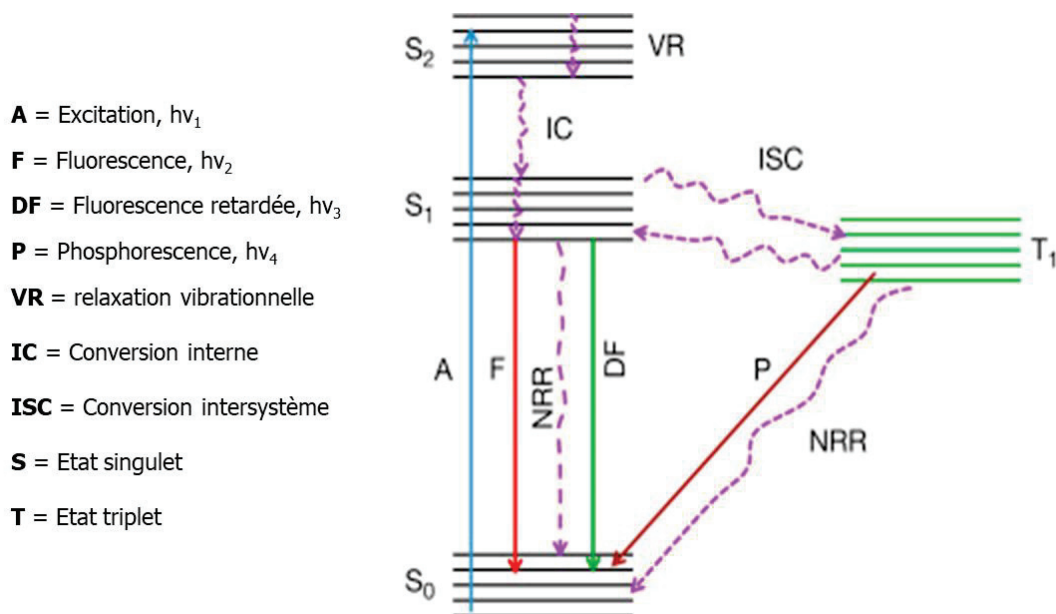


Figure 1-11. Diagramme de Jablonski, adapté à partir de Savoca et al., 2016

La mesure de fluorescence peut être acquise avec une méthode de i) fluorescence d'émission simple, c'est-à-dire se basant sur un couple de longueurs d'onde d'excitation-émission fixées, ii) fluorescence synchrone, en utilisant un écart fixe entre longueurs d'onde d'excitation et d'émission et iii) fluorescence 3D, qui correspond à un balayage simultané sur un domaine de longueurs d'onde d'excitation et d'émission (EEM) (Goffin, 2017). La fluorescence 3D, ou EEM de fluorescence, est très souvent utilisée dans la littérature pour caractériser les échantillons de MOD. L'application à large échelle de cette technique a permis la mise en place de bases de données sur lesquelles sont renseignés les spectres d'excitation-émission de fluorescence de plusieurs milliers d'études (*e.g.* OpenFluor; Murphy et al., 2014).

Les matrices de fluorescence de la matière organique dissoute sont souvent séparées en composés fluorophores dits type-humique (pics "A" et "C"), type-tryptophane (pic "T") et type-tyrosine (pic "B") (Figure 1-12). Cette nomenclature a été introduite pour la première fois par Coble et al., en 1996 puis par Parlanti et al., en 2000. Chacun de ces fluorophores présente un maximum d'intensité de fluorescence visible sous la forme d'un pic. Ces fluorophores constituent des repères dans le traitement des données, car ils permettent de calculer des ratios d'intensités de fluorescence.

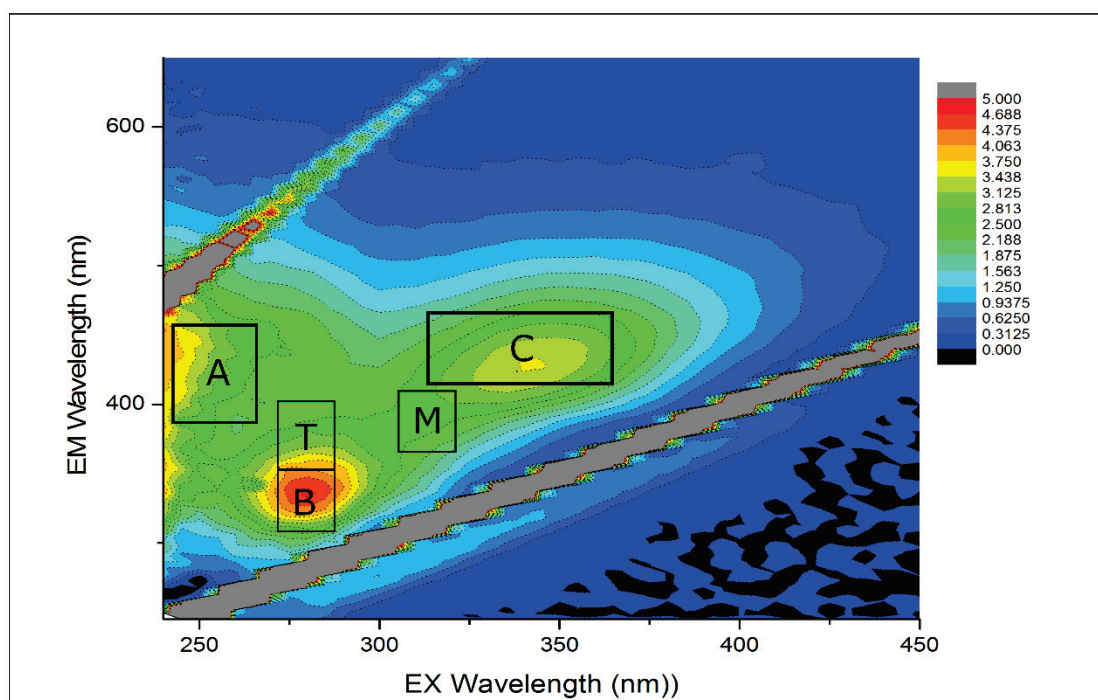


Figure 1-12. Exemple d'une matrice d'excitation émission de fluorescence avec la localisation des principaux fluorophores de la MOD selon Coble et al., 1996

Tableau 1-1. Principaux fluorophores de la MOD

Pic Coble (Parlanti)	Excitation (nm)	Emission (nm)	Type de composé
A (α')	260	380-460	Type-humique (matériel récent)
B (γ)	275	310	Type-protéique (Tyrosine)
C (α)	350	420-480	Type-humique
M (β)	312	380-420	Type-humique (marine)
T (δ)	275	340	Type-protéique (Tryptophane)

i. Ratios d'intensités de fluorescence

Plusieurs rapports d'intensités de fluorescence ont été introduits pour caractériser la matière organique dissoute. On peut par exemple citer :

- L'indicateur HIX, qui correspond à la somme des intensités d'émission entre les longueurs d'onde 435 à 480 nm divisée par la somme des intensités entre 300 et 345 nm, correspondant à la longueur d'onde d'excitation de 254 nm. Il indique le degré d'humification, et de ce fait le degré de décomposition de la MOD (Zsolnay et al., 1999).
- Le ratio β/α , où β correspond au ratio entre le maximum d'excitation de 310 à 320 nm et le maximum d'émission de 380 à 420 nm, et α correspond au ratio entre le maximum d'excitation de 330 à 350 nm et le maximum d'émission de 420 à 480 nm. Le ratio β/α reflète la matière organique récemment produite par l'activité des microorganismes (Huguet et al., 2009).
- L'indice de fluorescence FI, qui correspond au ratio d'intensité d'émission de la longueur d'onde 450 à 500 nm pour une longueur d'onde d'excitation de 370 nm. Cet indice est utilisé pour différencier les sources de MOD. En effet, une valeur de FI élevée indique une MOD autochtone, produite par les microorganismes, contrairement à une valeur de FI faible qui indique une origine terrestre et allochtone de la MOD (Mcknight et al., 2001). Heo et al. (2016) proposent également l'indicateur YFI, qui améliore l'interprétation de l'indicateur FI dans des matrices chargées en MOD (e.g. eaux de STEU). Celui-ci correspond au ratio des intensités de fluorescence moyennes corrigées dans l'intervalle 350-400 sur 400-500 pour la longueur d'onde d'excitation 280 nm.

ii. Modèle de décomposition PARAFAC

En plus du calcul des ratios d'intensités de fluorescence, les matrices d'excitation et d'émission peuvent être traitées mathématiquement par différents algorithmes de décomposition. Ces approches statistiques permettent d'identifier des fluorophores spécifiques dans des matrices complexes constituées de plusieurs fluorophores. La méthode de décomposition la plus utilisée à ce jour est la *PARAllel FACtor analysis* (PARAFAC). Le principe de la méthode repose sur l'utilisation d'un jeu de données sur lequel sera construit un modèle trinéaire le plus proche possible du jeu des données expérimentales (Luciani, 2008). Pour cela, PARAFAC utilise

un algorithme d'alternance des moindres carrés en minimisant la somme des résidus associés aux modèles 3D (Eq. 1-6) (Bro and Kiers, 2003).

$$x_{ijk} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} + \varepsilon_{ijk} \quad \text{Eq. 1-6}$$

où x_{ijk} représente l'intensité de fluorescence de l'échantillon i pour les longueurs d'ondes d'excitation et d'émission (j,k) . N le nombre de fluorophores présents dans le jeu de données. a_{in} représente le facteur de fluorescence du fluorophore n dans l'échantillon i . $b_{jn}c_{kn}$ représentent les valeurs des spectres d'émission et d'absorption du fluorophore n aux longueurs d'ondes j et k , et ε_{ijk} représente l'écart résiduel non pris en compte par le modèle.

La décomposition PARAFAC appliquée à l'étude de la matière organique dissoute a été popularisée dans les années 2000 (Stedmon et al., 2003). Depuis, cette technique a largement été utilisée pour étudier la MOD dans diverses matrices environnementales, y compris les eaux de surface, les eaux souterraines, les eaux résiduaires et le sol (*e.g.* Murphy et al., 2011; Guigue et al., 2014; Guéguen et al., 2015; Schittich et al., 2018). Elle a également été utilisée pour étudier les sources de MOD et son devenir dans les eaux de rivière et dans le milieu marin (Lambert et al., 2017; Lee et al., 2020). Cela a permis d'identifier certains fluorophores dits universels non spécifiques (*e.g.* composantes type-humiques, type-fulviques), ainsi que des fluorophores spécifiques à des types de sources (*e.g.* pesticides, produits pharmaceutiques). La décomposition PARAFAC a également été utilisée pour étudier les effets des facteurs environnementaux, tels que la température et le pH, sur les caractéristiques de la MOD (Lu et al., 2021; Groeneveld et al., 2022).

L'utilisation de PARAFAC dans le cadre d'études sur des matrices environnementales requiert des étapes de prétraitement, telles que la correction des phénomènes de diffusion de la lumière. En effet, lors de l'interaction des photons avec la matière, un phénomène de collision entre le flux de photons et des molécules de taille similaire ou proche intervient et induit une diffusion aléatoire de la lumière. Deux types de diffusions peuvent se produire : la diffusion élastique de Rayleigh et la diffusion inélastique de Raman. Si la lumière est diffusée sans variation d'énergie (pas de changement de fréquence du photon lorsqu'il est dévié par la molécule), il s'agit alors du phénomène de diffusion Rayleigh. Dans le cas où l'énergie du photon subit une perturbation (gain ou perte d'énergie) suite à la collision, il s'agit de la

diffusion Raman. Les deux diffusions sont visibles sur une matrice EEM de fluorescence sous la forme de deux bandes qui rendent impossible l'application de la décomposition PARAFAC en leur présence. Il est donc primordial de les supprimer (Zepp et al., 2004; Nouhi, 2017).

La technique d'EEM de fluorescence s'est révélée être un outil efficace dans la caractérisation de la MOD, tant pour sa composition, à travers l'étude des matrices excitation-émission (Coble, 1996; Parlanti et al., 2000), que pour les propriétés de complexation à travers le phénomène de *quenching* de fluorescence (Elliott et al., 2006; Mounier et al., 2011). Cette technique a également été utilisée pour suivre la dynamique de la MOD en réponse aux changements environnementaux, dans les processus de transfert et de transformation de la MOD, ainsi que dans les procédés de traitement des eaux usées (Stedmon et al., 2003; Carstea et al., 2016; Ba-Haddou et al., 2022). De plus, tout comme la spectroscopie UV-Visible, la spectroscopie de fluorescence est une technique analytique non-destructive, rapide et présentant une grande sensibilité de mesure.

1.3.3.3 Spectroscopie infrarouge à transformation de Fourier FTIR

Contrairement aux techniques UV-Vis et de fluorescence, la spectroscopie infrarouge offre des informations plus spécifiques sur les liaisons chimiques, l'environnement moléculaire et la structure moléculaire de la MOD (Chen and Yu, 2021). La spectroscopie FTIR (*Fourier transform infrared spectroscopy*) est une technique d'analyse qui permet de mesurer l'interaction du rayonnement infrarouge (IR) avec un échantillon. Dans la spectroscopie FTIR, un faisceau de lumière infrarouge traverse un échantillon, et la quantité de lumière absorbée ou transmise est mesurée à l'aide d'un interféromètre, en fonction de la longueur d'onde. Le spectre résultant fournit des informations sur le nombre de vibrations des molécules de l'échantillon, qui peuvent être utilisées pour identifier la présence de groupes chimiques particuliers et déterminer la composition de l'échantillon.

Dans la littérature, la spectroscopie FTIR est une technique largement utilisée pour caractériser la matière organique d'échantillons liquides et solides. Plusieurs études ont ainsi utilisé cette technique pour étudier les propriétés de la MOD dans divers contextes. Par exemple, une étude de comparaison des propriétés de la MOD dans différents types d'écosystèmes aquatiques (*e.g.* rivières, lacs et estuaires) a mis en évidence une composition de MOD estuarienne contenant plus de groupes fonctionnels ester et acétate hautement

aliphatique que la MOD fluviale (Abdulla et al., 2010). D'autres travaux ont employé la spectroscopie FTIR pour étudier la composition de substances humiques extraites de sols impactés par différents usages (terres cultivées, prairies et forêts; Thai et al., 2021). Les résultats ont montré que la MOD des sols forestiers présentait une concentration plus élevée de groupes fonctionnels aromatiques et carboxyliques, tandis que la MOD des sols cultivés présentait une concentration plus élevée en groupes fonctionnels aliphatiques et en chaînes de polysaccharides. D'autres applications de la FTIR (e.g. procédés de traitement des eaux usées et eaux potables, études de biodégradation et interactions entre la MOD et les métaux traces) ont permis d'apporter des avancées majeures dans la compréhension des propriétés physico-chimiques de la MOD (Li et al., 2014; Chen et al., 2015; Rodríguez et al., 2016). Cependant, des contraintes liées à la préparation des échantillons, à la sensibilité relativement faible de la méthode ainsi qu'à l'interférence majeure de la matrice eau dans le domaine de l'infrarouge ont conduit à combiner cette technique avec d'autres outils de spectroscopie tels que la NMR et la fluorescence (Derrien et al., 2019).

1.3.4 Résonance magnétique nucléaire (NMR)

Le principe de la RMN repose sur la détection des ondes électromagnétiques générées lorsque le noyau d'un atome passe d'un état énergétique à un autre. Cette détection se fait en appliquant d'un champ magnétique externe : plus le champ magnétique est intense, plus la mesure RMN sera précise. Parmi les techniques RMN existantes (Figure 1-13), celles du ^{13}C et ^1H sont les plus fréquemment utilisées pour l'analyse structurale des composés de la matière organique.

+ Analyse structurale des molécules par rapport au molécules d'hydrogène	+ Structure carboxylique de la MO, carbone aromatique, aliphatique non substitué, technique puissante avec la RMN ^1H	+ Identification des liaisons carbone-azote, donne une vision plus large de la structure de la MOD	+ Technique multidimensionnelle permet la séparation du spectre en 2 dimensions, possibilité de voir les pics croisés.
RMN ^1H	RMN ^{13}C	RMN ^{14}N	RMN ^2D
- Interprétation des spectres limitée	- Sensibilité pour le carbone inférieure à la RMN ^1H , nécessite l'isolation des échantillons dans un état solide	- sensibilité faible, détection en faible pourcentage du ^{15}N	- Possibilité de ne pas détecter les grosses molécules

Figure 1-13. Techniques RMN utilisées pour caractériser la MOD.

L'essor de la technique à partir des années 1990 a par ailleurs permis de renforcer les connaissances sur les phénomènes de dégradation de la MOD et sa composition en fonction des types de sols (Hedges and Oades, 1997; Kalbitz et al., 2003; Weishaar et al., 2003). Dans les cours d'eau, l'étude de la composition de la MOD par spectroscopie RMN a permis d'identifier certains types de composés tels que les composés allyliques riches en carboxyles (McCaul et al., 2011), des polymères sous forme de peptidoglycanes et des lipoprotéines principalement issues des microorganismes (Rosario-Ortiz, 2014). Les spectres ^{13}C , et ^1H ont également montré que la MOD issue du milieu marin est principalement aliphatique, d'origine autochtone, et la contribution des groupements fonctionnels aromatiques n'est pas significative au niveau de sa structure totale (Opsahl and Benner, 1997).

La spectroscopie RMN à haute résolution permet d'identifier des composés chimiques individuels dans la MOD, ce qui donne un aperçu des sources spécifiques de la MOD dans un environnement donné. Cette technique peut également fournir des informations sur la structure chimique et la réactivité des molécules de MOD, ce qui peut aider à élucider les processus régissant le devenir et le transport des MOD dans les environnements aquatiques.

La technique RMN a mis en évidence la présence de structures de lignine dans la matière organique du sol (Simpson, 2002), et a également permis d'étudier les interactions entre la fraction minérale de l'environnement et les composants aliphatiques et aromatiques de la MOD (Mitchell et al., 2018). Bien qu'il existe un potentiel d'amélioration pour l'avenir, l'application de la RMN reste cependant limitée, principalement en raison des difficultés de préparation des échantillons (étapes d'extraction fastidieuses, nécessité d'utiliser des solvants appropriés) et de la résolution relativement faible des signaux ne permettant pas de caractériser le mélange complexe de composés moléculaires constituant la MOD (Mitschke et al., 2022).

1.3.5 Chromatographie d'exclusion stérique (SEC)

La chromatographie d'exclusion stérique, également appelée chromatographie par perméation de gel, est une technique séparative basée sur la taille ou le volume hydrodynamique des molécules. Contrairement aux autres techniques de chromatographie liquide dont les interactions avec la phase stationnaire sont de type adsorption, partition ou échange d'ions, la SEC repose sur l'utilisation d'une colonne garnie de particules poreuses de

différentes tailles de pores. Ainsi, lors du passage de l'éluant à travers la colonne, les molécules organiques de l'échantillon qui ont une taille trop importante pour pénétrer dans les pores éluent en premier, et à mesure que la taille moléculaire de l'échantillon diminue par rapport à la taille moyenne des pores de la colonne, les molécules pénètrent de plus en plus dans les pores et, par conséquent, éluent plus tardivement (Figure 1-14).

Différents détecteurs peuvent être couplés à la SEC, comme les détecteurs de carbone et d'azote OCD et OND, ou plus récemment des détecteurs UV et fluorescence. La caractérisation de la matière organique dissoute par chromatographie d'exclusion stérique a débuté dans les années 1960 (Gjessing, 1965), mais ce n'est qu'à partir des années 2000, lorsqu'elle a été développée en HPSEC (exclusion stérique couplée à la chromatographie liquide haute performance), que le nombre de publications a fortement augmenté. Dans un premier temps, il a été question de déterminer la distribution des masses moléculaires de la matière organique naturelle (Her et al., 2003; Lanxiu et al., 2004). Dans ce contexte, trois gammes de masses moléculaires ont été identifiées, la première concernant des composés de type protéines/polysaccharides ayant une masse moléculaire élevée (>10 000 Da), la deuxième des composés de type acides humiques/fulviques de masse moléculaire moyenne (1000-5000 Da) et la dernière gamme, inférieure à 1000 Da, étant principalement constituée de composés de type acides/protéines.

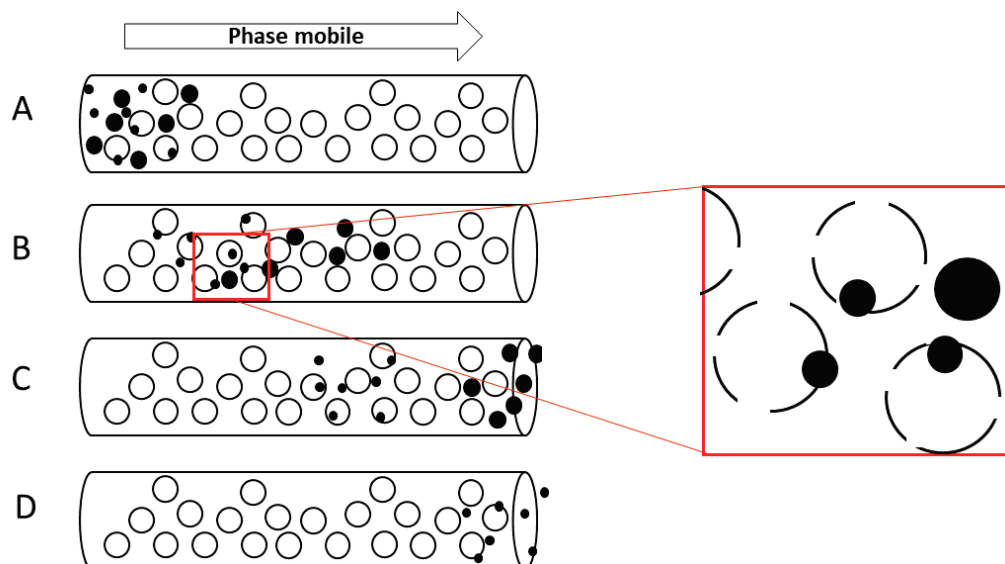


Figure 1-14. Représentation schématique de la séparation par chromatographie d'exclusion stérique. Les molécules de petite taille diffusent à travers les particules poreuses alors que les molécules de taille plus importante éluent sans rétention. Adapté de (Mori and Barth, 1999)

Cette caractérisation a permis de mieux comprendre la composition complexe de la matière organique, et ce dans un contexte environnemental mais également sanitaire par la suite, notamment dans le domaine du traitement des eaux potables (Chow et al., 2009; Fabris et al., 2008). En effet, l'étude de la composition de la matière organique par HPSEC, à travers les processus de traitement des eaux, a permis d'identifier les composés responsables du changement de couleur, d'odeur et de goût dans les eaux potables, mais également des sous-produits de désinfection de la MOD, et ainsi d'adapter les traitements (Brezinski and Gorczyca, 2019). La détection plus récente par UV-DAD (barrettes de diode) et fluorescence permet de cibler spécifiquement les composés chromophoriques et fluorescents de la matière organique (Chon et al., 2017).

1.3.6 Techniques de spectrométrie de masse

La spectrométrie de masse est une technique d'analyse souvent utilisée pour déterminer la composition élémentaire ainsi que d'autres aspects de la structure moléculaire d'un composé préalablement ionisé. Un spectromètre de masse est constitué principalement d'une source d'ions, d'un analyseur masse sur charges (m/z) ainsi qu'un détecteur. Les ions étant des particules chargées, leur position dans l'espace peut donc être manipulée avec l'utilisation de champs électriques et magnétiques. En effet, un spectromètre de masse va détecter les ions présents dans un échantillon selon leurs propriétés de masse et nombre de charges (m/z). Cette entité mesurée en spectrométrie de masse (m/z) des ions est différente de la masse de ces derniers.

1.3.6.1 Pyrolyse-chromatographie en phase gazeuse couplée à la spectrométrie de masse (Py-GC-MS)

La Py-GC-MS consiste à chauffer la matière organique dissoute à haute température (> 500°C) en l'absence d'oxygène, ce qui décompose les molécules organiques complexes (e.g. protéines, lignine, glucides) en composés volatils plus simples qui peuvent être analysés par chromatographie en phase gazeuse et détectés par spectrométrie de masse (Song and Peng, 2010). Le principal avantage de la Py-GC-MS est qu'elle permet l'identification et la quantification d'une large gamme de composés organiques dans la MOD, et donc de fournir des informations précieuses sur la composition et la structure moléculaires de la MOD (Matilainen et al., 2011).

La Py-GC-MS a été utilisée pour étudier les caractéristiques de la MOD dans différents milieux naturels et anthropiques (*e.g.* lacs, rivières, effluents de STEU) (Christy, 1999; Greenwood et al., 2012). La Py-GC-MS a également été utilisée pour étudier les sources et les transformations de la MOD dans les systèmes aquatiques, par exemple dans le cadre de l'identification de biomarqueurs spécifiques de MOD caractéristiques de différentes sources, telles que le couvert végétal, les algues et les bactéries (Greenwood et al., 2012; Kaal et al., 2017). La Py-GC-MS est une méthode efficace et utile pour la caractérisation moléculaire rapide d'échantillons de MOD (Zhang et al., 2021). Cependant, la technique souffre de certaines limitations lors de l'analyse des groupes fonctionnels polaires (-COOH, -OH) typiquement présents dans les matières organiques naturelles. L'interprétation des données pyrolytiques peut ainsi parfois être incorrecte, car dans de nombreux cas, les unités naturellement présentes dans les matériaux organiques naturels (-COOH, -OH) sont des groupes fonctionnels polaires (Song and Peng, 2010).

1.3.6.2 Spectrométrie de masse à résonance cyclotronique ionique (FTICR-MS)

La FTICR-MS est une technique qui fait appel à un champ magnétique qui piège les ions dans un mouvement similaire à celui d'un cyclotron (Marshall et al., 1998). Une tension d'excitation fait ensuite osciller les ions piégés à une certaine fréquence. Le rapport masse/charge (m/z) peut alors être facilement calculé en mesurant la fréquence des ions. Le signal obtenu à partir de la résonance cyclotronique des ions correspond à un domaine de spectres très complexes en raison de la superposition des signaux. Pour retrouver l'information initiale, l'algorithme de la transformation de Fourier est appliqué, ce qui génère un spectre de fréquence facilement convertible en spectre de masse (Buchanan et al., 1987). La transformation de Fourier présente un autre avantage, celui de réduire considérablement l'acquisition des données car les masses des différents ions sont détectées simultanément.

La spectrométrie de masse FTICR est un outil utilisé pour analyser des mélanges complexes comme la MOD. Cette technique a ainsi été utilisée pour caractériser les molécules de la MOD dans le milieu marin et dans les écosystèmes aquatiques terrestres (Minor et al., 2012; Seidel et al., 2022). La spectrométrie de masse par résonance cyclotronique ionique à transformée de Fourier (FT-ICR-MS) a par exemple permis d'identifier plusieurs milliers de formules chimiques prédites dans des échantillons de MOD marins et d'eau douce. La précision de

masse (<0,3 ppm) et la très haute résolution de masse (>500 000 FWHM) de la FT-ICR-MS couplée à l'ionisation par électro spray (ESI) permettent d'extraire la composition élémentaire des ions. Il est ainsi possible de calculer les ratios des éléments H/C, O/C et N/C et d'observer

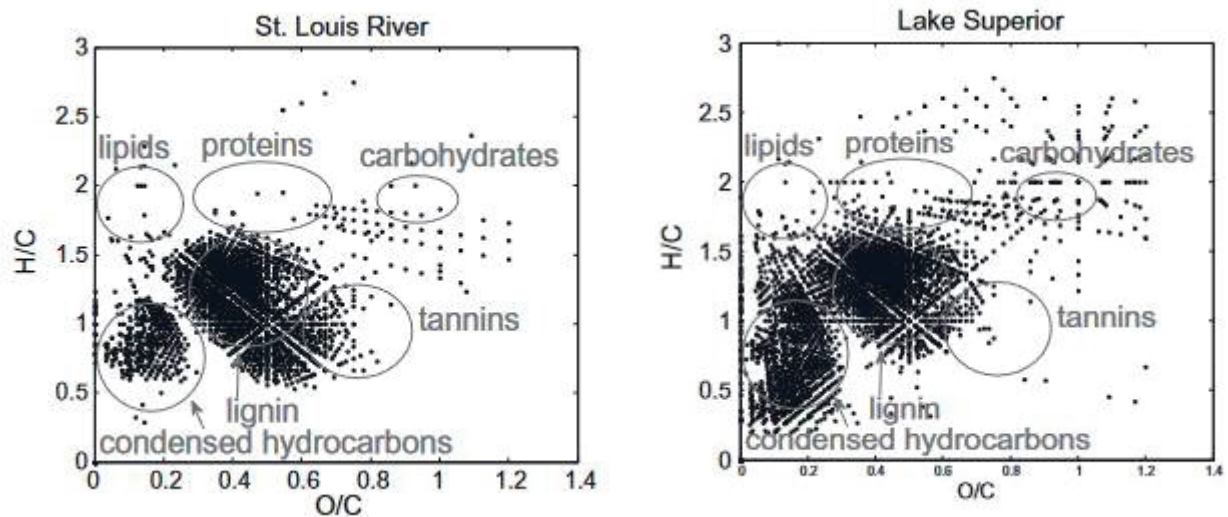


Figure 1-15. Diagramme de van Krevelen réalisés à partir des compositions élémentaires de deux échantillons caractéristiques d'une rivière et d'un lac. D'après Minor et al., 2012

la répartition de ces ratios pour l'ensemble des ions détectés au travers des diagrammes de van Krevelen (Figure 1-15). Ces diagrammes permettent alors de différencier les échantillons de MOD en fonction des structures chimiques (par exemple lipide, protéine, hydrate de carbone, lignine ou tanin) (Kamjunke et al., 2017; Zhu et al., 2019). Cela reste cependant sous la forme de graphiques, qui sont donc difficilement intégrables dans des analyses multivariées.

1.3.6.3 Chromatographie liquide couplée à la spectrométrie de masse haute résolution LC-HRMS

D'autres alternatives pour étudier les caractéristiques moléculaires des MOD peuvent être trouvées parmi les techniques HRMS couplées à la chromatographie liquide (LC). La combinaison de ces deux techniques permet d'identifier et de quantifier les composés (via leurs caractéristiques moléculaires) sur la base de leur rapport masse/charge (m/z), du temps de rétention chromatographique (RT) et de l'intensité. Cela a ouvert de nouvelles perspectives de recherche sur la composition complexe des MOD, car l'inclusion de la partie chromatographique permet une meilleure séparation des molécules organiques très hétérogènes qui composent les MOD (Brock et al., 2020).

Différents spectromètres de masse haute résolution peuvent être couplés à la LC, comme le spectromètre de masse à trappe orbitale (Orbitrap). Le principe de l'Orbitrap repose sur l'utilisation d'un champ électrostatique stable pour piéger les ions sur une orbite circulaire (Figure 1-16). Lorsque les ions piégés tournent autour du fuseau central, ils interagissent avec

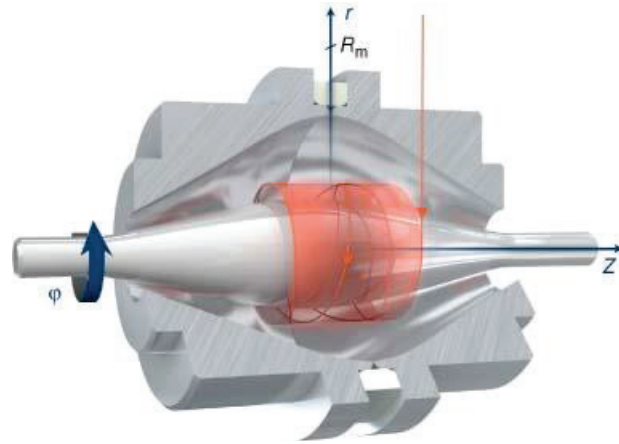


Figure 1-16. Schéma de l'analyseur de masse Orbitrap montrant une trajectoire en spirale stable d'un ion. D'après Hecht et al., 2019

le champ électrostatique environnant, ce qui les fait osciller à une fréquence spécifique. La fréquence d'oscillation des ions est ensuite mesurée et le rapport m/z des ions est déterminé sur la base de la fréquence d'oscillation (Hecht et al., 2019). Ce principe permet aux spectromètres de masse Orbitrap d'obtenir des mesures de masse précises à des résolutions pouvant atteindre les 480 000 FWHM pour les Orbitrap à ultra-haut champ.

Le spectromètre à temps de vol (QTOF/HRMS) est également de plus en plus utilisé dans les études de métabolomique et environnementales (Aguilar-Alarcón et al., 2020; Tenorio et al., 2022; Wang et al., 2021). Le principe des analyseurs TOF repose sur la mesure du temps requis pour qu'un ion se déplace jusqu'au détecteur localisé à 1-2 m de la source (Figure 1-17). Les ions reçoivent une énergie cinétique similaire au moment de l'accélération instantanée, mais la différence des m/z des ions va influencer la vitesse de traverser l'espace libre (Guilhaus, 1995). Contrairement aux autres analyseurs m/z , le principe du TOF ne présente en théorie pas de limite supérieure dans la valeur m/z , les ions avec les valeurs de m/z plus faibles atteignent le détecteur avant les ions ayant des m/z plus élevés car ces derniers ont une vitesse de traversée plus réduite (Mamyryn, 2001).

Les développements récents de l'analyse par LC-HRMS ont ouvert de nouvelles perspectives de recherche (Paszkiwicz et al., 2022). Qu'il s'agisse de rechercher spécifiquement des

composés organiques, de identifier des composés suspects ou plus globalement d'établir des empreintes moléculaires, l'analyse par HRMS a permis des avancées majeures dans les domaines de la métabolomique, de la surveillance des aliments, des études environnementales et de la santé humaine (Blum et al., 2017; Fraser et al., 2014; Wang et al., 2021).

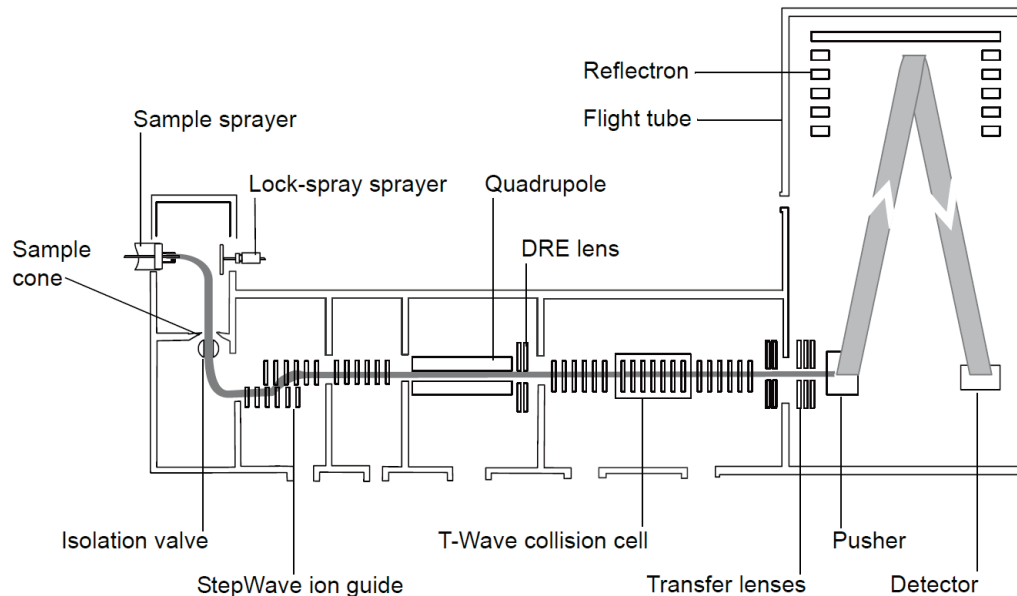


Figure 1-17. Schéma d'un instrument QTOF (waters)

Trois approches principales sont utilisées en HRMS :

- (i) L'analyse ciblée, qui permet la détection précise de composés chimiques sur la base d'informations telles que le temps de rétention (RT), le spectre MS/MS et l'ion de fragmentation des standards analytiques. Des instruments tels que le triple quadripôle et le piège à ions quadripolaire sont couramment utilisés dans l'analyse HRMS ciblée car ils offrent une sensibilité et une sélectivité élevées (Krauss et al., 2010).
- (ii) L'analyse suspectée, qui permet d'identifier des composés tels que les nouveaux composés émergents et les produits de transformation sans se référer à un standard analytique (Nihemaiti et al., 2022; Rocco et al., 2022). Le principe de l'approche suspectée repose sur l'utilisation de bases de données contenant des informations sur les temps de rétention, les masses, les isotopes et les profils de fragments de centaines de molécules.
- (iii) Enfin, l'analyse non-ciblée, qui est une approche plus globale pour le criblage de composés inconnus (Hollender et al., 2017). L'approche non ciblée (NTS) est basée sur

l'identification de *features* moléculaires, qui correspondent à des signaux caractérisés par un temps de rétention ainsi qu'un rapport masse sur charge (m/z) (Schulze et al., 2020).

En analyse LC-HRMS non-ciblée (NTS), la génération de la liste de *features* repose sur l'utilisation d'outils de retraitement de données disponibles en libre accès (Smith et al., 2006; Pluskal et al., 2010) ou bien fournis par les constructeurs (*e.g.* UNIFI Waters, Masshunter Agilent, Compound Discoverer Thermo). L'approche NTS pour la caractérisation de la MOD a commencé à émerger au début des années 2000, avec les progrès de la technologie de la spectrométrie de masse à haute résolution (HRMS). L'approche NTS est actuellement largement utilisée pour la caractérisation de la MOD dans diverses matrices environnementales, telles que les eaux naturelles, les eaux usées et le sol. Elle est utilisée pour identifier et quantifier une large gamme de composés, y compris ceux qui ne sont pas ciblés par les méthodes analytiques traditionnelles. Malgré son utilité, l'approche NTS présente également un certain nombre de limitations. L'un des principaux défis qu'elle pose est le traitement et l'interprétation de la grande quantité de données générées. En effet, les spectres HRMS générés par les instruments peuvent donner lieu à des ensembles de données volumineux, dont le traitement nécessite des logiciels spécialisés. Les étapes de traitement des données (*workflow*) nécessitent un contrôle de qualité minutieux afin de réduire l'apparition de faux positifs et de faux négatifs et de garantir la reproductibilité des données. En outre, la multitude de logiciels de retraitement peut entraîner des variations dans l'analyse des données, ce qui complique la comparaison des résultats obtenus dans le cadre de différentes études (Myers et al., 2017). Les différents outils logiciels peuvent par ailleurs appliquer des algorithmes différents pour gérer le traitement des données, la sélection et l'intégration des pics, ce qui peut entraîner des résultats variables. Il est donc essentiel de normaliser les processus de traitement des données pour garantir la comparabilité des résultats obtenus dans les différentes études.

1.3.7 Caractérisation de la MOD à travers une approche intégrée

Les précédentes sections montrent que les verrous liés aux techniques de caractérisation de la MOD résident dans l'information partielle apportée par chaque technique. Les différentes techniques précédemment passées en revue fournissent en effet des informations sur des

aspects spécifiques de la MOD comme les propriétés de fluorescence, la distribution en poids moléculaire ou encore la diversité des composés chimiques (Figure 1-18). L'intégration de techniques multiples est donc nécessaire pour établir une compréhension plus complète des MOD (Mangal et al., 2020).

De précédentes études se sont intéressés à combiner les techniques UV-visible et EEM de fluorescence avec la HRMS, ou plus récemment avec la HPSEC/UV-fluorescence, pour étudier la composition de la MOD, son origine et ses mécanismes de transformations (He et al., 2019; Mangal et al., 2020; Hansen et al., 2016). Cependant, la complexité et l'hétérogénéité de la MOD peuvent rendre difficile l'intégration des données. De plus, l'utilisation de différentes techniques nécessite souvent l'utilisation de méthodes de préconcentration et d'extraction de la MOD (e.g. extraction en phase solide SPE, ultrafiltration), ce qui peut introduire des biais ou altérer ses propriétés (Liao et al., 2021; Butturini et al., 2022). De plus, la caractérisation de la MOD par des techniques comme la spectrométrie de masse fait face à de nombreux défis, comme la concentration relativement faible des molécules dans les échantillons naturels ainsi que l'absence de standards analytiques, qui rendent difficile l'établissement d'une approche quantitative et l'assignation des *features* détectés à des molécules connues (Duarte, 2020).

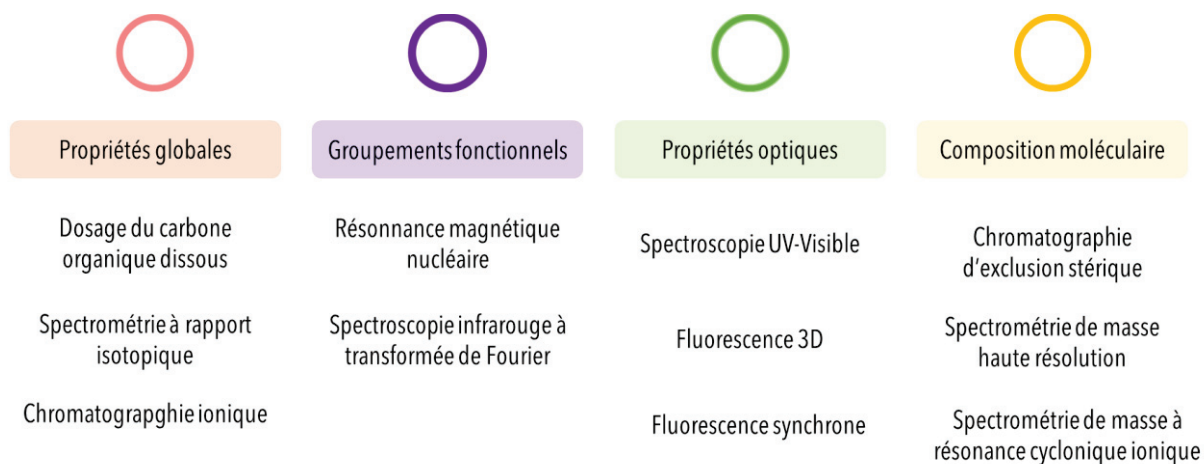


Figure 1-18. Principales techniques utilisées pour caractériser les différentes fractions de la MOD

Dans une optique d'étude des sources de MOD, construire des empreintes de MOD à partir des données issues d'une ou de plusieurs techniques est toutefois possible à l'aide d'approches multivariées (e.g. analyses en composantes principales, qui intègrent des données issues des toutes les techniques) (Tang et al., 2014; Liao et al., 2021). Pour pouvoir différencier plus finement les sources de MOD, il semble essentiel d'utiliser des indicateurs

variés qui reflètent les principales caractéristiques physico-chimiques de la MOD (*e.g.* taille, composition, aromaticité). Cependant, pour des techniques telles que la spectrométrie de masse à haute résolution, il n'existe actuellement aucun indicateur connu qui puisse être intégré facilement dans ce type d'analyse de données. De plus, la construction d'empreintes de MOD nécessite également la mise en place de bases de données avec un large éventail d'échantillons, diversifiés en termes de typologie de sources. Ces bases de données permettraient la création d'empreintes complexes de la MOD et faciliteraient le développement d'une typologie de la MOD basée sur son origine, ses liens ou ses effets avec la microbiologie ou les micropolluants.

1.4 Références

- Abdulla, H.A.N., Minor, E.C., Dias, R.F., Hatcher, P.G., 2010. Changes in the compound classes of dissolved organic matter along an estuarine transect: A study using FTIR and ¹³C NMR. *Geochimica et Cosmochimica Acta* 74, 3815–3838. <https://doi.org/10.1016/j.gca.2010.04.006>
- Aguilar-Alarcón, P., Gonzalez, S.V., Simonsen, M.A., Borrero-Santiago, A.R., Sanchís, J., Meriac, A., Kolarevic, J., Asimakopoulos, A.G., Mikkelsen, Ø., 2020. Characterizing changes of dissolved organic matter composition with the use of distinct feeds in recirculating aquaculture systems via high-resolution mass spectrometry. *Science of The Total Environment* 749, 142326. <https://doi.org/10.1016/j.scitotenv.2020.142326>
- Aiken, G.R., 2014. 1.11 - Dissolved Organic Matter in Aquatic Systems, in: Ahuja, S. (Ed.), *Comprehensive Water Quality and Purification*. Elsevier, Waltham, pp. 205–220. <https://doi.org/10.1016/B978-0-12-382182-9.00014-1>
- Aitkenhead-Peterson, J.A., McDowell, W.H., Neff, J.C., 2003. 2 - Sources, Production, and Regulation of Allochthonous Dissolved Organic Matter Inputs to Surface Waters, in: Findlay, S.E.G., Sinsabaugh, R.L. (Eds.), *Aquatic Ecosystems, Aquatic Ecology*. Academic Press, Burlington, pp. 25–70. <https://doi.org/10.1016/B978-012256371-3/50003-2>
- Alkhoury, J., 2022. Caractérisation et quantification des apports urbains de contaminants particuliers dans le Rhône. Rapport de Master 2 Sciences de l'eau, Université de Franche-Comté, INRAE. 63pp.
- Amin, S.A., Parker, M.S., Armbrust, E.V., 2012. Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews* 76, 667–684. <https://doi.org/10.1128/MMBR.00007-12>
- Arnosti, C., 2003. 13 - Microbial Extracellular Enzymes and their Role in Dissolved Organic Matter Cycling, in: Findlay, S.E.G., Sinsabaugh, R.L. (Eds.), *Aquatic Ecosystems, Aquatic Ecology*. Academic Press, Burlington, pp. 315–342. <https://doi.org/10.1016/B978-012256371-3/50014-7>
- Ba-Haddou, H., Hassoun, H., Foudeil, S., EL Bakkali, A., Ait Lyazidi, S., Haddad, M., Masson, M., Coquery, M., Margoum, C., 2022. Combination of 3D Fluorescence/PARAFAC and UV–Vis Absorption for the Characterization of Agricultural Soils from Morocco. *J Fluoresc* 32, 2141–2149. <https://doi.org/10.1007/s10895-022-03011-3>
- Baldock, J.A., Masiello, C.A., Gélinas, Y., Hedges, J.I., 2004. Cycling and composition of organic matter in terrestrial and marine ecosystems. *Marine Chemistry, New Approaches in Marine Organic Biogeochemistry: A Tribute to the Life and Science of John I. Hedges* 92, 39–64. <https://doi.org/10.1016/j.marchem.2004.06.016>
- Bertilsson, S., Jones, J.B., 2003. 1 - Supply of Dissolved Organic Matter to Aquatic Ecosystems: Autochthonous Sources, in: Findlay, S.E.G., Sinsabaugh, R.L. (Eds.), *Aquatic Ecosystems, Aquatic Ecology*. Academic Press, Burlington, pp. 3–24. <https://doi.org/10.1016/B978-012256371-3/50002-0>
- Bianchi, S.R., Miyazawa, M., Oliveira, E.L. de, Pavan, M.A., 2008. Relationship between the mass of organic matter and carbon in soil. *Braz. arch. biol. technol.* 51, 263–269. <https://doi.org/10.1590/S1516-89132008000200005>
- Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. *Science of The Total Environment* 575, 265–275. <https://doi.org/10.1016/j.scitotenv.2016.09.135>
- Boram, K., 2014. Devenir du phosphore dans les filtres plantés de roseaux : étude de sa rétention / libération et des facteurs d'influence.
- Brezinski, K., Gorczyca, B., 2019. An overview of the uses of high performance size exclusion chromatography (HPSEC) in the characterization of natural organic matter (NOM) in potable

- water, and ion-exchange applications. *Chemosphere* 217, 122–139. <https://doi.org/10.1016/j.chemosphere.2018.10.028>
- Brezonik, P.L., Olmanson, L.G., Finlay, J.C., Bauer, M.E., 2015. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. *Remote Sensing of Environment, Special Issue: Remote Sensing of Inland Waters* 157, 199–215. <https://doi.org/10.1016/j.rse.2014.04.033>
- Bro, R., Kiers, H.A.L., 2003. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics* 17, 274–286. <https://doi.org/10.1002/cem.801>
- Brock, O., Helmus, R., Kalbitz, K., Jansen, B., 2020. Non-target screening of leaf litter-derived dissolved organic matter using liquid chromatography coupled to high-resolution mass spectrometry (LC-QTOF-MS). *European Journal of Soil Science* 71, 420–432. <https://doi.org/10.1111/ejss.12894>
- Brown, J.N., Peake, B.M., 2006. Sources of heavy metals and polycyclic aromatic hydrocarbons in urban stormwater runoff. *Science of The Total Environment* 359, 145–155. <https://doi.org/10.1016/j.scitotenv.2005.05.016>
- Butturini, A., Herzsprung, P., Lechtenfeld, O.J., Alcorlo, P., Benaiges-Fernandez, R., Berlanga, M., Boadella, J., Freixinos Campillo, Z., Gomez, R.M., Sanchez-Montoya, M.M., Urmeneta, J., Romaní, A.M., 2022. Origin, accumulation and fate of dissolved organic matter in an extreme hypersaline shallow lake. *Water Research* 221, 118727. <https://doi.org/10.1016/j.watres.2022.118727>
- Carrey, R., Ballesté, E., Blanch, A.R., Lucena, F., Pons, P., López, J.M., Rull, M., Solà, J., Micola, N., Fraile, J., Garrido, T., Munné, A., Soler, A., Otero, N., 2021. Combining multi-isotopic and molecular source tracking methods to identify nitrate pollution sources in surface and groundwater. *Water Research* 188, 116537. <https://doi.org/10.1016/j.watres.2020.116537>
- Carstea, E.M., Bridgeman, J., Baker, A., Reynolds, D.M., 2016. Fluorescence spectroscopy for wastewater monitoring: A review. *Water Research* 95, 205–219. <https://doi.org/10.1016/j.watres.2016.03.021>
- Cartwright, I., Hofmann, H., Gilfedder, B., Smyth, B., 2014. Understanding parafluvial exchange and degassing to better quantify groundwater inflows using ²²²Rn: The King River, southeast Australia. *Chemical Geology* 380, 48–60. <https://doi.org/10.1016/j.chemgeo.2014.04.009>
- Chen, M., Price, R.M., Yamashita, Y., Jaffé, R., 2010. Comparative study of dissolved organic matter from groundwater and surface water in the Florida coastal Everglades using multi-dimensional spectrofluorometry combined with multivariate statistics. *Applied Geochemistry* 25, 872–880. <https://doi.org/10.1016/j.apgeochem.2010.03.005>
- Chen, W., Habibul, N., Liu, X.-Y., Sheng, G.-P., Yu, H.-Q., 2015. FTIR and Synchronous Fluorescence Heterospectral Two-Dimensional Correlation Analyses on the Binding Characteristics of Copper onto Dissolved Organic Matter. *Environ. Sci. Technol.* 49, 2052–2058. <https://doi.org/10.1021/es5049495>
- Chen, W., Yu, H.-Q., 2021. Advances in the characterization and monitoring of natural organic matter using spectroscopic approaches. *Water Research* 190, 116759. <https://doi.org/10.1016/j.watres.2020.116759>
- Chianese, S., Fenti, A., Iovino, P., Musmarra, D., Salvestrini, S., 2020. Sorption of Organic Pollutants by Humic Acids: A Review. *Molecules* 25, 918. <https://doi.org/10.3390/molecules25040918>
- Chon, Kangmin, Chon, Kyongmi, Cho, J., 2017. Characterization of size fractionated dissolved organic matter from river water and wastewater effluent using preparative high performance size exclusion chromatography. *Organic Geochemistry* 103, 105–112. <https://doi.org/10.1016/j.orggeochem.2016.11.003>
- Chow, C.W.K., van Leeuwen, J.A., Fabris, R., Drikas, M., 2009. Optimised coagulation using aluminium sulfate for the removal of dissolved organic carbon. *Desalination* 245, 120–134. <https://doi.org/10.1016/j.desal.2008.06.014>
- Christy, A., 1999. Characterization of natural organic matter by pyrolysis/GC-MS. *Environment International* 25, 181–189. [https://doi.org/10.1016/S0160-4120\(98\)00117-2](https://doi.org/10.1016/S0160-4120(98)00117-2)

- Claret, F., Schäfer, T., Bauer, A., Buckau, G., 2003. Generation of humic and fulvic acid from Callovo-Oxfordian clay under high alkaline conditions. *Science of The Total Environment* 317, 189–200. [https://doi.org/10.1016/S0048-9697\(03\)00337-1](https://doi.org/10.1016/S0048-9697(03)00337-1)
- Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry* 51, 325–346. [https://doi.org/10.1016/0304-4203\(95\)00062-3](https://doi.org/10.1016/0304-4203(95)00062-3)
- Denis, M., Jeanneau, L., Pierson-Wickman, A.-C., Humbert, G., Petitjean, P., Jaffrézic, A., Gruau, G., 2017. A comparative study on the pore-size and filter type effect on the molecular composition of soil and stream dissolved organic matter. *Organic Geochemistry* 110, 36–44. <https://doi.org/10.1016/j.orggeochem.2017.05.002>
- Derrien, M., Brogi, S.R., Gonçalves-Araujo, R., 2019. Characterization of aquatic organic matter: Assessment, perspectives and research priorities. *Water Research* 163, 114908. <https://doi.org/10.1016/j.watres.2019.114908>
- Du, Y., Chen, F., Xiao, K., Song, C., He, H., Zhang, Q., Zhou, Y., Jang, K.-S., Zhang, YiBo, Xing, P., Liu, Z., Zhang, YunLin, Lu, Y., 2021. Water Residence Time and Temperature Drive the Dynamics of Dissolved Organic Matter in Alpine Lakes in the Tibetan Plateau. *Global Biogeochemical Cycles* 35, e2020GB006908. <https://doi.org/10.1029/2020GB006908>
- Duarte, R., 2020. *Multidimensional Analytical Techniques in Environmental Research*. Elsevier. <https://doi.org/10.1016/C2018-0-04842-2>
- Elliott, S., Lead, J.R., Baker, A., 2006. Thermal quenching of fluorescence of freshwater, planktonic bacteria. *Analytica Chimica Acta* 564, 219–225. <https://doi.org/10.1016/j.aca.2006.01.087>
- Evangelou, M.W.H., Ebel, M., Schaeffer, A., 2007. Chelate assisted phytoextraction of heavy metals from soil. Effect, mechanism, toxicity, and fate of chelating agents. *Chemosphere* 68, 989–1003. <https://doi.org/10.1016/j.chemosphere.2007.01.062>
- Fabris, R., Chow, C.W.K., Drikas, M., Eikebrokk, B., 2008. Comparison of NOM character in selected Australian and Norwegian drinking waters. *Water Research* 42, 4188–4196. <https://doi.org/10.1016/j.watres.2008.06.023>
- Fellman, J.B., Petrone, K.C., Grierson, P.F., 2011. Source, biogeochemical cycling, and fluorescence characteristics of dissolved organic matter in an agro-urban estuary. *Limnology and Oceanography* 56, 243–256. <https://doi.org/10.4319/lo.2011.56.1.0243>
- Fraser, K., Lane, G.A., Otter, D.E., Harrison, S.J., Quek, S.-Y., Hemar, Y., Rasmussen, S., 2014. Non-targeted analysis by LC–MS of major metabolite changes during the oolong tea manufacturing in New Zealand. *Food Chemistry* 151, 394–403. <https://doi.org/10.1016/j.foodchem.2013.11.054>
- Galgani, L., Tognazzi, A., Rossi, C., Ricci, M., Galvez, J.A., Dattilo, A.M., Cozar, A., Bracchini, L., Loiselle, S.A., 2011. Assessing the optical changes in dissolved organic matter in humic lakes by spectral slope distributions. *J Photochem Photobiol B* 102, 132–139. <https://doi.org/10.1016/j.jphotobiol.2010.10.001>
- Garnier, J., Cébron, A., Tallec, G., Billen, G., Sebilo, M., Martinez, A., 2006. Nitrogen Behaviour and Nitrous Oxide Emission in the Tidal Seine River Estuary (France) as Influenced by Human Activities in the Upstream Watershed. *Biogeochemistry* 77, 305–326. <https://doi.org/10.1007/s10533-005-0544-4>
- Gasperi, J., Le Roux, J., Deshayes, S., Ayrault, S., Bordier, L., Boudahmane, L., Budzinski, H., Caupos, E., Caubrière, N., Flanagan, K., Guillon, M., Huynh, N., Labadie, P., Meffray, L., Neveu, P., Partibane, C., Paupardin, J., Saad, M., Varnede, L., Gromaire, M.-C., 2022. Micropollutants in Urban Runoff from Traffic Areas: Target and Non-Target Screening on Four Contrasted Sites. *Water* 14, 394. <https://doi.org/10.3390/w14030394>
- Gevao, B., Semple, K.T., Jones, K.C., 2000. Bound pesticide residues in soils: a review. *Environmental Pollution* 108, 3–14. [https://doi.org/10.1016/S0269-7491\(99\)00197-9](https://doi.org/10.1016/S0269-7491(99)00197-9)
- Gjessing, E.T., 1965. Use of “sephadex” gel for the estimation of molecular weight of humic substances in natural water [11]. *Nature* 208, 1091–1092. <https://doi.org/10.1038/2081091a0>

- Goffin, A., 2017. Potentiel d'utilisation de la spectrofluorimétrie 3D pour la caractérisation en ligne de la matière organique dissoute: de la station d'épuration au milieu récepteur.
- Goffin, A., Guérin, S., Rocher, V., Varrault, G., 2018. Towards a better control of the wastewater treatment process: excitation-emission matrix fluorescence spectroscopy of dissolved organic matter as a predictive tool of soluble BOD5 in influents of six Parisian wastewater treatment plants. *Environ Sci Pollut Res* 25, 8765–8776. <https://doi.org/10.1007/s11356-018-1205-1>
- Gorini, D., Choubert, J.-M., le Pimpec, P., Heduit, A., 2011. Concentrations and fate of sugars, proteins and lipids during domestic and agro-industrial aerobic treatment. *Water Science and Technology* 63, 1669–1677. <https://doi.org/10.2166/wst.2011.334>
- Green, H., Weller, D., Johnson, S., Michalenko, E., 2019. Microbial Source-Tracking Reveals Origins of Fecal Contamination in a Recovering Watershed. *Water* 11, 2162. <https://doi.org/10.3390/w11102162>
- Greenwood, P.F., Berwick, L.J., Croué, J.-P., 2012. Molecular characterisation of the dissolved organic matter of wastewater effluents by MSSV pyrolysis GC–MS and search for source markers. *Chemosphere* 87, 504–512. <https://doi.org/10.1016/j.chemosphere.2011.12.051>
- Groeneveld, M., Catalán, N., Einarsdottir, K., G. Bravo, A., N. Kothawala, D., 2022. The influence of pH on dissolved organic matter fluorescence in inland waters. *Analytical Methods* 14, 1351–1360. <https://doi.org/10.1039/D1AY01702K>
- Guéguen, C., Itoh, M., Kikuchi, T., Eert, J., Williams, W.J., 2015. Variability in dissolved organic matter optical properties in surface waters in the Amerasian Basin. *Front. Mar. Sci.* 2. <https://doi.org/10.3389/fmars.2015.00078>
- Guigue, J., Mathieu, O., Lévêque, J., Mounier, S., Laffont, R., Maron, P.-A., Navarro, N., Chateau-Smith, C., Suchet, P., Lucas, Y., 2014. A comparison of extraction procedures for water-extractable organic matter in soils. *European Journal of Soil Science* 65, 520–530. <https://doi.org/10.1111/ejss.12156>
- Guilhaus, M., 1995. Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry* 30, 1519–1532. <https://doi.org/10.1002/jms.1190301102>
- Guo, W., Stedmon, C.A., Han, Y., Wu, F., Yu, X., Hu, M., 2007. The conservative and non-conservative behavior of chromophoric dissolved organic matter in Chinese estuarine waters. *Marine Chemistry, 9th International Estuarine Biogeochemistry Symposium Estuaries and Enclosed Seas under Changing Environmental Conditions* 107, 357–366. <https://doi.org/10.1016/j.marchem.2007.03.006>
- Halim, M., Conte, P., Piccolo, A., 2003. Potential availability of heavy metals to phytoextraction from contaminated soils induced by exogenous humic substances. *Chemosphere* 52, 265–275. [https://doi.org/10.1016/S0045-6535\(03\)00185-1](https://doi.org/10.1016/S0045-6535(03)00185-1)
- Han, L., Yang, Y., Sun, K., Zhang, B., Chen, Y., Fang, L., Xing, B., 2021. Different mechanisms driving the preferential adsorption of dissolved organic matter by goethite and montmorillonite. *Chemical Geology* 585, 120560. <https://doi.org/10.1016/j.chemgeo.2021.120560>
- Hansen, A.M., Kraus, T.E.C., Pellerin, B.A., Fleck, J.A., Downing, B.D., Bergamaschi, B.A., 2016. Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. *Limnology and Oceanography* 61, 1015–1032. <https://doi.org/10.1002/lno.10270>
- Hartnett, H.E., 2018. Dissolved Organic Matter (DOM), in: White, W.M. (Ed.), *Encyclopedia of Geochemistry: A Comprehensive Reference Source on the Chemistry of the Earth*, *Encyclopedia of Earth Sciences Series*. Springer International Publishing, Cham, pp. 375–378. https://doi.org/10.1007/978-3-319-39312-4_155
- He, D., He, C., Li, P., Zhang, X., Shi, Q., Sun, Y., 2019. Optical and Molecular Signatures of Dissolved Organic Matter Reflect Anthropogenic Influence in a Coastal River, Northeast China. *Journal of Environmental Quality* 48, 603–613. <https://doi.org/10.2134/jeq2018.09.0330>
- Hecht, E.S., Scigelova, M., Eliuk, S., Makarov, A., 2019. Fundamentals and Advances of Orbitrap Mass Spectrometry, in: *Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Ltd, pp. 1–40. <https://doi.org/10.1002/9780470027318.a9309.pub2>

- Hedges, J.I., Oades, J.M., 1997. Comparative organic geochemistries of soils and marine sediments. *Organic Geochemistry* 27, 319–361. [https://doi.org/10.1016/S0146-6380\(97\)00056-9](https://doi.org/10.1016/S0146-6380(97)00056-9)
- Heinz, M., Graeber, D., Zak, D., Zwirnmann, E., Gelbrecht, J., Pusch, M.T., 2015. Comparison of Organic Matter Composition in Agricultural versus Forest Affected Headwaters with Special Emphasis on Organic Nitrogen. *Environ. Sci. Technol.* 49, 2081–2090. <https://doi.org/10.1021/es505146h>
- Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008a. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography* 53, 955–969. <https://doi.org/10.4319/lo.2008.53.3.0955>
- Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008b. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnol. Oceanogr.* 53, 955–969. <https://doi.org/10.4319/lo.2008.53.3.0955>
- Her, N., Amy, G., McKnight, D., Sohn, J., Yoon, Y., 2003. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. *Water Research* 37, 4295–4303. [https://doi.org/10.1016/S0043-1354\(03\)00317-8](https://doi.org/10.1016/S0043-1354(03)00317-8)
- Hertkorn, N., Harir, M., Cawley, K.M., Schmitt-Kopplin, P., Jaffé, R., 2016. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* 13, 2257–2277. <https://doi.org/10.5194/bg-13-2257-2016>
- Hocquet, C., Golla, G., 2008. Bilan 2008 de l'assainissement en France.
- Hoikkala, L., Kortelainen, P., Soinnie, H., Kuosa, H., 2015. Dissolved organic matter in the Baltic Sea. *Journal of Marine Systems* 142, 47–61. <https://doi.org/10.1016/j.jmarsys.2014.10.005>
- Hollender, J., Schymanski, E.L., 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>
- Holm, P.E., Rootzén, H., Borggaard, O.K., Møberg, J.P., Christensen, T.H., 2003. Correlation of cadmium distribution coefficients to soil characteristics. *J. Environ. Qual.* 32, 138–145. <https://doi.org/10.2134/jeq2003.1380>
- Hood, E., 2005. Tap Water and Trihalomethanes: Flow of Concerns Continues. *Environ Health Perspect* 113, A474.
- Hu, A., Choi, M., Tanentzap, A.J., Liu, J., Jang, K.-S., Lennon, J.T., Liu, Y., Soininen, J., Lu, X., Zhang, Y., Shen, J., Wang, J., 2022. Ecological networks of dissolved organic matter and microorganisms under global change. *Nat Commun* 13, 3600. <https://doi.org/10.1038/s41467-022-31251-1>
- Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J.-M., Parlanti, E., 2009. Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Organic Geochemistry* 40, 706–719. <https://doi.org/10.1016/j.orggeochem.2009.03.002>
- Ikeya, K., Sleighter, R.L., Hatcher, P.G., Watanabe, A., 2015. Characterization of the chemical composition of soil humic acids using Fourier transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta* 153, 169–182. <https://doi.org/10.1016/j.gca.2015.01.002>
- Jaffé, R., Boyer, J.N., Lu, X., Maie, N., Yang, C., Scully, N.M., Mock, S., 2004. Source characterization of dissolved organic matter in a subtropical mangrove-dominated estuary by fluorescence analysis. *Marine Chemistry* 84, 195–210. <https://doi.org/10.1016/j.marchem.2003.08.001>
- Jiao, N., Herndl, G.J., Hansell, D.A., Benner, R., Kattner, G., Wilhelm, S.W., Kirchman, D.L., Weinbauer, M.G., Luo, T., Chen, F., Azam, F., 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat Rev Microbiol* 8, 593–599. <https://doi.org/10.1038/nrmicro2386>
- Jiao, N., Robinson, C., Azam, F., Thomas, H., Baltar, F., Dang, H., Hardman-Mountford, N.J., Johnson, M., Kirchman, D.L., Koch, B.P., Legendre, L., Li, C., Liu, J., Luo, T., Luo, Y.-W., Mitra, A., Romanou, A., Tang, K., Wang, X., Zhang, C., Zhang, R., 2014. Mechanisms of microbial carbon

- sequestration in the ocean & future research directions. *Biogeosciences* 11, 5285–5306. <https://doi.org/10.5194/bg-11-5285-2014>
- Kaal, J., Cortizas, A.M., Biester, H., 2017. Downstream changes in molecular composition of DOM along a headwater stream in the Harz mountains (Central Germany) as determined by FTIR, Pyrolysis-GC-MS and THM-GC-MS. *Journal of Analytical and Applied Pyrolysis* 126, 50–61. <https://doi.org/10.1016/j.jaap.2017.06.025>
- Kalbitz, K., Schmerwitz, J., Schwesig, D., Matzner, E., 2003. Biodegradation of soil-derived dissolved organic matter as related to its properties. *Geoderma, Ecological aspects of dissolved organic matter in soils* 113, 273–291. [https://doi.org/10.1016/S0016-7061\(02\)00365-8](https://doi.org/10.1016/S0016-7061(02)00365-8)
- Kalbitz, K., Solinger, S., Park, J.-H., Michalzik, B., Matzner, E., 2000. CONTROLS ON THE DYNAMICS OF DISSOLVED ORGANIC MATTER IN SOILS: A REVIEW. *Soil Science* 165, 277–304.
- Kamjunke, N., von Tümpling, W., Hertkorn, N., Harir, M., Schmitt-Kopplin, P., Norf, H., Weitere, M., Herzsprung, P., 2017. A new approach for evaluating transformations of dissolved organic matter (DOM) via high-resolution mass spectrometry and relating it to bacterial activity. *Water Research* 123, 513–523. <https://doi.org/10.1016/j.watres.2017.07.008>
- Kaule, R., Gilfedder, B.S., 2021. Groundwater Dominates Water Fluxes in a Headwater Catchment During Drought. *Frontiers in Water* 3.
- Kim, C., Eom, J.B., Jung, S., Ji, T., 2016. Detection of Organic Compounds in Water by an Optical Absorbance Method. *Sensors (Basel)* 16, 61. <https://doi.org/10.3390/s16010061>
- Kissman, C.E.H., Williamson, C.E., Rose, K.C., Saros, J.E., 2017. Nutrients associated with terrestrial dissolved organic matter drive changes in zooplankton:phytoplankton biomass ratios in an alpine lake. *Freshwater Biology* 62, 40–51. <https://doi.org/10.1111/fwb.12847>
- Kracht, D.G., 2001. 1. Prof. Dr. G. Büchel (Friedrich-Schiller-Universität Jena) 146.
- Krauss, M., Singer, H., Hollender, J., 2010. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal Bioanal Chem* 397, 943–951. <https://doi.org/10.1007/s00216-010-3608-9>
- Lambert, T., 2013. Sources, production et transfert du carbone organique dissous dans les bassins versants élémentaires sur socle: apports des isotopes stables du carbone 166.
- Lambert, T., Bouillon, S., Darchambeau, F., Morana, C., Roland, F.A.E., Descy, J.-P., Borges, A.V., 2017. Effects of human land use on the terrestrial and aquatic sources of fluvial organic matter in a temperate river basin (The Meuse River, Belgium). *Biogeochemistry* 136, 191–211. <https://doi.org/10.1007/s10533-017-0387-9>
- Lanxiu, Y., Fengchang, W., Congqiang, L., Wen, L., Jing, W., Yi, M., 2004. Molecular weight distribution of dissolved organic matter in Lake Hongfeng determined by high performance size exclusion chromatography (HPSEC) with on-line UV-vis absorbance and fluorescence detection. *Chin. J. Geochem.* 23, 275–283. <https://doi.org/10.1007/BF02842075>
- Lapworth, D.J., Goody, D.C., Butcher, A.S., Morris, B.L., 2008. Tracing groundwater flow and sources of organic carbon in sandstone aquifers using fluorescence properties of dissolved organic matter (DOM). *Applied Geochemistry* 23, 3384–3390. <https://doi.org/10.1016/j.apgeochem.2008.07.011>
- Le, N.D., France, X., Pontvianne, S., Poirot, H., Leclerc, J.P., Pons, M.N., 2017. Daily wastewater pollutant dynamics with respect to catchment population structure. *Urban Water Journal* 14, 1016–1022. <https://doi.org/10.1080/1573062X.2017.1325498>
- Lee, S.-A., Kim, T.-H., Kim, G., 2020. Tracing terrestrial versus marine sources of dissolved organic carbon in a coastal bay using stable carbon isotopes. *Biogeosciences* 17, 135–144. <https://doi.org/10.5194/bg-17-135-2020>
- Lennon, J.T., Pfaff, L.E., 2005. Source and supply of terrestrial organic matter affects aquatic microbial metabolism. *Aquatic Microbial Ecology* 39, 107–119. <https://doi.org/10.3354/ame039107>
- Li, C., Yan, F., Kang, S., Chen, P., Qu, B., Hu, Z., Sillanpää, M., 2016. Concentration, sources, and flux of dissolved organic carbon of precipitation at Lhasa city, the Tibetan Plateau. *Environ Sci Pollut Res* 23, 12915–12921. <https://doi.org/10.1007/s11356-016-6455-1>

- Li, P., Hur, J., 2017. Utilization of UV-Vis spectroscopy and related data analyses for dissolved organic matter (DOM) studies: A review. *Critical Reviews in Environmental Science and Technology* 47. <https://doi.org/10.1080/10643389.2017.1309186>
- Li, X., Dai, X., Takahashi, J., Li, N., Jin, J., Dai, L., Dong, B., 2014. New insight into chemical changes of dissolved organic matter during anaerobic digestion of dewatered sewage sludge using EEM-PARAFAC and two-dimensional FTIR correlation spectroscopy. *Bioresource Technology* 159, 412–420. <https://doi.org/10.1016/j.biortech.2014.02.085>
- Li, Y., Sun, L.-L., Sun, Y.-Y., Cha, Q.-Q., Li, C.-Y., Zhao, D.-L., Song, X.-Y., Wang, M., McMinn, A., Chen, X.-L., Zhang, Y.-Z., Qin, Q.-L., 2019. Extracellular Enzyme Activity and Its Implications for Organic Matter Cycling in Northern Chinese Marginal Seas. *Frontiers in Microbiology* 10.
- Liao, Z., Chu, J., Luo, C., Chen, H., 2021. Revealing the characteristics of dissolved organic matter in urban runoff at three typical regions via optical indices and molecular composition. *Journal of Environmental Sciences* 108, 8–21. <https://doi.org/10.1016/j.jes.2021.02.012>
- Liu, C., Li, Z., Berhe, A.A., Xiao, H., Liu, L., Wang, D., Peng, H., Zeng, G., 2019. Characterizing dissolved organic matter in eroded sediments from a loess hilly catchment using fluorescence EEM-PARAFAC and UV-Visible absorption: Insights from source identification and carbon cycling. *Geoderma* 334, 37–48. <https://doi.org/10.1016/j.geoderma.2018.07.029>
- Loiselle, S., Vione, D., Minero, C., Maurino, V., Tognazzi, A., Dattilo, A.M., Rossi, C., Bracchini, L., 2012. Chemical and optical phototransformation of dissolved organic matter. *Water Res* 46, 3197–3207. <https://doi.org/10.1016/j.watres.2012.02.047>
- Lu, Y., Shang, P., Chen, S., Du, Y., Bonizzoni, M., Ward, A.K., 2021. Discharge and Temperature Controls of Dissolved Organic Matter (DOM) in a Forested Coastal Plain Stream. *Water* 13, 2919. <https://doi.org/10.3390/w13202919>
- Luciani, X., 2008. Analyse numérique des spectres de fluorescence 3D issus de mélanges non linéaires 175.
- Lupon, A., Catalán, N., Martí, E., Bernal, S., 2020. Influence of Dissolved Organic Matter Sources on In-Stream Net Dissolved Organic Carbon Uptake in a Mediterranean Stream. *Water* 12, 1722. <https://doi.org/10.3390/w12061722>
- Ma, B., Ma, F., Li, Z., Wu, F., 2014. Effect of crops on rainfall redistribution processes under simulated rainfall. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering* 30, 136–146. <https://doi.org/10.3969/j.issn.1002-6819.2014.16.019>
- Mamyryn, B.A., 2001. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry* 206, 251–266. [https://doi.org/10.1016/S1387-3806\(00\)00392-4](https://doi.org/10.1016/S1387-3806(00)00392-4)
- Mangal, V., DeGasparro, S., Beresford, D.V., Guéguen, C., 2020. Linking molecular and optical properties of dissolved organic matter across a soil-water interface on Akimiski Island (Nunavut, Canada). *Science of The Total Environment* 704, 135415. <https://doi.org/10.1016/j.scitotenv.2019.135415>
- Matilainen, A., Gjessing, E.T., Lahtinen, T., Hed, L., Bhatnagar, A., Sillanpää, M., 2011. An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. *Chemosphere* 83, 1431–1442. <https://doi.org/10.1016/j.chemosphere.2011.01.018>
- Mayora, G., Devercelli, M., Frau, D., 2016. Spatial variability of chromophoric dissolved organic matter in a large floodplain river: control factors and relations with phytoplankton during a low water period. *Ecohydrology* 9, 487–497. <https://doi.org/10.1002/eco.1651>
- McCaul, M.V., Sutton, D., Simpson, A.J., Spence, A., McNally, D.J., Moran, B.W., Goel, A., O'Connor, B., Hart, K., Kelleher, B.P., 2011. Composition of dissolved organic matter within a lacustrine environment. *Environ. Chem.* 8, 146. <https://doi.org/10.1071/EN10109>
- McDonough, L.K., Andersen, M.S., Behnke, M.I., Rutledge, H., Oudone, P., Meredith, K., O'Carroll, D.M., Santos, I.R., Marjo, C.E., Spencer, R.G.M., McKenna, A.M., Baker, A., 2022. A new conceptual framework for the transformation of groundwater dissolved organic matter. *Nat Commun* 13, 2153. <https://doi.org/10.1038/s41467-022-29711-9>

- McElmurry, S.P., Long, D.T., Voice, T.C., 2014. Stormwater Dissolved Organic Matter: Influence of Land Cover and Environmental Factors. *Environ. Sci. Technol.* 48, 45–53. <https://doi.org/10.1021/es402664t>
- Mcknight, D., Boyer, E., Westerhoff, P., Doran, P., Kulbe, T., Andersen, D., 2001. Spectrofluorometric Characterization of Dissolved Organic Matter for Indication of Precursor Organic Material and Aromaticity. *Limnology and Oceanography* 46, 38–48. <https://doi.org/10.4319/lo.2001.46.1.0038>
- Menzel, D.W., Vaccaro, R.F., 1964. The Measurement of Dissolved Organic and Particulate Carbon in Seawater. *Limnology and Oceanography* 9, 138–142. <https://doi.org/10.4319/lo.1964.9.1.0138>
- Meyer, J.L., Paul, M.J., Taulbee, W.K., 2005. Stream ecosystem function in urbanizing landscapes. *Journal of the North American Benthological Society* 24, 602–612. <https://doi.org/10.1899/04-021.1>
- Michael-Kordatou, I., Michael, C., Duan, X., He, X., Dionysiou, D.D., Mills, M.A., Fatta-Kassinos, D., 2015. Dissolved effluent organic matter: Characteristics and potential implications in wastewater treatment and reuse applications. *Water Research* 77, 213–248. <https://doi.org/10.1016/j.watres.2015.03.011>
- Minor, E.C., Steinbring, C.J., Longnecker, K., Kujawinski, E.B., 2012. Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. *Organic Geochemistry* 43, 1–11. <https://doi.org/10.1016/j.orggeochem.2011.11.007>
- Mitchell, P.J., Simpson, A.J., Soong, R., Simpson, M.J., 2018. Nuclear Magnetic Resonance Analysis of Changes in Dissolved Organic Matter Composition with Successive Layering on Clay Mineral Surfaces. *Soil Systems* 2, 8. <https://doi.org/10.3390/soils2010008>
- Mitschke, N., Vemulapalli, S.P.B., Dittmar, T., 2022. NMR spectroscopy of dissolved organic matter: a review. *Environ Chem Lett.* <https://doi.org/10.1007/s10311-022-01528-4>
- Mjejjra, M., 2015. Étude de l'évapotranspiration dans le bassin versant de Mejerda (en Tunisie): apport de la télédétection satellitaire et des Systèmes d'Information Géographique 311.
- Molle, P., 2012. Les filtres plantés de roseaux : évolution de la recherche et tendances actuelles. *Sciences Eaux & Territoires Numéro 9*, 24. <https://doi.org/10.3917/set.009.0024>
- Moore, T.R., 1989. Dynamics of dissolved organic carbon in forested and disturbed catchments, Westland, New Zealand: 1. Maimai. *Water Resources Research* 25, 1321–1330. <https://doi.org/10.1029/WR025i006p01321>
- Moran, M.A., Sheldon, W., Zepp, R., 2000. Carbon Loss and Optical Property Changes during Long-Term Photochemical and Biological Degradation of Estuarine Dissolved Organic Matter. *Limnology and Oceanography - LIMNOL OCEANOGR* 45, 1254–1264. <https://doi.org/10.4319/lo.2000.45.6.1254>
- Mounier, S., Zhao, H., Garnier, C., Redon, R., 2011. Copper complexing properties of dissolved organic matter: PARAFAC treatment of fluorescence quenching. *Biogeochemistry* 106, 107–116. <https://doi.org/10.1007/s10533-010-9486-6>
- Murphy, K.R., Hambly, A., Singh, S., Henderson, R.K., Baker, A., Stuetz, R., Khan, S.J., 2011. Organic Matter Fluorescence in Municipal Water Recycling Schemes: Toward a Unified PARAFAC Model. *Environ. Sci. Technol.* 45, 2909–2916. <https://doi.org/10.1021/es103015e>
- Murphy, K.R., Stedmon, C.A., Waite, T.D., Ruiz, G.M., 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Marine Chemistry* 108, 40–58. <https://doi.org/10.1016/j.marchem.2007.10.003>
- Murphy, K.R., Stedmon, C.A., Wenig, P., Bro, R., 2014. OpenFluor– an online spectral library of auto-fluorescence by organic compounds in the environment. *Anal. Methods* 6, 658–661. <https://doi.org/10.1039/C3AY41935E>
- Myers, O.D., Sumner, S.J., Li, S., Barnes, S., Du, X., 2017. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection

- Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Anal. Chem.* 89, 8689–8695. <https://doi.org/10.1021/acs.analchem.7b01069>
- Nihemaiti, M., Huynh, N., Mailler, R., Mèche-Ananit, P., Rocher, V., Barhdadi, R., Moillon, R., Le Roux, J., 2022. High-Resolution Mass Spectrometry Screening of Wastewater Effluent for Micropollutants and Their Transformation Products during Disinfection with Performic Acid. *ACS EST Water* 2, 1225–1233. <https://doi.org/10.1021/acsestwater.2c00075>
- Nouhi, A., 2017. Caractérisation Spectrale et Temporelle par Quenching de Fluorescence des Interactions Matière Organique-Éléments Métalliques 197.
- Opsahl, S., Benner, R., 1997. Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature* 386, 480–482. <https://doi.org/10.1038/386480a0>
- Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M., 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Organic Geochemistry* 31, 1765–1781. [https://doi.org/10.1016/S0146-6380\(00\)00124-8](https://doi.org/10.1016/S0146-6380(00)00124-8)
- Paszkiwicz, M., Godlewska, K., Lis, H., Caban, M., Białk-Bielińska, A., Stepnowski, P., 2022. Advances in suspect screening and non-target analysis of polar emerging contaminants in the environmental monitoring. *TrAC Trends in Analytical Chemistry* 154, 116671. <https://doi.org/10.1016/j.trac.2022.116671>
- Peleato, N.M., Legge, R.L., Andrews, R.C., 2017. Investigation of fluorescence methods for rapid detection of municipal wastewater impact on drinking water sources. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 171, 104–111. <https://doi.org/10.1016/j.saa.2016.07.031>
- Penna, D., Oliviero, O., Assendelft, R., Zuecco, G., Meerveld, I. (H. J.) van, Anfodillo, T., Carraro, V., Borga, M., Fontana, G.D., 2013. Tracing the Water Sources of Trees and Streams: Isotopic Analysis in a Small Pre-Alpine Catchment. *Procedia Environmental Sciences, Four Decades of Progress in Monitoring and Modeling of Processes in the Soil-Plant-Atmosphere System: Applications and Challenges* 19, 106–112. <https://doi.org/10.1016/j.proenv.2013.06.012>
- Petitjean, P., Henin, O., Gruau, G., 2004. DOSAGE DU CARBONE ORGANIQUE DISSOUS DANS LES EAUX DOUCES NATURELLES.
- Pluskal, T., Castillo, S., Villar-Briones, A., Oresic, 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>
- Reader, H.E., Stedmon, C.A., Nielsen, N.J., Kritzberg, E.S., 2015. Mass and UV-visible spectral fingerprints of dissolved organic matter: sources and reactivity. *Front. Mar. Sci.* 2. <https://doi.org/10.3389/fmars.2015.00088>
- Riedel, T., Zark, M., Vähätalo, A.V., Niggemann, J., Spencer, R.G.M., Hernes, P.J., Dittmar, T., 2016. Molecular Signatures of Biogeochemical Transformations in Dissolved Organic Matter from Ten World Rivers. *Frontiers in Earth Science* 4.
- Ritter, W.F., Shirmohammadi, A., 2000. Agricultural Nonpoint Source Pollution: Watershed Management and Hydrology. CRC Press.
- Rocco, K., Margoum, C., Richard, L., Coquery, M., 2022. Enhanced database creation with in silico workflows for suspect screening of unknown tebuconazole transformation products in environmental samples by UHPLC-HRMS. *Journal of Hazardous Materials* 440, 129706. <https://doi.org/10.1016/j.jhazmat.2022.129706>
- Rodríguez, F.J., Schlenger, P., García-Valverde, M., 2016. Monitoring changes in the structure and properties of humic substances following ozonation using UV-Vis, FTIR and ¹H NMR techniques. *Science of The Total Environment* 541, 623–637. <https://doi.org/10.1016/j.scitotenv.2015.09.127>
- Rosario-Ortiz, F. (Ed.), 2014. Advances in the Physicochemical Characterization of Dissolved Organic Matter: Impact on Natural and Engineered Systems, ACS Symposium Series. American Chemical Society, Washington, DC. <https://doi.org/10.1021/bk-2014-1160>

- Rutledge, H., McDonough, L.K., Oudone, P., Andersen, M.S., Meredith, K., Chinu, K., Peterson, M., Baker, A., 2021. Characterisation of groundwater dissolved organic matter using LCOCD: Implications for water treatment. *Water Research* 188, 116422. <https://doi.org/10.1016/j.watres.2020.116422>
- Sachse, A., Henrion, R., Gelbrecht, J., Steinberg, C.E.W., 2005. Classification of dissolved organic carbon (DOC) in river systems: Influence of catchment characteristics and autochthonous processes. *Organic Geochemistry* 36, 923–935. <https://doi.org/10.1016/j.orggeochem.2004.12.008>
- Salonen, K., Kairesalo, T., Jones, R.I. (Eds.), 1992. *Dissolved Organic Matter in Lacustrine Ecosystems: Energy Source and System Regulator*. Springer Netherlands, Dordrecht. <https://doi.org/10.1007/978-94-011-2474-4>
- Sanderman, J., Lohse, K.A., Baldock, J.A., Amundson, R., 2009. Linking soils and streams: Sources and chemistry of dissolved organic matter in a small coastal watershed. *Water Resources Research* 45. <https://doi.org/10.1029/2008WR006977>
- Sankar, M.S., Dash, P., Singh, S., Lu, Y., Mercer, A.E., Chen, S., 2019. Effect of photo-biodegradation and biodegradation on the biogeochemical cycling of dissolved organic matter across diverse surface water bodies. *Journal of Environmental Sciences* 77, 130–147. <https://doi.org/10.1016/j.jes.2018.06.021>
- Schittich, A.-R., Wünsch, U.J., Kulkarni, H.V., Battistel, M., Bregnhøj, H., Stedmon, C.A., McKnight, U.S., 2018. Investigating Fluorescent Organic-Matter Composition as a Key Predictor for Arsenic Mobility in Groundwater Aquifers. *Environ. Sci. Technol.* 52, 13027–13036. <https://doi.org/10.1021/acs.est.8b04070>
- 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>
- Seidel, M., Vemulapalli, S.P.B., Mathieu, D., Dittmar, T., 2022. Marine Dissolved Organic Matter Shares Thousands of Molecular Formulae Yet Differs Structurally across Major Water Masses. *Environ. Sci. Technol.* 56, 3758–3769. <https://doi.org/10.1021/acs.est.1c04566>
- Shen, Y., Benner, R., Robbins, L.L., Wynn, J.G., 2016. Sources, Distributions, and Dynamics of Dissolved Organic Matter in the Canada and Makarov Basins. *Frontiers in Marine Science* 3.
- Sheng, H., Cai, T., 2019. Influence of Rainfall on Canopy Interception in Mixed Broad-Leaved—Korean Pine Forest in Xiaoxing'an Mountains, Northeastern China. *Forests* 10, 248. <https://doi.org/10.3390/f10030248>
- Simpson, A.J., 2002. Determining the molecular weight, aggregation, structures and interactions of natural organic matter using diffusion ordered spectroscopy. *Magnetic Resonance in Chemistry* 40, S72–S82. <https://doi.org/10.1002/mrc.1106>
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78, 779–787. <https://doi.org/10.1021/ac051437y>
- Song, J., Peng, P., 2010. Characterisation of black carbon materials by pyrolysis–gas chromatography–mass spectrometry. *Journal of Analytical and Applied Pyrolysis* 87, 129–137. <https://doi.org/10.1016/j.jaap.2009.11.003>
- Sophocleous, M., 2002. Interactions between groundwater and surface water: the state of the science. *Hydrogeology Journal* 10, 52–67. <https://doi.org/10.1007/s10040-001-0170-8>
- Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry* 82, 239–254. [https://doi.org/10.1016/S0304-4203\(03\)00072-0](https://doi.org/10.1016/S0304-4203(03)00072-0)
- Tang, X., Wu, Q.-Y., Zhao, X., Huang, H., Shi, X.-J., Hu, H.-Y., 2014. A fingerprint analysis method for characterization of dissolved organic matter in secondary effluents of municipal wastewater treatment plant. *Environ Sci Pollut Res* 21, 14211–14218. <https://doi.org/10.1007/s11356-014-3336-3>

- Tenorio, R., Maizel, A.C., Schaefer, C.E., Higgins, C.P., Strathmann, T.J., 2022. Application of High-Resolution Mass Spectrometry to Evaluate UV-Sulfite-Induced Transformations of Per- and Polyfluoroalkyl Substances (PFASs) in Aqueous Film-Forming Foam (AFFF). *Environ. Sci. Technol.* 56, 14774–14787. <https://doi.org/10.1021/acs.est.2c03228>
- Thai, S., Pavlů, L., Tejnecký, V., Vokurková, P., Nozari, S., Borůvka, L., 2021. Comparison of soil organic matter composition under different land uses by DRIFT spectroscopy. *Plant Soil Environ.* 67, 255–263. <https://doi.org/10.17221/11/2021-PSE>
- Toming, K., Tuvikene, L., Vilbaste, S., Agasild, H., Viik, M., Kisand, A., Feldmann, T., Martma, T., Jones, R.I., Nõges, T., 2013. Contributions of autochthonous and allochthonous sources to dissolved organic matter in a large, shallow, eutrophic lake with a highly calcareous catchment. *Limnology and Oceanography* 58, 1259–1270. <https://doi.org/10.4319/lo.2013.58.4.1259>
- 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>
- Twardowski, M.S., Boss, E., Sullivan, J.M., Donaghay, P.L., 2004. Modeling the spectral shape of absorption by chromophoric dissolved organic matter. *Marine Chemistry* 89, 69–88. <https://doi.org/10.1016/j.marchem.2004.02.008>
- Unkovich, M.J., Pate, J.S., McNeill, A., Gibbs, J., 2013. *Stable Isotope Techniques in the Study of Biological Processes and Functioning of Ecosystems*. Springer Science & Business Media.
- 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>
- Villafañe, V.E., Paczkowska, J., Andersson, A., Durán Romero, C., Valiñas, M.S., Helbling, E.W., 2018. Dual role of DOM in a scenario of global change on photosynthesis and structure of coastal phytoplankton from the South Atlantic Ocean. *Science of The Total Environment* 634, 1352–1361. <https://doi.org/10.1016/j.scitotenv.2018.04.121>
- Wang, T., Duedahl-Olesen, L., Lauritz Frandsen, H., 2021. Targeted and non-targeted unexpected food contaminants analysis by LC/HRMS: Feasibility study on rice. *Food Chemistry* 338, 127957. <https://doi.org/10.1016/j.foodchem.2020.127957>
- Webb, J.R., Santos, I.R., Maher, D.T., Tait, D.R., Cyronak, T., Sadat-Noori, M., Macklin, P., Jeffrey, L.C., 2019. Groundwater as a source of dissolved organic matter to coastal waters: Insights from radon and CDOM observations in 12 shallow coastal systems. *Limnology and Oceanography* 64, 182–196. <https://doi.org/10.1002/lno.11028>
- Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Fram, M.S., Fujii, R., Mopper, K., 2003. Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. *Environ. Sci. Technol.* 37, 4702–4708. <https://doi.org/10.1021/es030360x>
- Williams, C.J., Yamashita, Y., Wilson, H.F., Jaffé, R., Xenopoulos, M.A., 2010. Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. *Limnology and Oceanography* 55, 1159–1171. <https://doi.org/10.4319/lo.2010.55.3.1159>
- Wu, X., Wu, L., Liu, Y., Zhang, P., Li, Q., Zhou, J., Hess, N.J., Hazen, T.C., Yang, W., Chakraborty, R., 2018. Microbial Interactions With Dissolved Organic Matter Drive Carbon Dynamics and Community Succession. *Front Microbiol* 9. <https://doi.org/10.3389/fmicb.2018.01234>
- Wünsch, U.J., Geuer, J.K., Lechtenfeld, O.J., Koch, B.P., Murphy, K.R., Stedmon, C.A., 2018. Quantifying the impact of solid-phase extraction on chromophoric dissolved organic matter composition. *Marine Chemistry* 207, 33–41. <https://doi.org/10.1016/j.marchem.2018.08.010>
- Xia, F., Liu, Z., Zhao, M., Li, Q., Li, D., Cao, W., Zeng, C., Hu, Y., Chen, B., Bao, Q., Zhang, Y., He, Q., Lai, C., He, X., Ma, Z., Han, Y., He, H., 2022. High stability of autochthonous dissolved organic matter in karst aquatic ecosystems: Evidence from fluorescence. *Water Research* 220, 118723. <https://doi.org/10.1016/j.watres.2022.118723>

- Zepp, R.G., Sheldon, W.M., Moran, M.A., 2004. Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation–emission matrices. *Marine Chemistry, CDOM in the Ocean: Characterization, Distribution and Transformation* 89, 15–36. <https://doi.org/10.1016/j.marchem.2004.02.006>
- Zhang, H., Zheng, Y., Wang, X.C., Wang, Y., Dzakpasu, M., 2021. Characterization and biogeochemical implications of dissolved organic matter in aquatic environments. *Journal of Environmental Management* 294, 113041. <https://doi.org/10.1016/j.jenvman.2021.113041>
- Zhu, Y., Vieth-Hillebrand, A., Noah, M., Poetz, S., 2019. Molecular characterization of extracted dissolved organic matter from New Zealand coals identified by ultrahigh resolution mass spectrometry. *International Journal of Coal Geology* 203, 74–86. <https://doi.org/10.1016/j.coal.2019.01.007>
- Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere* 38, 45–50. [https://doi.org/10.1016/S0045-6535\(98\)00166-0](https://doi.org/10.1016/S0045-6535(98)00166-0)

Chapitre 2.

Objectifs et questionnement scientifique

Chapitre 2. Objectifs et questionnement scientifique

L'objectif général de ma thèse est de caractériser la MOD issue de différents types de sources à l'échelle d'un bassin versant, dans le but de mettre en évidence des marqueurs physico-chimiques propres à ces sources. Ces marqueurs sont une série d'indicateurs (e.g. aromaticité, distribution en taille, humification) issus d'un large panel de techniques analytiques.

Suite à l'étude bibliographique et pour répondre à cet objectif général, j'ai identifié trois principales questions de recherche qui m'amènent à définir des objectifs plus spécifiques.

1. Est-il possible de mettre en évidence de nouveaux indicateurs optiques et moléculaires pour caractériser la MOD ?

Hypothèse :

L'apport de certaines techniques couplées, comme la HPSEC/UV-fluorescence ou la LC-HRMS, à la caractérisation de la MOD reste aujourd'hui relativement limité par rapport aux techniques classiques comme la spectroscopie UV-Visible et l'EMM de fluorescence. Cela est dû au fait que les données obtenues par les techniques HPSEC/UV-fluorescence et LC-HRMS sont complexes et nécessitent plusieurs étapes de traitement (*i.e.* décomposition des spectres UV et de fluorescence, tri des signaux HRMS ...) avant de pouvoir être interprétées. Notre faisons ici l'hypothèse qu'une étude approfondie des données issues de ces deux couplages devrait permettre de mettre en lumière de nouveaux indicateurs de caractérisation de la MOD, de différencier les types de sources définies *a priori* et, ainsi, de pallier certaines lacunes rencontrées avec les indicateurs classiques.

Objectif :

L'objectif est de mettre en évidence des indicateurs innovants permettant de décrire la MOD et qui soient pertinents pour discriminer les sources à l'échelle d'un bassin versant.

Démarche scientifique :

Le couplage HPSEC/UV-fluorescence présente un avantage considérable dans la caractérisation de la MOD. En effet, la séparation en taille de molécules couplée à la double détection UV et fluorescence permet d'enrichir l'information sur les fractions chromophoriques et fluorescentes par rapport aux techniques de spectroscopie UV-visible et EEM de fluorescence. Afin d'exploiter au mieux cette information, des indicateurs basés sur la

décomposition gaussienne des chromatogrammes HPSEC/UV-fluorescence ont ainsi été développés dans cette thèse. Ces indicateurs, qui reflètent des contributions de petites, moyennes et grosses molécules, ont été calculés à des longueurs d'ondes spécifiques des composés de types humique et protéique.

Concernant la spectrométrie de masse haute résolution (couplage LC-HRMS), la mise en place d'une méthodologie de traitement des données a constitué un premier défi. En effet, la complexité de l'information générée par cette technique (à savoir des milliers de composés moléculaires) nécessite la mise en place d'étapes de filtration de la donnée (*workflow* de retraitement et de filtration) au moyen de contrôles de qualité (e.g. blancs d'injections, triplicats d'injections randomisés). Le but de cette méthodologie est d'obtenir une donnée épurée, robuste et fiable qui ne soit pas parasitée par des centaines de composés moléculaires non-pertinents. Après la mise en place de ces étapes de filtration de la donnée, le deuxième défi a été l'interprétation de l'information obtenue. Pour cela, des indicateurs basés sur la répartition tridimensionnelle des composés moléculaires (polarité, masse, intensité) ont été développés

Pour tester la pertinence des nouveaux indicateurs HPSEC/UV-fluorescence et LC-HRMS, je me suis basé sur différents jeux de données diversifiées en types de sources (apports diffus terrigènes, rejets de STEU). J'ai ensuite confronté ces indicateurs à ceux classiquement utilisés pour caractériser la MOD (e.g. SUVA, HIX, FI), dans le but d'améliorer l'interprétation des nouveaux indicateurs ainsi développés, mais également d'apporter un regard critique sur les indicateurs classiques.

2. Quelle méthodologie faut-il adopter pour échantillonner des sources diffuses terrigènes de MOD ?

Hypothèses :

La MOD terrigène prélevée dans le cours d'eau garde une marque de son origine en termes d'occupation des sols. Ainsi, le prélèvement dans des cours d'eau en aval de sous-bassins versants homogènes de ces bassins versants, devrait fournir une empreinte spécifique de l'occupation des sols. Ces empreintes spécifiques pourront alors être réutilisées dans de futurs modèles de mélange dans le but de mieux tracer ces sources.

Par ailleurs, les conditions hydrologiques ont un impact sur le transfert de MOD. Un prélèvement en période d'étiage serait par exemple moins représentatif d'une MOD terrigène, contrairement à un prélèvement en période humide avec des précipitations atmosphériques (entraînant un lessivage des sols). De plus, un écoulement de surface de sol permettrait le transfert d'une matière organique fraîche riche en composés labiles (peu dégradés), contrairement à un écoulement vertical en profondeur qui favoriserait le transfert d'une MOD dégradée et humifiée.

Objectif :

L'objectif est donc d'évaluer la représentativité de l'échantillonnage d'une source diffuse terrigène de MOD dans les cours d'eau. Plus précisément, il s'agit d'étudier le lien (caractéristiques physicochimiques similaires) entre la MOD extraite de sols caractérisés par une occupation particulière, et la MOD collectée dans les cours d'eau adjacents à ces sols.

Démarche scientifique :

Des sous-bassins versants homogènes en termes d'occupation des sols (forêts, pâturages et vignes) ont été sélectionnés dans le bassin versant de l'Ardières-Morcille (au nord de Lyon). Au niveau de chaque sous-bassin, trois échantillonnages ont été effectués : un échantillon de sol représentatif d'une occupation bien identifiée et deux échantillons d'eau prélevés dans les cours d'eau adjacents au sol. Les deux prélèvements d'eau ont été réalisés à la fois en période de hautes eaux et en période de basses eaux. La MOD a ensuite été extraite des sols, dégradée afin d'éliminer le *pool* de molécules labiles, puis comparée à la MOD des cours d'eaux adjacents. La comparaison a été réalisée en utilisant trois techniques : la spectrophotométrie UV-Visible, la fluorescence 3D et la chromatographie d'exclusion stérique HPSEC/UV-fluorescence.

Dans une seconde partie, d'autres sources diffuses ont été prélevées à l'échelle de plusieurs bassins versants, afin de valider la méthode d'échantillonnage et de comparer, entre les différents bassins, les empreintes de sources diffuses similaires en termes d'usage des sols.

3. L'approche intégrée multi-analytique, multi-sources, permet-elle d'utiliser la MOD comme traceur de l'origine des masses d'eaux et des pollutions associées dans les cours d'eaux ?

Hypothèses :

La combinaison de plusieurs techniques (caractérisation intégrée) devrait permettre d'avoir une information plus complète et plus pertinente sur la MOD étudiée. En effet, chaque technique utilisée dans le cadre de ce travail apporte une information partielle, caractéristique d'une fraction de la MOD (e.g. fraction chromophorique, fraction fluorescente, fractions des différentes tailles ...). Ainsi, l'originalité de l'approche intégrée consiste justement en la prise en compte de toutes ces informations partielles pour présenter un portrait détaillé de la MOD.

En construisant une base d'échantillons caractéristiques de plusieurs types de sources de MOD, définis *a priori*, il devrait être possible de mettre en évidence des différences marquées entre ces types de sources.

Objectif :

L'objectif est de déterminer s'il existe un lien entre les caractéristiques de la MOD et les types de sources rencontrées à l'échelle d'un bassin versant périurbain, caractérisé par des zones naturelles et des zones à plus ou moins fortes pressions anthropiques, avec des activités agricoles et urbaines.

Démarche scientifique :

Cette dernière problématique a été traitée en mettant en place une base de données résultant de la caractérisation intégrée multi-analytique de plus de 150 échantillons représentant les principales sources de MOD dans les cours d'eaux. Ces sources incluent les sources diffuses étudiées dans la question 2 (échantillons prélevés dans le bassin versant de l'Ardières-Morcille) et également des sources ponctuelles de rejets urbains (eaux usées domestiques non traitées, rejets de STEU, rejets d'eaux pluviales urbaines). Les indicateurs optiques classiques ainsi que les indicateurs développés pour la question 1 ont été testés, via des analyses statistiques multivariées, pour discriminer les différents types de sources inclus dans la base de données.

Chapitre 3.

Matériels et Méthodes

Chapitre 3. Matériels et méthodes

La méthodologie employée dans le cadre de la thèse a impliqué la construction d'une base de données multi-analytiques, utilisant un nombre important d'échantillons représentatifs de sources diffuses et d'échantillons d'eaux usées ou urbaines. Les échantillons des sources diffuses ont été prélevés dans différents cours d'eau et retenues d'eau de trois bassins versants, tandis que les échantillons d'eaux usées et urbaines ont été prélevés dans des STEU et dans des systèmes de rejets d'eaux de ruissellement urbain. Les analyses réalisées sur ces échantillons (DOC, UV-Visible, EEM de fluorescence, HPSEC/UV-fluorescence et LC-HRMS) ont été intégrées dans une base de données qui a été utilisée pour répondre aux différentes questions posées dans le cadre de la thèse.

3.1 Stratégie d'échantillonnage

3.1.1 Présentation des sites d'étude

3.1.1.1 Bassin versant de l'Ardières-Morcille

Le bassin versant (BV) de l'Ardières-Morcille (152 km²) est situé dans la région du Beaujolais, au nord-est de la ville de Lyon (Rhône). La partie aval du BV est caractérisée par des activités viticoles représentant plus de 32 % du paysage total (Peyrard et al., 2016). D'autres occupations des sols, telles que les forêts (37 %) et les prairies (21 %), sont également présentes dans la partie amont (Figure 3-1).

La pluviométrie annuelle moyenne est de 940 mm avec des chutes de neige épisodiques en hiver. Les températures maximales moyennes en été sont d'environ 30°C et les températures minimales moyennes en hiver sont d'environ -5°C (Gouy et al., 2021). Les caractéristiques géologiques sont celles d'un socle cristallin hercynien altéré composé principalement de granite biotite porphyrique. Les sols sont majoritairement sableux et érodables. Le BV est caractérisé par des pentes très fortes, comprises entre 5 et 30-35 %, favorisant le ruissellement de surface. De plus, des événements hydrologiques nombreux et rapides font que le lit de la rivière est rarement à sec (Rabiet et al., 2015).

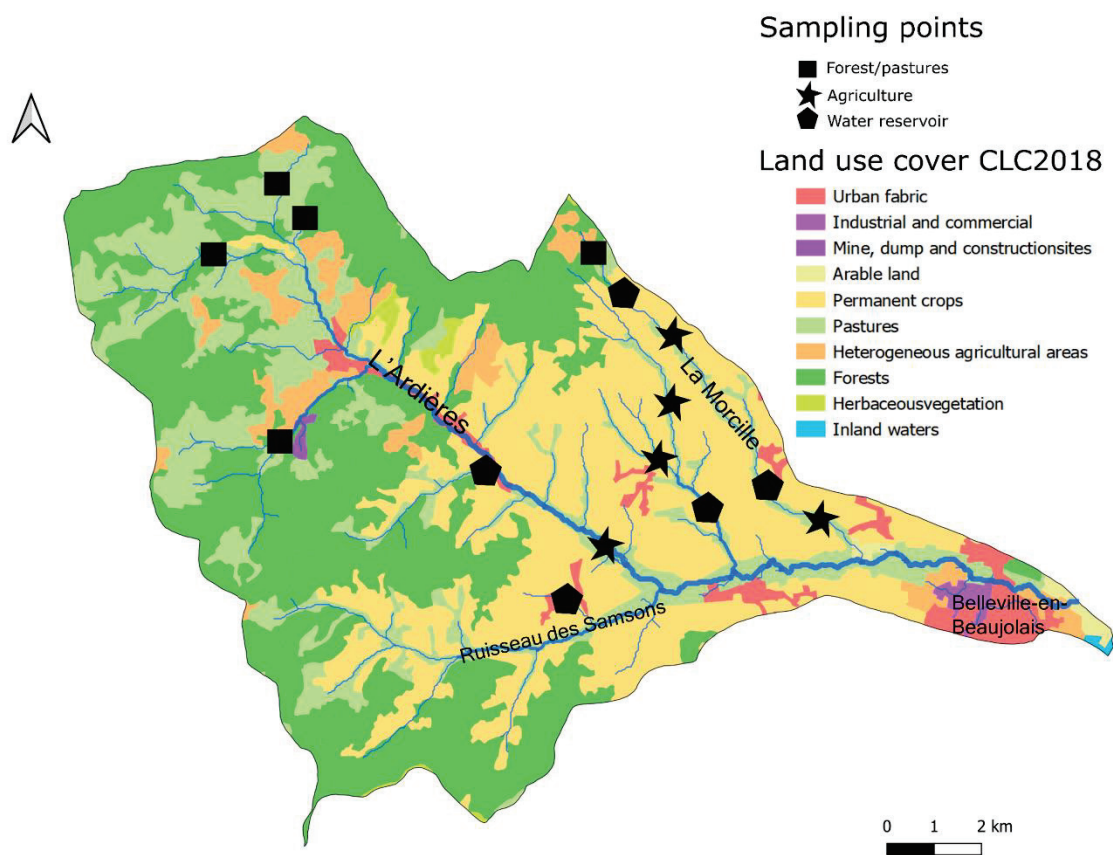


Figure 3-1. Carte du bassin versant de l’Ardières-Morcille indiquant la localisation des sites d’échantillonnage et les limites de l’occupation des sols (*Corine Land Cover CLC 2018*).

3.1.1.2 Bassin versant de l’Yzeron

Le BV de l’Yzeron (147 km²) est un BV péri-urbain situé à l’ouest de la ville de Lyon (Rhône). Le BV de l’Yzeron fait partie de l’Observatoire de Terrain en Hydrologie Urbaine (OTHU) depuis 1997 pour la réalisation d’un suivi des flux d’eau ainsi que des sources de contamination urbaine et leur impact sur les milieux aquatiques (<http://www.graie.org/othu/index.htm>). Depuis 2016, il fait partie de l’infrastructure de recherche Observatoires de la Zone Critique Applications et Recherche (OZCAR). L’exutoire de l’Yzeron se situe à 162 m d’altitude dans le Rhône et le point le plus élevé se situe à 917 m dans les monts du Lyonnais (Braud et al., 2013). La géologie du site est dominée par des roches cristallines et métamorphiques de type gneiss, granite et schiste à l’ouest du bassin. A l’est du BV, sur la partie aval, se trouvent des alluvions quaternaires sur base cristalline et métamorphique (Gnouma, 2006.; Labbas, 2017). L’Yzeron est caractérisé par un climat semi-continental avec des températures maximales en été de 17°C et minimales en hiver de 8°C et les précipitations annuelles moyenne sont de 830 mm (Labbas, 2017).

Les usages et occupations des sols dans le BV de l'Yzeron (Figure 3-2) sont très hétérogènes. En amont du bassin, la couverture végétale est partagée entre forêts (feuillus et conifères) et exploitations agricoles (zones de prairies et diverses cultures) avec la présence de quelques zones urbaines. Au centre du BV, les cultures agricoles et les prairies représentent la majorité des usages des sols. Enfin, en aval, la présence d'une forte densité urbaine vient compléter le gradient naturel-anthropique de l'amont vers l'aval.

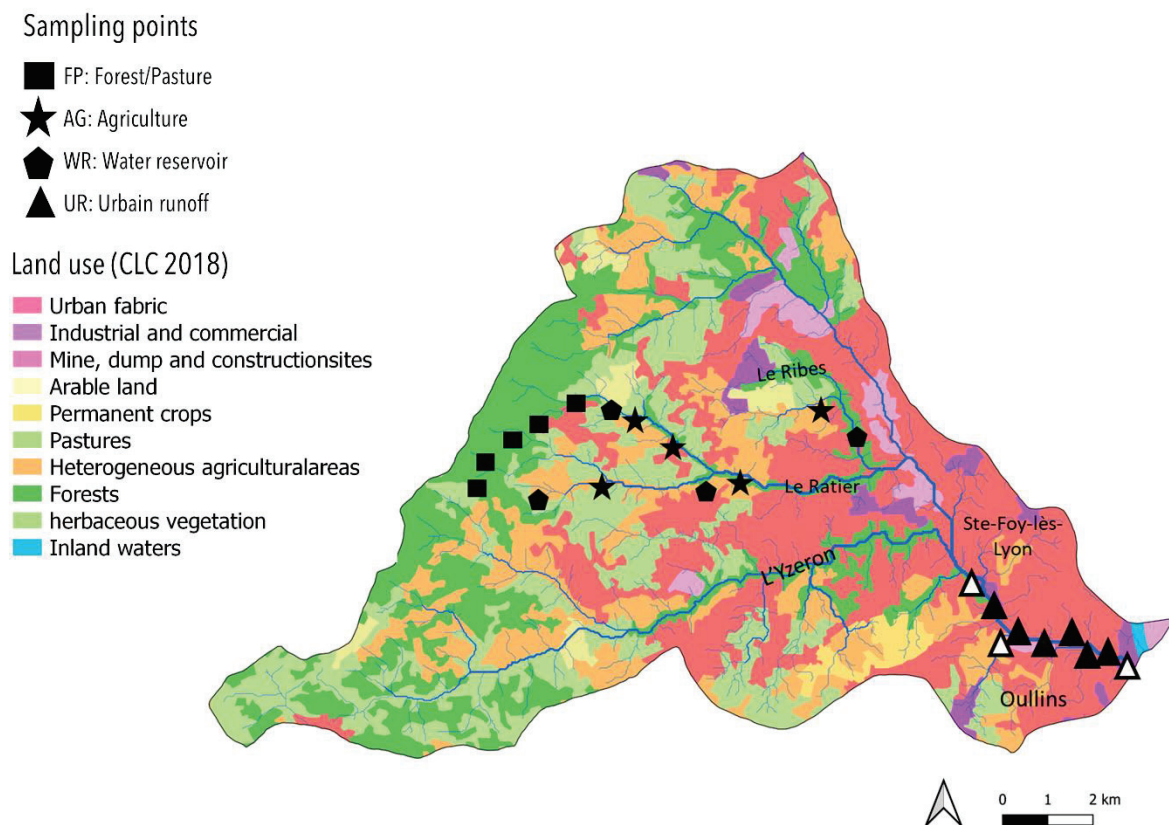


Figure 3-2. Carte du bassin versant de l'Yzeron indiquant la localisation des sites d'échantillonnage et les limites de l'occupation des sols (Corine Land Cover CLC 2018). Les triangles blancs représentent les points ruissellement urbains prélevés deux fois pour la variabilité temporelle (voir partie 1.5)

3.1.1.3 Bassin versant du Gier-Grand Malval

Le BV du Grand Malval (80 km²) est un sous-bassin versant du Gier (425 km²). Situé au sud de la ville de Lyon (Rhône), le bassin est caractérisé par des occupations des sols de forêts (mélanges de conifères et de feuillus) et prairies avec quelques activités agricoles et d'élevage, ainsi qu'une forte densité urbaine et industrielle en aval (Figure 3-3). Le cours d'eau principal du BV (Grand Malval) est l'un des affluents de la rivière du Gier qui trouve sa source dans le Pilat à 1 400 m d'altitude et se jette dans le Rhône au niveau de la ville de Givors. La géologie du site est principalement caractérisée par des roches plutoniques, métamorphiques ou sédimentaires anciennes qui forment des massifs rocheux imperméables. L'absence de stockage souterrain permet donc une mobilisation très importante des eaux de surface, avec la construction de nombreuses retenues collinaires et barrages pour le maintien des ressources en période d'étiage (Rapport EVP, 2010). La pluviométrie annuelle moyenne est de 800 mm, les températures maximales moyennes en été sont d'environ 30°C et les températures minimales moyennes en hiver sont d'environ -5°C.

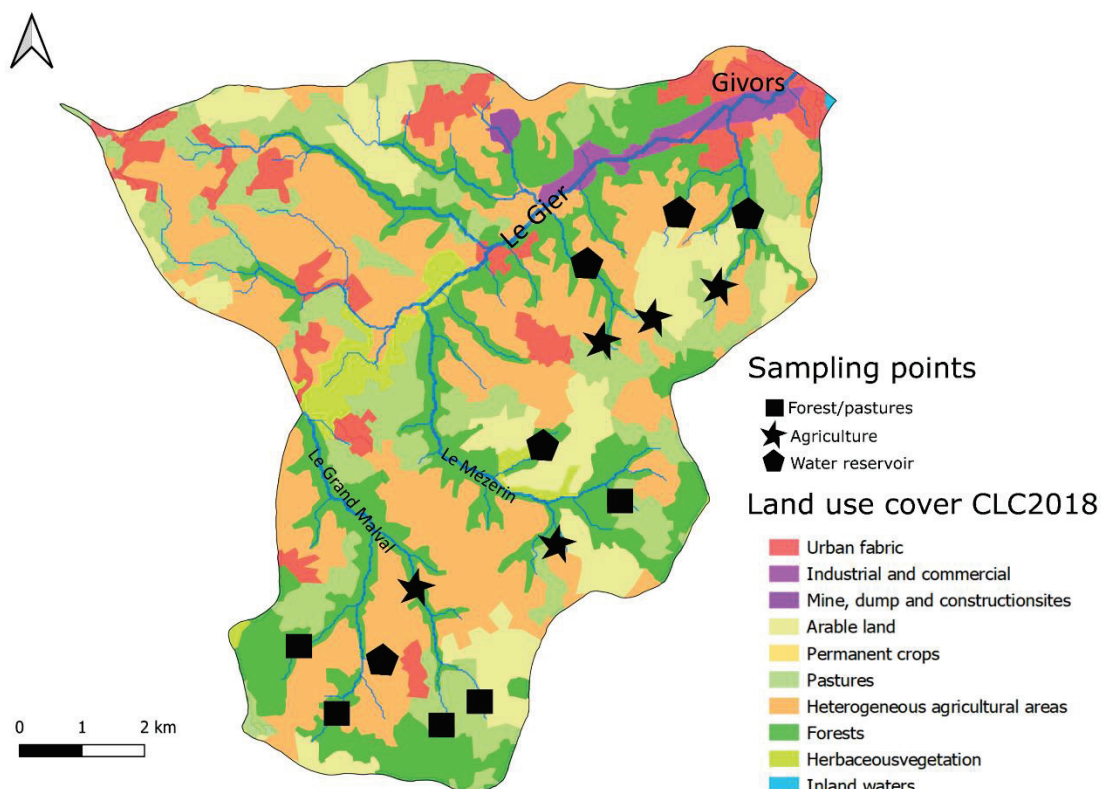


Figure 3-3. Carte du bassin versant du Gier-Grand Malval indiquant la localisation des sites d'échantillonnage et les limites de l'occupation des sols (*Corine Land Cover CLC 2018*).

3.1.2 Choix des sous-bassins homogènes et des sites de prélèvement des eaux associés

La première étape a été de sélectionner des points dans les cours d'eau à l'exutoire des sous-bassins versants les plus homogènes possible en terme d'occupation des sols. Ce travail géographique a été réalisé à l'aide d'un logiciel d'information géographique (QGIS version 3.10.0). Pour ce faire, les modèles numériques de terrain (MNT) des sites d'étude ont été téléchargés (<https://www.usgs.gov/>) puis utilisés pour définir les reliefs, les sens des écoulements, les cours d'eau et enfin les sous-bassins en amont des points d'échantillonnage. Une fois ces sous-bassins versants délimités, la dernière étape a été de croiser cette information avec la carte d'occupation des sols et de calculer la surface des occupations de sols drainée (Figure 3-4). Un exemple d'application de ces étapes est illustré sur la figure 3-5.

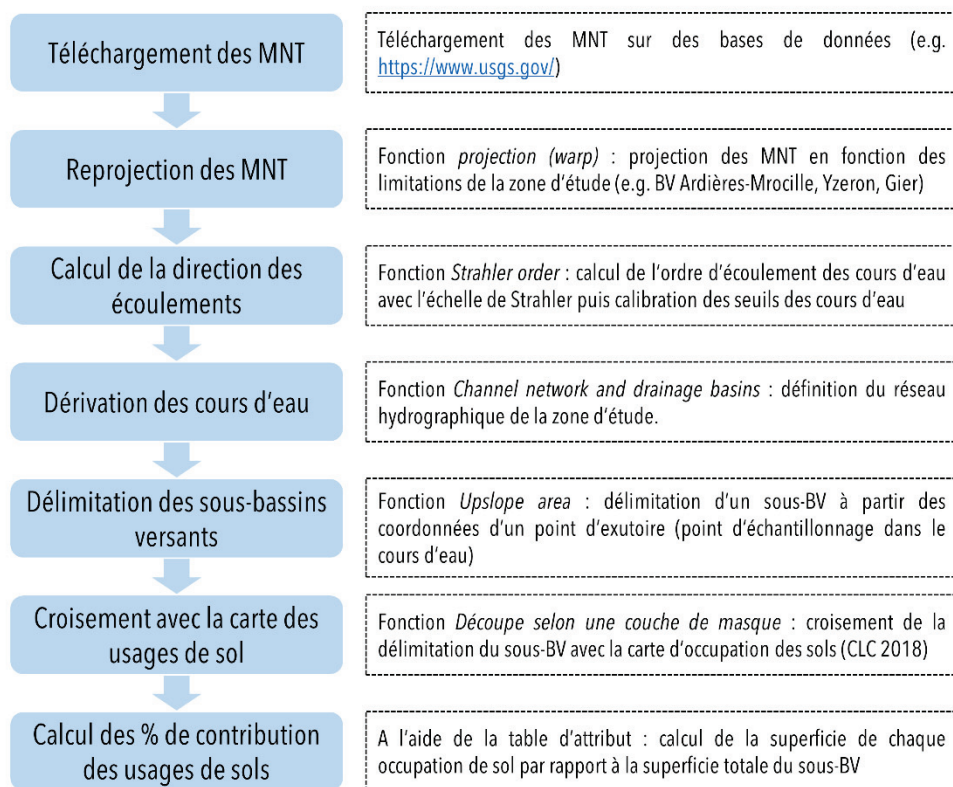


Figure 3-4. Principales étapes traitement sur le logiciel QGIS pour la sélection d'un sous-bassin versant et le calcul du pourcentage de contribution des usages et occupations des sols à l'exutoire.

La stratégie d'échantillonnage adoptée dans cette étude est basée sur le prélèvement de trois types d'occupation des sols, à savoir les forêts/prairies, l'agriculture et les retenues d'eau. Ces occupations et usages des sols ont été choisis car ils sont représentatifs des utilisations dominantes des sols dans les bassins versants étudiés.

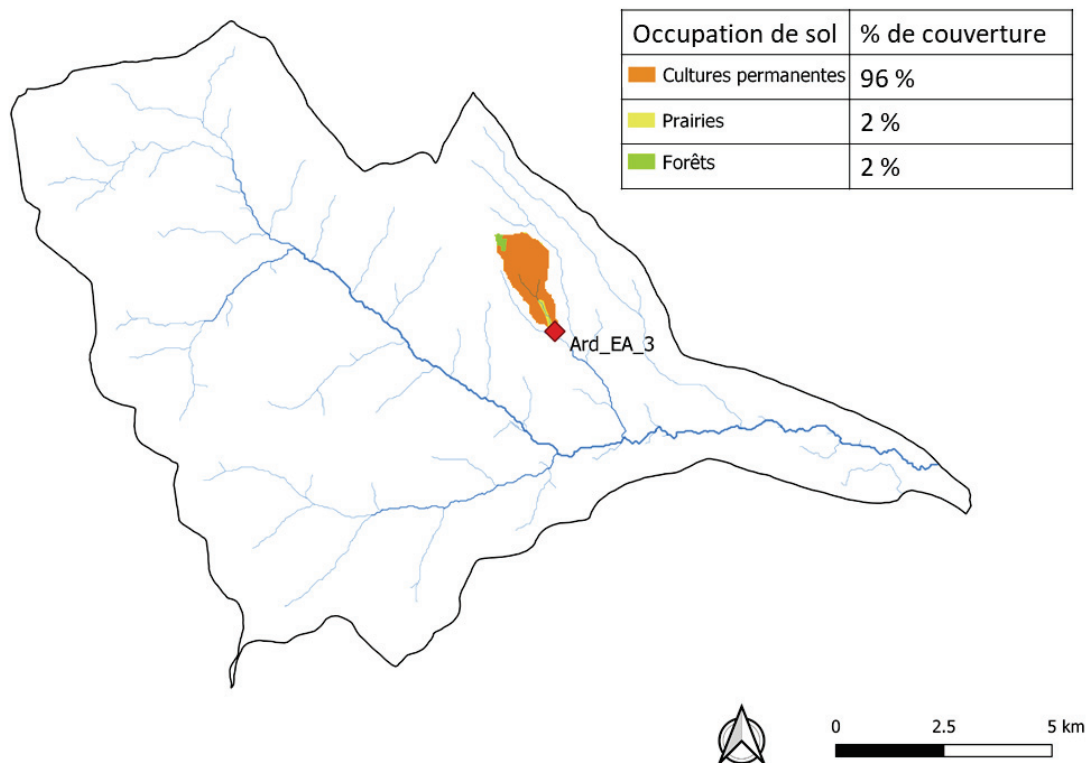


Figure 3-6. Exemple de l'identification d'un sous-bassin versant agricole (1.8 km²) sur le bassin de l'Ardières avec les pourcentages de contribution des usages de sols associés. La station de prélèvement de ce sous-bassin homogène est définie comme le point aval du cours d'eau drainant ce sous-bassin.

La catégorie « forêt/prairies » se réfère aux zones où la couverture végétale naturelle est dominée par des arbres de conifères et de feuillus ainsi que des prairies sans élevage. Cela a été vérifié à plusieurs reprises lors de campagnes de prospection de terrain. La catégorie « agriculture » se réfère aux zones où les terres sont principalement utilisées pour la production de cultures, l'élevage ou une combinaison des deux. L'usage de sol agricole était principalement axé sur la viticulture dans le BV de l'Ardières-Morcille et sur un mélange de cultures et d'élevage dans les BV de l'Yzeron et du Gier. Cette différenciation a également été faite pour tenir compte de la variabilité des pratiques agricoles dans les différents bassins versants, qui pourrait avoir des conséquences sur la qualité de l'eau. Enfin, les retenues d'eau

ont été incluses pour tenir compte des effets de ces milieux relativement stagnants, avec un temps de séjour élevé qui peut avoir un impact sur la qualité de la MOD.

Les échantillons d'eau ont été prélevés manuellement au niveau des cours d'eau, en aval des sous-bassins versants définis (partie 1.2.1). Les prélèvements ont été réalisés durant deux périodes en prenant en compte l'état hydrologique des cours d'eau : une première campagne d'échantillonnage a ainsi été réalisée en période de hautes eaux et une seconde en période de basses eaux. L'objectif de l'échantillonnage durant ces deux périodes était de mettre en évidence un apport terrigène pouvant marquer la MOD du cours d'eau récepteur en période de hautes eaux et un apport plus autochtone en période d'étiage. Ainsi, dans chaque bassin versant, 14 échantillons ont été prélevés en période de hautes eaux et 14 en période de basses eaux (Tableau 3-1).

Tableau 3-1. Récapitulatif des échantillons d'eau représentatifs de sources diffuses prélevés dans le cadre de la thèse (les coordonnées sont exprimées en degrés décimaux dans le système géodésique mondial WGS84).

Bassin versant	Type de source	Période de hautes eaux		Période de basses eaux		Coordonnées (Longitudes, Latitudes)
		Echantillon	Date	Echantillon	Date	
Ardières-Morcille	Forêt/prairie	Ard-FP-HW-01	26/01/2021	Ard-FP-LW-01	26/07/2021	46.18277, 4.63003
		Ard-FP-HW-02	26/01/2021	Ard-FP-LW-02	26/07/2021	46.19197, 4.54902
		Ard-FP-HW-03	26/01/2021	Ard-FP-LW-03	26/07/2021	46.18642, 4.52236
		Ard-FP-HW-04	26/01/2021	Ard-FP-LW-04	26/07/2021	46.15123, 4.5437
		Ard-FP-HW-05	26/01/2021	Ard-FP-LW-05	26/07/2021	46.19197, 4.54902
	Agriculture	Ard-AG-HW-01	26/01/2021	Ard-AG-LW-01	26/07/2021	46.13941, 4.67532
		Ard-AG-HW-02	26/01/2021	Ard-AG-LW-02	26/07/2021	46.14673, 4.65093
		Ard-AG-HW-03	26/01/2021	Ard-AG-LW-03	26/07/2021	46.14719, 4.64498
		Ard-AG-HW-04	26/01/2021	Ard-AG-LW-04	26/07/2021	46.13271, 4.6248
		Ard-AG-HW-05	26/01/2021	Ard-AG-LW-05	26/07/2021	46.17415, 4.64304
	Retenue d'eau	Ard-WR-HW-01	26/01/2021	Ard-WR-LW-01	26/07/2021	46.13808, 4.67656
		Ard-WR-HW-02	26/01/2021	Ard-WR-LW-02	26/07/2021	46.1447, 4.649
		Ard-WR-HW-03	26/01/2021	Ard-WR-LW-03	26/07/2021	46.13054, 4.62495
		Ard-WR-HW-04	26/01/2021	Ard-WR-LW-04	26/07/2021	46.14484, 4.59575
	Yzeron	Forêt/prairie	Yze-FP-HW-01	03/02/2021	Yze-FP-LW-01	19/07/2021
Yze-FP-HW-02			03/02/2021	Yze-FP-LW-02	19/07/2021	45.76718, 4.65037
Yze-FP-HW-03			03/02/2021	Yze-FP-LW-03	19/07/2021	45.76337, 4.64915
Yze-FP-HW-04			03/02/2021	Yze-FP-LW-04	19/07/2021	45.75889, 4.63642
Yze-FP-HW-05			03/02/2021	Yze-FP-LW-05	19/07/2021	45.75776, 4.63601
Agriculture		Yze-AG-HW-01	03/02/2021	Yze-AG-LW-01	19/07/2021	45.7603, 4.65874
		Yze-AG-HW-02	03/02/2021	Yze-AG-LW-02	19/07/2021	45.75699, 4.6672
		Yze-AG-HW-03	03/02/2021	Yze-AG-LW-03	19/07/2021	45.75496, 4.66622
		Yze-AG-HW-04	03/02/2021	Yze-AG-LW-04	19/07/2021	45.75402, 4.67858
		Yze-AG-HW-05	03/02/2021	Yze-AG-LW-05	19/07/2021	45.76796, 4.73738
Retenue d'eau		Yze-WR-HW-01	03/02/2021	Yze-WR-LW-01	19/07/2021	45.76188, 4.65353
		Yze-WR-HW-02	03/02/2021	Yze-WR-LW-02	19/07/2021	45.75724, 4.64233
		Yze-WR-HW-03	03/02/2021	Yze-WR-LW-03	19/07/2021	45.76012, 4.66443
		Yze-WR-HW-04	03/02/2021	Yze-WR-LW-04	19/07/2021	45.76719, 4.73824
Gier		Forêt/prairie	Gie-FP-HW-01	12/01/2021	Gie-FP-LW-01	22/07/2021
	Gie-FP-HW-02		12/01/2021	Gie-FP-LW-02	22/07/2021	45.49407, 4.69658
	Gie-FP-HW-03		12/01/2021	Gie-FP-LW-03	22/07/2021	45.49643, 4.6761
	Gie-FP-HW-04		12/01/2021	Gie-FP-LW-04	22/07/2021	45.50193, 4.65896
	Gie-FP-HW-05		12/01/2021	Gie-FP-LW-05	22/07/2021	45.50968, 4.6571
	Agriculture	Gie-AG-HW-01	12/01/2021	Gie-AG-LW-01	22/07/2021	45.50319, 4.6927
		Gie-AG-HW-02	12/01/2021	Gie-AG-LW-02	22/07/2021	45.51741, 4.7203
		Gie-AG-HW-03	12/01/2021	Gie-AG-LW-03	22/07/2021	45.55308, 4.72962
		Gie-AG-HW-04	12/01/2021	Gie-AG-LW-04	22/07/2021	45.55417, 4.73963
		Gie-AG-HW-05	12/01/2021	Gie-AG-LW-05	22/07/2021	45.55888, 4.75159
	Retenue d'eau	Gie-WR-HW-01	12/01/2021	Gie-WR-LW-01	22/07/2021	45.50227, 4.68307
		Gie-WR-HW-02	12/01/2021	Gie-WR-LW-02	22/07/2021	45.51891, 4.71924
		Gie-WR-HW-03	12/01/2021	Gie-WR-LW-03	22/07/2021	45.53317, 4.73804
		Gie-WR-HW-04	12/01/2021	Gie-WR-LW-04	22/07/2021	45.55531, 4.75065

3.1.3 Protocole de collecte des échantillons d'eau

Les échantillons d'eau ont été prélevés à l'aide de seringues en polypropylène d'une contenance de 50 mL puis filtrés avec des filtres en fluorure de polyvinylidène (PVDF) d'une porosité de 0,45 μm pour l'analyse du DOC et des propriétés optiques (UV-vis, EEM de fluorescence, HPSEC) et de 0,2 μm pour les analyses en LC-HRMS. Le filtrat a dans un premier temps été utilisé pour rincer trois fois les flacons avant de les remplir complètement. Les flacons utilisés étaient en verre borosilicaté de 250 mL pour les échantillons destinés aux analyses de DOC et propriétés optiques et en verre ambré de 25 mL pour les analyses en LC-HRMS. Tous les flacons ont été préalablement nettoyés par laveur automatique à 90°C puis calcinés à 500°C. L'ensemble des échantillons a ensuite été stocké dans une glacière (à $\pm 4^\circ\text{C}$) à l'abri de la lumière durant leur transport au laboratoire.

Une fois au laboratoire, les échantillons ont été répartis dans différents flacons en fonction des techniques d'analyses (Figure 3-6) :

- Les échantillons destinés aux analyses du DOC ont été acidifiés à pH 1 avec de l'acide chlorhydrique (HCl 32%) et stockés dans des tubes en verre de 25 mL, pré-calcinés à 500°C.
- Les échantillons destinés aux analyses optiques ont été stockés à 4°C dans des flacons en polypropylène de 40 mL.
- Les échantillons analysés en HRMS ont été aliquotés et congelés à -20 °C dans des vials en verre ambré de 2 mL.

L'ensemble des flaconnages et étapes de filtrations a été testé à l'aide de blancs d'eau ultrapure (résistivité à 18,2 M Ω .cm) sans détecter de contamination particulière.

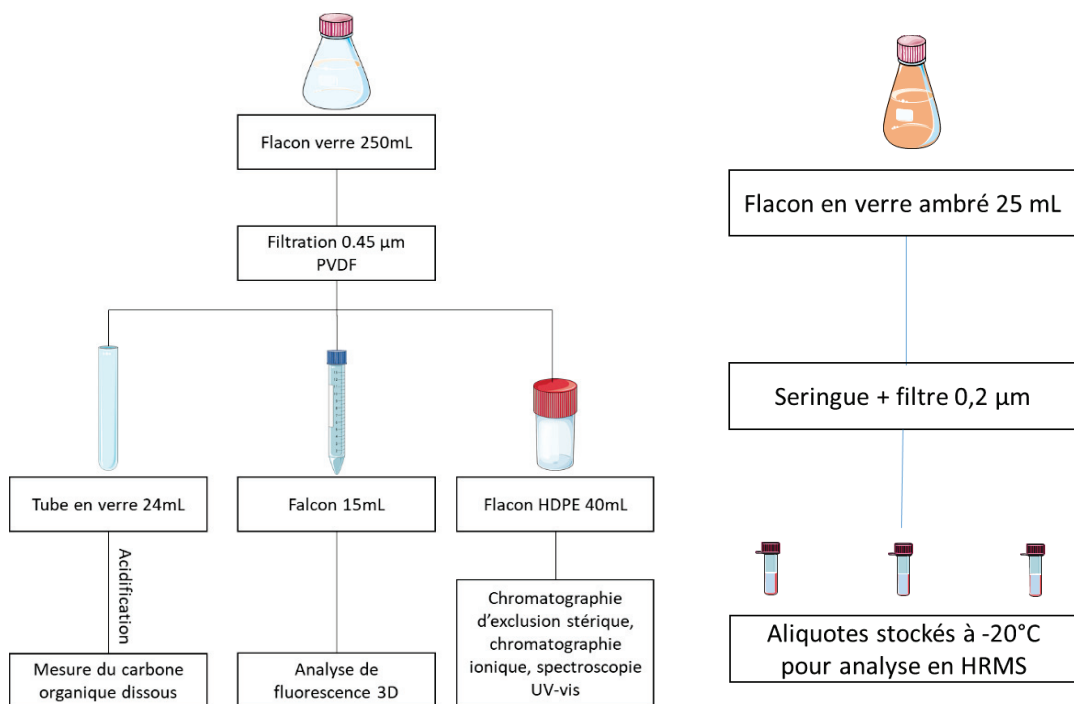


Figure 3-8. Conditionnement des échantillons d'eau avant analyse.

3.1.4 Méthode d'échantillonnage des sols et d'extraction de la MOD

3.1.4.1 Echantillonnage des sols

Les échantillons de sol ont été prélevés une seule fois dans le bassin de l'Ardières-Morcille pour une étude comparative entre les empreintes de la MOD des sols et celle des cours d'eau adjacents. En amont des points de prélèvement dans le cours d'eau du BV de l'Ardières-Morcilles, dix échantillons de sol (5 sols forestiers et 5 sols viticoles) ont été prélevés à l'aide d'une tarière dans des parcelles homogènes. Sur chaque parcelle, quatre points ont été choisis aléatoirement (périmètre de 50 m environ) pour échantillonner environ 1 kg de sol (Figure 3-7). Pour chaque point, la profondeur du sol prélevé était d'environ 20 cm. Chacun des 4 sous-échantillons a été placé dans un sac en polypropylène. Une fois les quatre points prélevés, les échantillons ont été regroupés en un seul échantillon composite qui a été mélangé dans un autre sac en polypropylène. L'échantillon composite a ensuite été transporté et conservé à 4°C à l'abri de la lumière jusqu'à préparation au laboratoire.

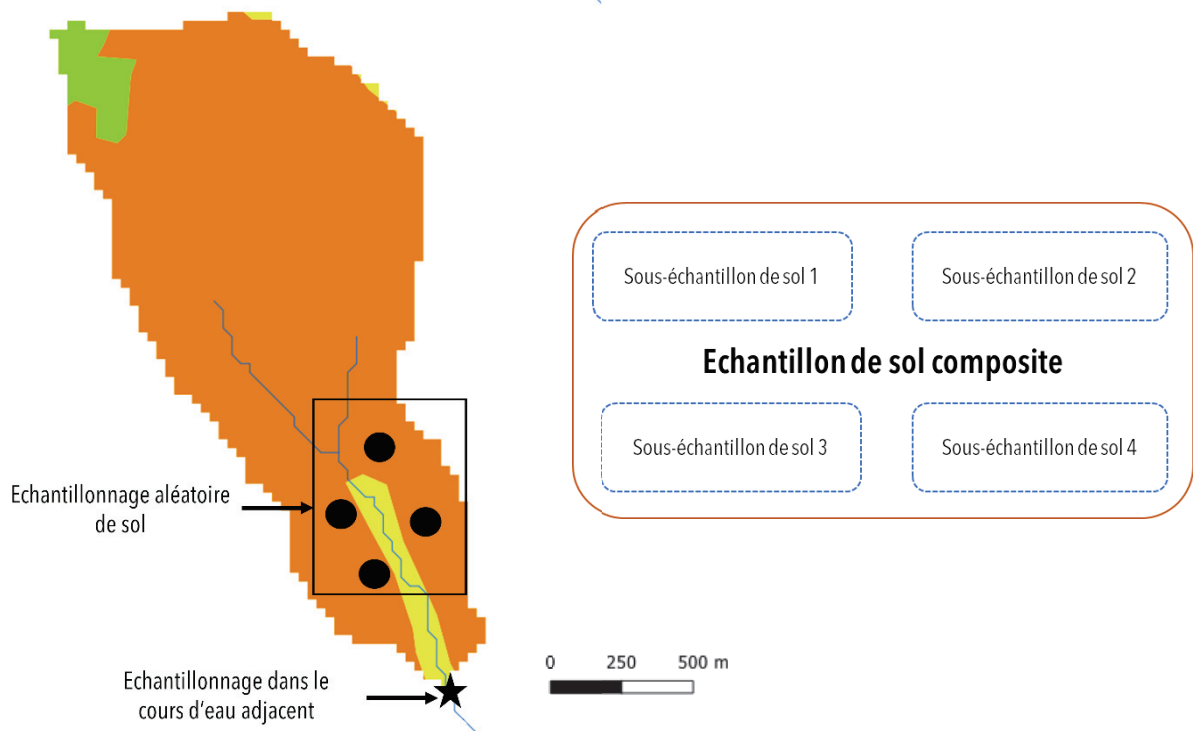


Figure 3-9. Illustration d'un échantillonnage de sol dans un sous-bassin versant à proximité en amont d'un point d'échantillonnage dans le cours

Au laboratoire, les échantillons de sols ont été tamisés à 2 mm avec un tamis en inox, puis placés dans des barquettes en aluminium (60 cm x 40 cm). Les échantillons ont été mis à sécher dans une étuve à 40°C pendant 96 h, puis stockés à 4°C dans des flacons en verre ambré de 1 L. Tous les échantillons de sol ont par la suite été caractérisés par analyse du carbone organique particulaire (Tableau 3-2), avant de procéder à l'extraction de la MOD.

Tableau 3-2. Coordonnées géographiques détaillées des zones d'échantillonnage des sols avec les concentrations particulières de carbone organique (Flash Dynamic Combustion) dans chaque sol (les coordonnées sont exprimées en degrés décimaux dans le système géodésique mondial WGS84).

Sol	Coordonnées (Longitudes, Latitudes)	POC = g/kg MS
Sol forestier 1	46.182, 4.63098	28
Sol forestier 2	46.19197, 4.54902	30
Sol forestier 3	46.18642, 4.52236	38
Sol forestier 4	46.16876, 4.52858	22
Sol forestier 5	46.15804, 4.58192	56
Sol viticole 1	46.13965, 4.67481	18
Sol viticole 2	46.14673, 4.65093	36
Sol viticole 3	46.14719, 4.64498	51
Sol viticole 4	46.13271, 4.6248	36
Sol viticole 5	46.17415, 4.64304	35

3.1.4.2 Préparation d'extraction et de dégradation de la MOD des sols

La MOD des sols a été extraite en agitant 20 g (poids sec) de sol dans 200 mL d'eau ultrapure enrichie avec 10 mM de CaCl₂ pour ajouter une force ionique minimale, ce qui permet de réaliser une extraction plus efficace (McDowell et al., 2006). L'agitation a été réalisée à 180 rotation par minute dans une cabine thermorégulatrice *Lovibond* à 20°C et en absence de lumière. Après 6 h d'agitation, les échantillons ont été centrifugés à 4500 g pendant 15 min, filtrés à 0,45 µm avec un filtre PVDF et un premier prélèvement T0 de 60 mL a été réalisé et conditionné comme mentionné dans la partie 1.2.2.

La dégradation de la MOD extraite des sols a été réalisée en ajoutant 35 µL d'inoculum et 1 mL de solution de nutriments aux extraits de sols après le prélèvement T0. Deux inoculums de sols forestiers et de sols viticoles ont été préparés 48 h avant le début des extractions de la

matière organique des sols. Pour obtenir une meilleure diversité microbienne, les deux inoculums ont été préparés à partir d'un mélange des 5 sols (Guigues et al., 2014). Pour l'inoculum « forêt », 5 g de chaque sol ont été ajoutés à 200 mL d'eau ultrapure. La même préparation a été réalisée pour l'inoculum « viticole ». Les deux préparations ont ensuite été incubées à 20°C et agitées pendant 48 h dans l'obscurité (Figure 3-8). Une solution de nutriment à base de $(\text{NH}_4)_2\text{SO}_4$ et de KH_2PO_4 (ratio molaire C:N:P:S:K de 5:1:1:1:1) a également été préparée avant de procéder à l'étude de la dégradation de la MOD extraite des sols (Guigues et al., 2014).

Des conditions de contrôle sans inoculum ni solution de nutriments ainsi que des contrôles sans sol (blancs d'extraction) ont également été réalisés. Tous les extraits de sols ont été incubés à 20°C dans l'obscurité, avec une agitation à 180 rotations par minute de façon continue pendant une période 28 jours. La période d'incubation de 28 jours a été choisie sur la base de travaux antérieurs qui ont étudié le temps nécessaire pour dégrader la fraction labile de la MOD à partir d'échantillons de sol (Guigues et al., 2014). Des sous-échantillons ont été prélevés à 3, 7 et 14 jours d'incubation pour suivre les changements dans la quantité et la qualité de la MOD (propriétés optiques).

La MOD des extraits de sols à T0, T3, T7, T14 et T28 a été caractérisée par une série de techniques d'analyses. Les propriétés globales, optiques et moléculaires ont été étudiées par le dosage du carbone organique dissous (DOC), la spectroscopie UV-Vis, l'EEM de fluorescence et la chromatographie d'exclusion stérique (HPSEC UV-fluorescence).

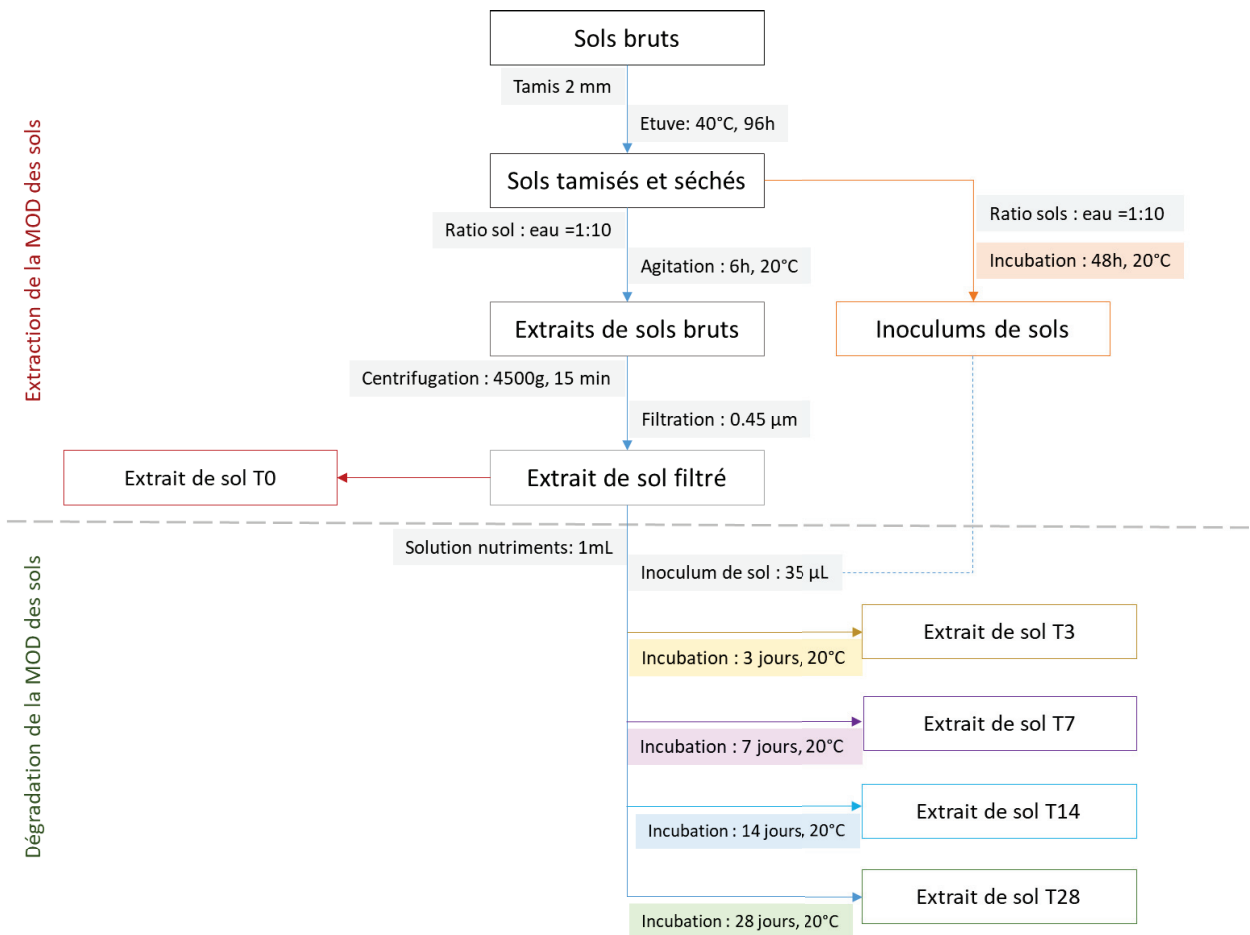


Figure 3-10. Schéma descriptif des étapes d'extraction et de dégradation de la MOD extraite des sols.

3.1.5 Prélèvement des effluents de stations de traitements d'eaux usées (STEU)

Les sources d'eaux usées de MOD ont été classées en fonction des rejets en amont des STEU. Deux STEU, de capacités de traitement respectives de 300 000 et 120 000 équivalent habitant (EH), utilisant un traitement par boues activées (BA) et deux STEU de capacité inférieure à 10 000 EH utilisant un traitement par filtres plantés de roseaux (PRF) ont été considérées. Les STEU avec traitement par boues activées (traitement intensif) utilisent un traitement biologique de type culture libre associé à des décanteurs longitudinaux. La STEU avec traitement PRF utilise quant à elle un traitement extensif basé sur la capacité de remédiation naturelle du sol qui est largement utilisé en France (Morvannou et al., 2015).

Tableau 3-3. Descriptif des échantillons prélevés dans les stations de traitement des eaux usées

Type de traitement	Capacité	Type d'eau	Echantillon	Date
Boues activées	Grosse collectivité (> 100 k EH)	Eau non traitée	BA-Vie-ESE-01	22/07/2020
			BA-Vie-ESE-02	28/10/2020
			BA-Vie-ESE-03	17/11/2020
			BA-Vie-ESE-04	29/11/2020
			BA-Vie-ESE-05	08/12/2020
			BA-Vie-ESE-06	15/12/2020
			BA-Vie-ESE-07	05/01/2021
			BA-Vie-ESE-08	19/01/2021
	Eau traitée	BA-Vie-SSE-01	22/07/2020	
		BA-Vie-SSE-02	28/10/2020	
		BA-Vie-SSE-03	17/11/2020	
		BA-Vie-SSE-04	29/11/2020	
		BA-Vie-SSE-05	08/12/2020	
		BA-Vie-SSE-06	15/12/2020	
		BA-Vie-SSE-07	05/01/2021	
		BA-Vie-SSE-08	19/01/2021	
	Grosse collectivité (> 300 k EH)	Eau non traitée	BA-Fey-ESE-01	04/03/2021
			BA-Fey-ESE-02	16/02/2021
BA-Fey-ESE-03			17/03/2021	
BA-Fey-ESE-04			26/03/2021	
BA-Fey-ESE-05			22/04/2021	
BA-Fey-ESE-06			16/06/2021	
Eau traitée		BA-Fey-SSE-01	04/03/2021	
		BA-Fey-SSE-02	16/02/2021	
		BA-Fey-SSE-03	17/03/2021	
		BA-Fey-SSE-04	26/03/2021	
		BA-Fey-SSE-05	22/04/2021	
		BA-Fey-SSE-06	16/06/2021	
Filtres plantés de roseaux	Petite collectivité (< 10 k EH)	Eau non traitée	FPR-CR-ESE-01	09/06/2021
			FPR-CR-ESE-02	10/06/2021
			FPR-CR-ESE-03	15/06/2021
			FPR-Due-ESE-01	22/04/2021
			FPR-Due-ESE-02	01/07/2021
			FPR-Due-ESE-03	22/04/2021
			FPR-Due-ESE-04	16/06/2021
			Eau traitée	FPR-CR-SSE-01
	FPR-CR-SSE-02	01/07/2021		
	FPR-CR-SSE-03	10/06/2021		
	FPR-CR-SSE-04	02/07/2021		
	FPR-CR-SSE-05	15/06/2021		
	FPR-Due-SSE-01	22/04/2021		
	FPR-Due-SSE-02	22/04/2021		
FPR-Due-SSE-03	16/06/2021			
FPR-Due-SSE-04	17/06/2021			

Des débordements du réseau d'assainissement observés notamment sur l'Yzeron nous ont amené à considérer la collecte d'eaux brutes non traitées afin d'obtenir la signature de la MOD associée à ce type de rejet. Ainsi, pour chaque STEU, une eau d'entrée non traitée et une eau de sortie traitée ont été prélevées sur 24 h à intervalles réguliers de 30 min à l'aide d'échantillonneurs automatiques. Les échantillons composites de 24 heures ont ensuite été stockés à 4°C dans des bouteilles en plastique en polyéthylène haute densité (HDPE) en vue de leur transport vers le laboratoire. Une fois au laboratoire, les échantillons ont été filtrés et stockés selon le protocole décrit dans la partie 1.3. Au total, 21 échantillons d'entrées de STEU et 22 échantillons de sorties de STEU ont été prélevés sur une période d'environ un an (Tableau 3-3).

3.1.6 Prélèvement des eaux de ruissellement urbain

La partie aval du bassin versant de l'Yzeron (au sud-est sur la Figure 3-2), caractérisée par une zone urbaine très dense, a été définie comme la zone d'échantillonnage préférentielle pour les sources de MOD liées au ruissellement urbain. Pour cela, neuf points de sorties du réseau d'eaux pluviales ont été définis comme des sources ponctuelles de rejet dans la rivière Yzeron. Les échantillons ont été prélevés lors d'un événement pluvieux (>15 mm sur une période de 3 heures). Les points caractérisés par un triangle blanc (Figure 3-2) ont été échantillonnés deux fois à des intervalles de 30 minutes pour évaluer la variation temporelle (Tableau 3-4). Les échantillons ont ensuite été filtrés et stockés à 4°C selon le protocole décrit dans la section 1.3.

Tableau 3-4. Descriptif des échantillons représentatifs des ruissellements urbains ((les coordonnées sont exprimées en degrés décimaux dans le système géodésique mondial WGS84).

Bassin versant	Date	Heure	Echantillon	Coordonnées (Longitude, Latitude)
Yzeron	14/02/2022	11:11	Lyo-RU-01	45.71764, 4.80922
		12:25	Lyo-RU-01-bis	45.71764, 4.80922
		11:17	Lyo-RU-02	45.71809, 4.80869
		11:21	Lyo-RU-03	45.71963, 4.80622
		11:35	Lyo-RU-04	45.72044, 4.80542
		11:43	Lyo-RU-05	45.72047, 4.80198
		11:50	Lyo-RU-06	45.72042, 4.80059
		12:07	Lyo-RU-07	45.72136, 4.79523
		13:20	Lyo-RU-07-bis	45.72136, 4.79523
		12:53	Lyo-RU-08	45.72646, 4.7801
		13:09	Lyo-RU-09	45.73303, 4.77605
		14:30	Lyo-RU-09-bis	45.73303, 4.77605

3.2 Stratégie analytique

3.2.1 Analyse du carbone organique dissous (DOC)

L'analyse du DOC a été réalisée à l'aide d'un analyseur de carbone organique total (TOC) *Analytik Jena multi N/C® 3100*, selon la norme NF EN 1484. Le principe de l'analyse repose sur une combustion catalytique à haute température. Les solutions d'étalonnage, de rinçage, de contrôles de qualité, ainsi que les échantillons, ont été introduits dans des flacons en verre pré-calcinés (à 500°C pendant 2 h) où ils ont été acidifiés à pH 1 avec de l'acide chlorhydrique (HCl 37% RPE). La limite de détection de l'appareil est de 0,2 mgC.L⁻¹ et l'incertitude analytique est de 12%. La précision des analyses (105%) a été systématiquement vérifiée en utilisant un étalon certifié d'hydrogénophthalate de potassium (5 mg.L⁻¹, fournisseur Fluka®) différent de celui utilisé pour la gamme d'étalonnage (0,2-25 mgL⁻¹, fournisseur Sigma-Aldrich®).

3.2.2 Spectroscopie UV-Visible

L'analyse par spectroscopie UV-Visible a été réalisée à l'aide d'un spectrophotomètre *Shimadzu UV 1900* avec une cellule en quartz de 1 cm de trajet (Hellma Analytics®). Les spectres d'absorption ont été acquis de 200 à 800 nm avec une résolution de 1 nm et une vitesse de balayage de 0.2 s.nm⁻¹ (Figure 3-9). Une correction de la ligne de base a été effectuée en utilisant un blanc d'eau ultrapure sur la même gamme de longueurs d'ondes. Des exemples de spectres UV-Visible obtenus pour des eaux brute et traitée d'une STEU ainsi qu'une eau naturelle prélevée en aval d'un sous-bassin versant forestier sont donnés dans la Figure 3-9. Une série d'indicateurs descriptifs de la nature de la MOD classiquement utilisés dans la bibliographie a été calculée à partir des spectres d'absorbance générés (Tableau 3-5).

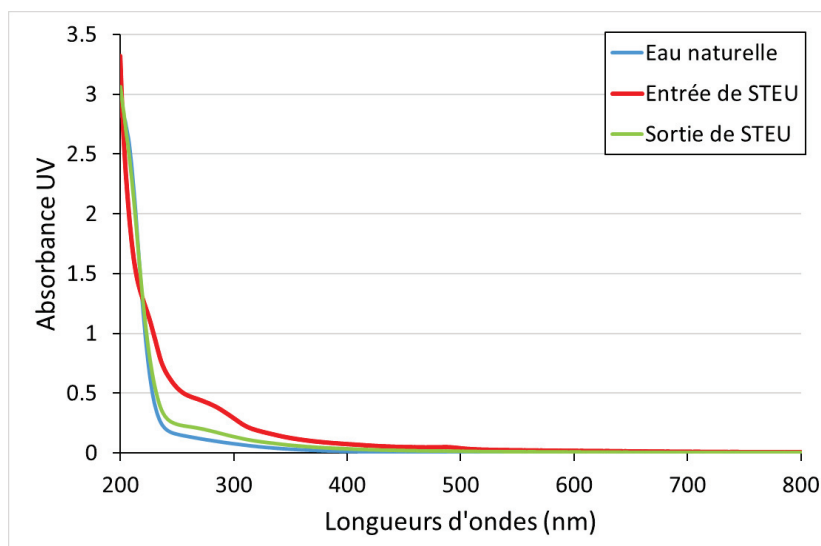


Figure 3-11. Exemple de spectres UV-Visible d'eaux de STEU brute et traitée ainsi que d'une eau naturelle prélevée en aval d'un sous-bassin versant forestier.

Tableau 3-5. Indicateurs UV-Visible calculés à partir des spectres d'absorbance

Indicateur	Calcul	Interprétation	Références
E2.E3	$\frac{Abs\ 250\ nm}{Abs\ 365\ nm}$	Inversement corrélé à la taille des molécules organiques	Helms et al., 2008
E2.E4	$\frac{Abs\ 254\ nm}{Abs\ 436\ nm}$	MOD autochtones vs MOD allochtones	Jaffé et al., 2004; Li and Hur, 2017
E3.E4	$\frac{Abs\ 300\ nm}{Abs\ 400\ nm}$	Inversement corrélé au degré d'humification	Claret et al., 2003; Li and Hur, 2017
S ₁ (275-295 nm)	$\alpha_\lambda = \alpha_{\lambda_0} e^{S(\lambda_0 - \lambda)}$ α_λ : coefficient absorption à une longueur d'onde λ λ_0 : longueur d'onde de référence S: pente exponentielle	Inversement corrélé à la teneur en carbone aromatique	Stedmon et al., 2003 Helms et al., 2008
S ₂ (350-400 nm)			
SR	$\frac{S_1}{S_2}$	Inversement corrélé au poids moléculaire de la MOD	Helms et al., 2008
Suva 254	$\frac{Abs\ 254\ nm}{[COD]}$	Corré à l'aromaticité de la MOD.	Weishaar et al., 2003

3.2.3 Spectroscopie d'excitation-émission de fluorescence

Les analyses des matrices d'excitation-émission (EEM) de fluorescence ont été réalisées à l'aide d'un spectrofluoromètre Aqualog (Horiba Scientific) dans une cuvette en quartz de 1 cm de chemin optique (Hellma Analytix®), thermostatée à 20 °C. L'instrument Aqualog était équipé d'une lampe à arc Xenon de 150 W et les longueurs d'onde d'excitation ont été balayées à l'aide d'un monochromateur à double grille de 240 à 800 nm à des intervalles de 5 nm, tandis que les spectres d'émission ont été obtenus à l'aide d'un détecteur couple de charge (CCD) à gain élevé à des intervalles d'environ 0,58 nm (1 pixel) pour les longueurs d'onde de 240 à 800 nm (Figure 3-10).

Tous les spectres EEM ont été corrigés des biais instrumentaux et normalisés par l'aire sous le pic Raman de l'eau, à 350 nm d'excitation, acquise quotidiennement. Les intensités de fluorescence sont donc exprimées en unités Raman (RU). Afin d'éviter les effets de filtre interne, tous les échantillons présentant une absorbance maximale > 0,1 ont été dilués. Les pics de diffusion Raman et Rayleigh ont été supprimés en soustrayant le spectre de l'eau ultrapure. Les spectres générés ont été traités et les indicateurs de fluorescence classiques calculés à l'aide du logiciel TreatEEM-V1 développé par Dario Omanović (Institut Ruđer Bošković, Zagreb, Croatie ; <https://sites.google.com/site/daromasoft/home/treateem>), comme résumé dans le Tableau 3-6.

Tableau 3-6. Indicateurs calculés à partir des matrices d'excitation-émission de fluorescence

Indicateur	Calcul	Interprétation	Références
HIX	$\frac{\sum IF_{435-480}}{\sum IF_{300-445}}$ à 255 nm IF : intensités de fluorescence	Corrélié à la teneur en carbone aromatique Indicateur de l'état d'humification de la MOD	Zsolnay et al., 1999
BIX	$\frac{IF_{380}}{IF_{430}}$ à 310 nm	Indicateur de la production autochtone récente de MOD due à l'activité biologique	Huguet et al., 2009
FI	$\frac{IF_{450}}{IF_{500}}$ à 370 nm	MOD terrigène vs MOD microbienne	Mcknight et al., 2001
YFI	$\frac{FEI_{350-400}}{FEI_{400-450}}$ à 280 nm FEI : intensités de fluorescence moyennes	Augmente la résolution de l'indicateur FI	Heo et al., 2016

L'analyse factorielle PARAFAC a été appliquée à deux *pools* d'échantillons différents en termes de composition de MOD. Le premier *pool* comprenait l'ensemble des échantillons des sources diffuses et de ruissellement urbain (n=96). Le deuxième *pool* de spectres EEM comprenait 43 échantillons, correspondant aux entrées et sorties de STEU. Le modèle PARAFAC a été exécuté à l'aide de la boîte à outils drEEM sous MatLab R2020a, pour 4 à 8 composantes avec des contraintes de non-négativité (Murphy et al., 2013). Cinq composantes ont été validées dans chaque *pool* de spectres EEM, sur la base d'une validation *half-split*. Trois des cinq composantes présentaient des longueurs d'onde d'excitation et d'émission (Ex-Em) similaires entre les deux *pools* et deux composantes étaient spécifiques à chaque *pool* d'échantillons. Toutes les composantes obtenues ont été comparées à celles précédemment rapportés en utilisant la base de données de fluorescence *OpenFluor*.

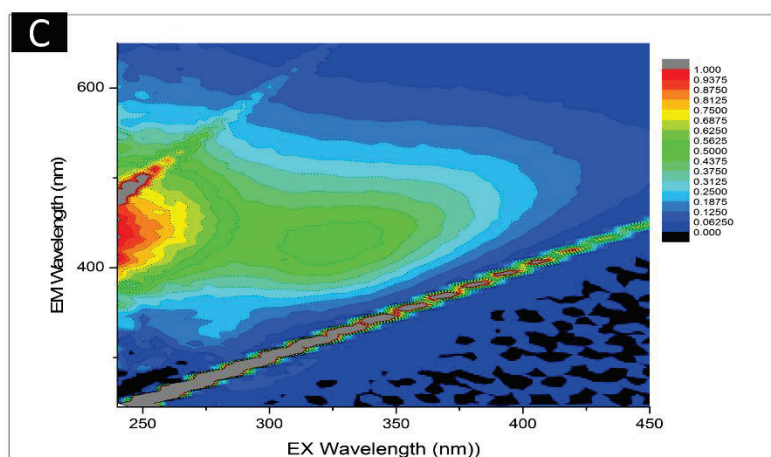
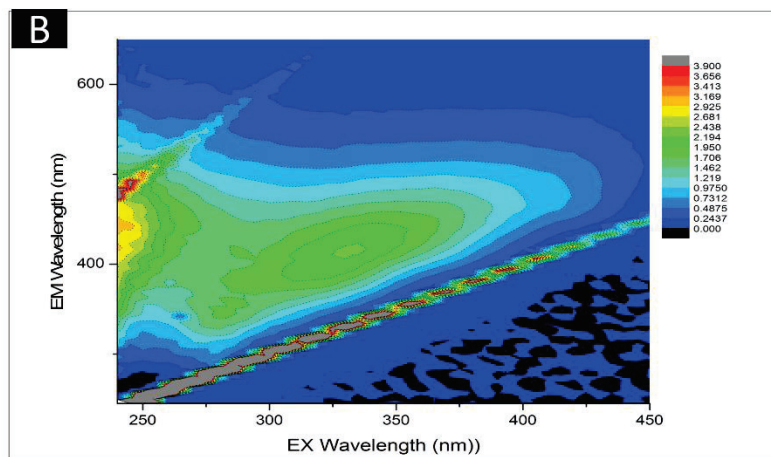
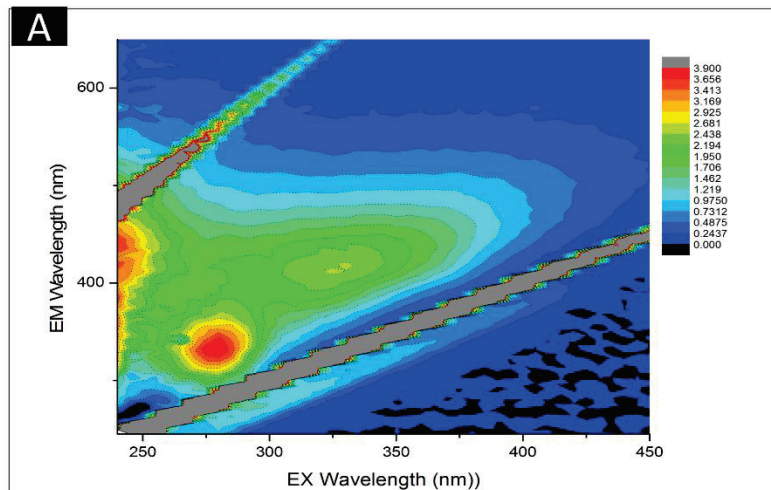


Figure 3-12. Exemples de matrices d'excitation-émission de fluorescence d'eaux de STEU brute et traitée ainsi que d'une eau naturelle prélevée en aval d'un sous-bassin versant forestier.

3.2.4 Chromatographie d'exclusion stérique HPSEC/UV-fluorescence

La chromatographie d'exclusion stérique à haute performance (HPSEC, modèle LC-2030C 3D Plus, Shimadzu) couplée à des détecteurs à réseau de photodiodes (UV-DAD, modèle PDA Deuterium D2 Lamp) et de fluorescence (modèle RF-20Axs) a été utilisée pour déterminer la distribution de la taille moléculaire de la MOD. Le volume d'injection était de 10 μL , avec un débit de 0,8 $\text{mL}\cdot\text{mn}^{-1}$ de phase mobile de tampon phosphate (NaCl , KH_2PO_4 et K_2HPO_4). Deux colonnes (Shodex OHpak SB-803 HQ et Shodex OHpak SB-802 HQ) couplées en série ont été utilisées pour la séparation chromatographique. La première colonne couvrait un domaine d'exclusion compris entre 1 kDa et 100 kDa et la seconde entre 0,2 kDa et 1 kDa. Des étalons analytiques PSS (polystyrene sulfonate) ont été utilisés comme références de poids moléculaires moyens (0,246 ; 1,6 ; 4,3 ; 6,8 ; 10 ; 17 ; 32 ; 77 kDa). D'autres étalons ont également été utilisés pour couvrir une plus large gamme de types de molécules, tels que les nitrates (0,063 kDa), l'albumine de sérum bovin BSA (66 kDa) et les étalons d'acide polyacrylique PAA (1 ; 2,6 et 3,6 kDa). Une courbe d'étalonnage du poids moléculaire (0,2 - 77 kDa) en fonction du volume d'élution a été modélisée à l'aide de ces étalons analytiques (Figure 3-11).

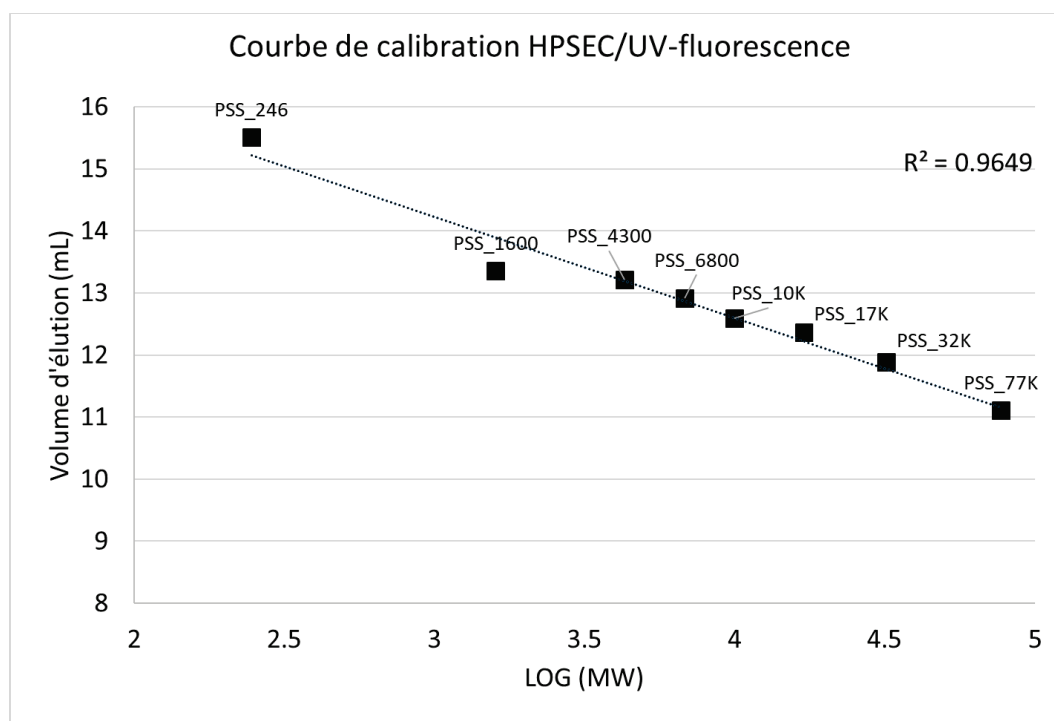


Figure 3-13. Courbe d'étalonnage (logarithme du poids moléculaire vs volume d'élution) des standards PSS utilisés pour évaluer le poids moléculaire (MW) moyen des échantillons de MOD

Le détecteur UV a permis l'acquisition de chromatogrammes sur une gamme d'absorbance comprise entre 200 et 450 nm, avec une incrémentation de 1,2 nm toutes les 300 ms. Dans le cadre de ce travail, seules les longueurs d'onde 254 nm et 280 nm ont été exploitées, correspondant respectivement aux pics d'absorption des composés organiques aromatiques et des composés type protéique (*e.g.* acides aminés aromatiques, tels que le tryptophane, la tyrosine et la phénylalanine) (Weishaar et al., 2003; Anthis and Clore, 2013). Concernant le détecteur de fluorescence, 3 couples de longueurs d'onde d'excitation-émission (Ex-Em) ont été retenus pour toutes les analyses. Les intensités de fluorescence aux longueurs d'ondes correspondant aux composants de type-humique (315-443 nm), aux composants de type-protéique (280-330 nm) et aux composants de type-acide aminé (278-354 nm) ont ainsi été mesurées toutes les 400 ms.

3.2.5 Chromatographie liquide couplée à la spectrométrie de masse haute résolution

3.2.5.1 Conception de la séquence d'analyse et mise en place des contrôles de qualité

Chaque séquence ou *batch* analytique a été conçu en tenant compte d'un ensemble d'échantillons de MOD, ainsi que d'une série de contrôles qualité (QC). Les échantillons de MOD étaient injectés en triplicats randomisés afin d'éliminer les sources de biais systématiques dans les données et de réduire le risque de contamination par rémanence. En plus des injections aléatoires d'échantillons en triplicat, d'autres QC ont été mis en place afin de disposer d'un suivi des dérives instrumentales ou des anomalies particulières au cours de l'analyse :

- QC *pool* (QCP) : correspond à un échantillon composite d'un volume égal de tous les échantillons (100 µL de chaque échantillon) injectés tout au long d'une séquence. Le QCP a été utilisé pour corriger les dérives d'intensité au cours d'une analyse.
- Blancs d'injection : eau ultrapure (résistivité à 18,2 MΩ.cm) injectée tout au long du *batch*. Les blancs d'injection ont été utilisés dans le post-traitement des données afin d'éliminer les pics dont l'intensité est proche du bruit de fond.

- Injection de traceurs deutérés : un mélange de diuron-d6 et de tebuconazole-d6 a été ajouté ($10 \mu\text{g}\cdot\text{L}^{-1}$) dans chaque échantillon, blanc d'injection et QCP pour évaluer la dérive du temps de rétention (RT) pendant l'acquisition (Tableau 3-7).
- QC mix : un mélange de 13 standards analytiques (Tableau 3-7) injectés en un triplicat randomisé dans chaque séquence d'analyse, afin d'évaluer la dérive intra-séquence et inter-séquence du RT et de la masse.

Tableau 3-7. Traceurs deutérés d'injection (en bleu) et composés du contrôle qualité des étalons en mélange (QCmix)

Nom du composé	Masse neutre (Da)	Masse neutre observée (Da)	m/z observée	RT attendu (min)	RT observé (min)
Aténolol	266.16304	266.163	267.1703	4.57	4.54
Azoxystrobine	403.11682	403.1162	404.1235	10.6	10.5
Bezafibrate	361.10809	361.1083	362.1156	9.64	9.57
Céliprolol	379.24711	379.2471	380.2544	6.44	6.51
Cétirizine	388.15537	388.1566	389.1639	8.26	8.35
Diuron d6	238.05468	238.0543	239.0616	8.99	8.97
Epoxiconazole	329.07312	329.0737	330.081	10.22	10.53
Erythromycine	733.46124	733.4608	734.4681	7.47	7.63
Metformine	129.10145	129.1016	130.1089	0.55	0.59
Montelukast	585.21044	585.2102	586.2174	13.44	13.02
Netilmicine	475.3006	475.299	476.3062	5.8	5.88
Ritonavir	720.31276	720.3138	721.3211	10.96	10.93
Roxythromycine	836.52457	836.5238	837.531	8.35	8.59
Simazine	201.07812	201.0777	202.085	7.78	7.74
Tébuconazole (TBZ)	307.14514	307.1456	308.1529	10.76	10.69
TBZ d6	313.1549	313.1658	314.19216	10.65	10.80
Telmisartan	514.23688	514.2383	515.2456	8.21	8.36
Tiaprïde	328.14568	328.1453	329.1526	5.09	5.05
Venlafaxine	277.20418	277.2041	278.2113	6.72	6.82

3.2.5.2 Analyses LC-HRMS non-ciblées

Les analyses LC-HRMS ont été mises en œuvre au moyen d'une chaîne de chromatographie liquide ultra-haute performance ACQUITY UPLC H-Class de Waters (Milford, MA, USA) couplée à un spectromètre de masse haute résolution QToF (Xevo G2-S Q-TOF ; Waters). Les échantillons ont été injectés sur une colonne de silice greffée C18, avec un mécanisme de chromatographie en phase inverse (HSS T3 C18, 10 mm x 2,1 mm, 1,8 μm , Waters). La phase mobile était composée de (A) eau ultra pure + 0,1 % d'acide formique et (B)

acétonitrile + 0,1 % d'acide formique. Les analyses ont été effectuées à un débit de 0,5 mL.min⁻¹ à 30°C en utilisant le gradient suivant : phase mobile initiale de 2 % de B pendant 2 min, suivie d'une augmentation jusqu'à 99 % de B pendant 13 min et d'un maintien pendant 2 min, puis un retour à la condition initiale de 2 % de B en 2 min avec un maintien de 2 min, ce qui donne une durée totale d'acquisition de 20 min. L'ionisation était en mode positif ESI+ avec les réglages suivants : une tension capillaire de 2,3 kV, une température de la source à 150 °C, un débit de gaz de désolvatation de 700 L/h, des débits d'infusion de l'échantillon et de *lockSpray* de 20 µLmin⁻¹. La gamme de masse d'acquisition, notée m/z (masse / charge, la charge étant égale à 1) s'étend de 100 à 1500 Da avec une fréquence de balayage de référence de 10 secondes.

Les données générées ont été traitées à l'aide de la plateforme Workflow4Metabolomics (W4M) (package xcms) (Giacomini et al., 2015). Les données brutes ont été préalablement converties au format mzXML avec le logiciel MSconvert de Proteowizard. Les paramètres de conversion étaient les suivants : peakPicking cwt snr = 0,1, peakSpace = 0,1 et msLevel = 1. Cette étape permet de transformer les données profiles en format centroïde. Une fois les données converties et chargées sur la plateforme W4M, un *workflow* automatisé a été appliqué pour produire la liste des *features* (Figure 3-12).

3.2.5.3 Etapes de génération de la liste de features

La première étape du *workflow* a consisté à extraire les masses et les pics chromatographiques. Pour ce faire, l'algorithme CentWave a été utilisé avec une tolérance maximale dans l'écart de m/z à 8,0 ppm et une largeur de pic fixée à 5,2 secondes. Le seuil signal/bruit a été fixé à 3 pour générer un maximum de pics. La seconde étape a consisté à regrouper les pics avec l'algorithme PeakDensity basé sur une largeur de bande de 10,0, une fraction minimale d'échantillons et un nombre minimal d'échantillons de 1. Avant de procéder à l'intégration des zones de pics, nous avons appliqué une étape de correction RT avec la fonction *retcor*. Pour ce faire, nous avons appliqué un alignement loess avec un degré de lissage de 0,5. Après la correction du RT, il a été nécessaire de répéter le regroupement des pics en réduisant la largeur de bande à 5,0. L'étape suivante a consisté à intégrer les zones de pics manquants à l'aide de la fonction fillPeaks. Le nombre de décimales pour les valeurs de masse a été fixé à 5. L'annotation des adduits et des isotopes a été réalisée avec l'algorithme

CAMERA (Kuhl et al., 2012), avec les paramètres suivants : une FWHM de 0,6, une erreur générale en ppm fixée à 5,0 et une erreur absolue en m/z fixée à 0,005. Pour l'annotation des isotopes, la charge maximale des ions a été fixée à 4 et le nombre d'isotopes attendus à 6. Les paramètres par défaut ont été utilisés pour l'annotation des adduits. La dernière étape du *workflow* a été la correction de la dérive instrumentale des intensités réalisée à partir des intensités des contrôles qualités composites (QCP) et en utilisant un modèle de régression de loess (span de 0,7).

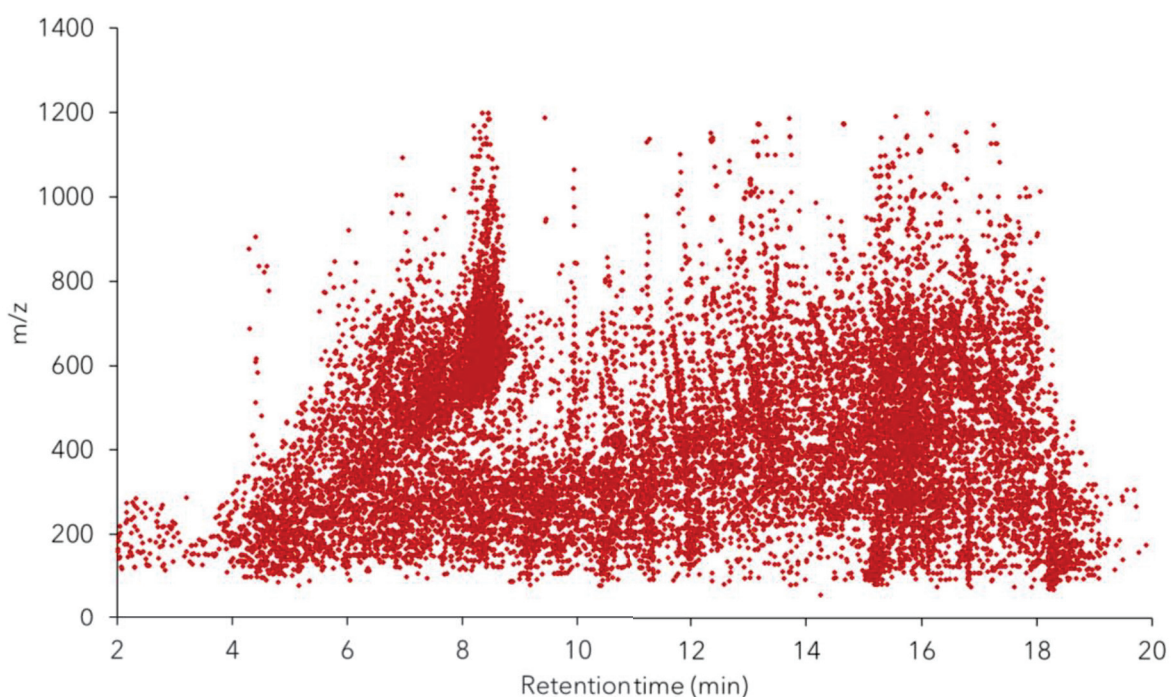


Figure 3-14. Illustration de la répartition des features (signaux caractérisés par un rapport masse sur charge m/z en Da et un temps de rétention RT en min) générés dans une séquence d'analyse en LC-HRMS non-ciblée.

3.3 Analyses statistiques des résultats

Les analyses statistiques ont été réalisées avec le logiciel *R* (version 4.0.2). Des analyses de variances (ANOVA) et des tests post-hoc Tuckey-HSD (packages *Stats* et *Rstatix*) ont été réalisés pour établir les différences entre séries d'individus (échantillons) pour un paramètre donné, par exemple pour déterminer si les valeurs de HIX sont significativement différentes d'une source à une autre. Des analyses multivariées (package *Factoextra*) avec des analyses de correspondance (AC) et des analyses en composantes principales (ACP) ont été réalisées pour visualiser la répartition des échantillons en fonction de l'ensemble des indicateurs

optiques et moléculaires. Des analyses supervisées, à savoir des analyses linéaires discriminantes (LDA) et des analyses discriminantes par les moindres carrés partiels (PLS-DA), ont été effectuées pour sélectionner les indicateurs les plus discriminants. Elles ont également été mobilisées dans l'analyse des données LC-HRMS pour la sélection des composés moléculaires les plus pertinents.

3.4 Références

- Anthis, N.J., Clore, G.M., 2013. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci. Publ. Protein Soc.* 22, 851–858. <https://doi.org/10.1002/pro.2253>
- Braud, I., Breil, P., Thollet, F., Lagouy, M., Branger, F., Jacqueminet, C., Kermadi, S., Michel, K., 2013. Evidence of the impact of urbanization on the hydrological regime of a medium-sized periurban catchment in France. *J. Hydrol.* 485, 5–23. <https://doi.org/10.1016/j.jhydrol.2012.04.049>
- Claret, F., Schäfer, T., Bauer, A., Buckau, G., 2003. Generation of humic and fulvic acid from Callovo-Oxfordian clay under high alkaline conditions. *Sci. Total Environ.* 317, 189–200. [https://doi.org/10.1016/S0048-9697\(03\)00337-1](https://doi.org/10.1016/S0048-9697(03)00337-1)
- Giacomini, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., Duperier, C., Tremblay-Franco, M., Martin, J.-F., Jacob, D., Goullitquer, S., Thévenot, E.A., Caron, C., 2015. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinforma. Oxf. Engl.* 31, 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Gnouma, R., n.d. Aide à la calibration d'un modèle hydrologique distribué au moyen d'une analyse des processus hydrologiques: application au bassin versant de l'Yzeron 449.
- Gouy, V., Liger, L., Ahrouch, S., Bonnineau, C., Carluet, N., Chaumot, A., Coquery, M., Dabrin, A., Margoum, C., Pesce, S., 2021. Ardières-Morcille in the Beaujolais, France: A research catchment dedicated to study of the transport and impacts of diffuse agricultural pollution in rivers. *Hydrol. Process.* 35, e14384. <https://doi.org/10.1002/hyp.14384>
- Guigue, J., Mathieu, O., Lévêque, J., Mounier, S., Laffont, R., Maron, P.-A., Navarro, N., Chateau-Smith, C., Suchet, P., Lucas, Y., 2014. A comparison of extraction procedures for water-extractable organic matter in soils. *Eur. J. Soil Sci.* 65, 520–530. <https://doi.org/10.1111/ejss.12156>
- Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnol. Oceanogr.* 53, 955–969. <https://doi.org/10.4319/lo.2008.53.3.0955>
- Heo, J., Yoon, Y., Kim, D.-H., Lee, H., Lee, D., Her, N., 2016. A new fluorescence index with a fluorescence excitation-emission matrix for dissolved organic matter (DOM) characterization. *Desalination Water Treat.* 57, 20270–20282. <https://doi.org/10.1080/19443994.2015.1110719>
- Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J.-M., Parlanti, E., 2009. Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Org. Geochem.* 40, 706–719. <https://doi.org/10.1016/j.orggeochem.2009.03.002>
- Jaffé, R., Boyer, J.N., Lu, X., Maie, N., Yang, C., Scully, N.M., Mock, S., 2004. Source characterization of dissolved organic matter in a subtropical mangrove-dominated estuary by fluorescence analysis. *Mar. Chem.* 84, 195–210. <https://doi.org/10.1016/j.marchem.2003.08.001>
- Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T.R., Neumann, S., 2012. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* 84, 283–289. <https://doi.org/10.1021/ac202450g>
- Labbas, M., 2017. Modélisation hydrologique de bassins versants périurbains et influence de l'occupation du sol et de la gestion des eaux pluviales: Application au bassin de l'Yzeron (130km²) 395.
- Li, P., Hur, J., 2017. Utilization of UV-Vis spectroscopy and related data analyses for dissolved organic matter (DOM) studies: A review. *Crit. Rev. Environ. Sci. Technol.* 47. <https://doi.org/10.1080/10643389.2017.1309186>

- McDowell, W.H., Zsolnay, A., Aitkenhead-Peterson, J.A., Gregorich, E.G., Jones, D.L., Jödemann, D., Kalbitz, K., Marschner, B., Schwesig, D., 2006. A comparison of methods to determine the biodegradable dissolved organic carbon from different terrestrial sources. *Soil Biol. Biochem.* 38, 1933–1942. <https://doi.org/10.1016/j.soilbio.2005.12.018>
- Morvannou, A., Forquet, N., Michel, S., Troesch, S., Molle, P., 2015. Treatment performances of French constructed wetlands: results from a database collected over the last 30 years. *Water Sci. Technol.* 71, 1333–1339. <https://doi.org/10.2166/wst.2015.089>
- Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Anal. Methods* 5, 6557–6566. <https://doi.org/10.1039/C3AY41160E>
- Peyrard, X., Liger, L., Guillemain, C., Gouy, V., 2016. A trench study to assess transfer of pesticides in subsurface lateral flow for a soil with contrasting texture on a sloping vineyard in Beaujolais. *Environ. Sci. Pollut. Res.* 23, 14–22. <https://doi.org/10.1007/s11356-015-4917-5>
- Rabiet, M., Coquery, M., Carlier, N., Gahou, J., Gouy, V., 2015. Transfer of metal(loid)s in a small vineyard catchment: contribution of dissolved and particulate fractions in river for contrasted hydrological conditions. *Environ. Sci. Pollut. Res.* 22, 19224–19239. <https://doi.org/10.1007/s11356-015-5079-1>
- Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* 82, 239–254. [https://doi.org/10.1016/S0304-4203\(03\)00072-0](https://doi.org/10.1016/S0304-4203(03)00072-0)
- Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Fram, M.S., Fujii, R., Mopper, K., 2003. Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. *Environ. Sci. Technol.* 37, 4702–4708. <https://doi.org/10.1021/es030360x>
- Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere* 38, 45–50. [https://doi.org/10.1016/S0045-6535\(98\)00166-0](https://doi.org/10.1016/S0045-6535(98)00166-0)

Chapitre 4.

**Mise en place de nouveaux indicateurs
HPSEC/UV-fluorescence et LC-HRMS de
caractérisation de la MOD.**

Chapitre 4. Mise en place de nouveaux indicateurs HPSEC/UV-fluorescence et LC-HRMS de caractérisation de la MOD.

Avant-propos

Parmi le panel de techniques utilisées dans le cadre de cette thèse, la chromatographie d'exclusion stérique HPSEC couplée à des détections UV et fluorescence d'une part, et la spectrométrie de masse haute résolution couplée à la chromatographie en phase liquide LC-HRMS d'autre part sont des techniques relativement moins utilisées que les techniques classiques classique d'UV-Visible et EEM de fluorescence pour l'étude des sources de MOD. Pourtant, elles présentent des avantages majeurs en apportant une information complémentaire sur la structure et la composition moléculaire de la MOD. Le développement de nouveaux indicateurs à partir de ces deux techniques peut donc être un outil précieux pour mieux décrire la composition de la MOD dans les écosystèmes aquatiques. Ces nouveaux indicateurs, basés sur la répartition des composants de la MOD en poids moléculaire, masse et polarité, peuvent ainsi permettre d'améliorer les connaissances sur les sources, les processus de transformation et le devenir de la MOD dans les cours d'eau.

L'étude de la littérature a toutefois montré l'existence de verrous liés au traitement des données HPSEC et LC-HRMS, quand bien même il s'agit de deux techniques de plus en plus utilisées pour décrire la composition et les propriétés physicochimiques de la MOD. En effet, la complexité des données produites ainsi que la multitude d'outils possibles pour le retraitement, dont le développement est plus ou moins avancé, peuvent parfois compliquer l'exploitation de ces données. Dans ce contexte, ce chapitre a pour but de présenter les étapes suivies pour le traitement de données HPSEC/UV-fluorescence et LC-HRMS ainsi que les indicateurs pertinents pour l'étude des sources de MOD. Ce chapitre s'articule autour de deux parties : la première partie concerne le traitement de la donnée HPSEC et la présentation des indicateurs LMW, MMW et HMW, incluant une comparaison avec les indicateurs classiques de distribution de poids moléculaire (M_n , M_w et P). La deuxième partie aborde le traitement et l'interprétation de la donnée LC-HRMS avec : i) le *workflow* de retraitement de données (génération de la liste préliminaire de *features*) et les étapes de consolidation de la donnée (sous la forme d'un article sur l'application de filtres qualités et la génération de la liste finale

de *features*) ; ii) la présentation des indicateurs globaux LC-HRMS, la confrontation avec des indicateurs optiques et l'identification de marqueurs moléculaires spécifiques.

4.1 Proposition de nouveaux indicateurs HPSEC/UV-fluorescence

4.1.1 Indicateurs classiques de distribution de poids moléculaire : M_n , M_w et ρ

A l'origine, la distribution du poids moléculaire (MWD ; *molecular weight distribution*) est un paramètre utilisé pour mesurer la gamme de masses présentes dans un échantillon de polymères, fournissant ainsi une indication de la taille et de la forme des molécules de polymères, ce qui permet de comprendre leurs propriétés physico-chimiques. En utilisant des standards analytiques, il est alors possible d'établir une courbe de calibration basée sur le temps d'éluion des molécules et de calculer le MWD (Chapitre 3, Partie 3.2.4). Ce paramètre MWD est défini à travers des moyennes statistiques de distribution des composés en nombre (M_n), en masse (M_w) et leur distribution ou polydispersité (ρ). Plus précisément, le paramètre M_n reflète le poids moléculaire moyen en nombre de composés d'un échantillon, calculé en divisant la masse totale d'un échantillon par le nombre total de molécules de l'échantillon (Eq 4.1). Le M_w représente quant à lui la masse moléculaire moyenne des composés d'un échantillon, calculée en divisant la masse totale de l'échantillon par le nombre total de molécules dans l'échantillon, pondérée par la taille des molécules (Eq 4.2). Enfin, la polydispersité ρ représente la distribution des poids moléculaires d'un échantillon. Elle est obtenue en calculant le rapport M_w/M_n (Eq 4.3). Dans la caractérisation de polymères, une valeur de ρ égale à 1 signifie que l'échantillon de polymère est constitué d'un seul et unique composant. On parle alors d'échantillon monodispersé.

$$M_n = \frac{\sum_{i=1}^n h_i}{\sum_{i=1}^n \left(\frac{h_i}{M_i}\right)} \quad \text{Eq 4.1}$$

$$M_w = \frac{\sum_{i=1}^n h_i M_i}{\sum_{i=1}^n (h_i)} \quad \text{Eq 4.2}$$

$$\rho = \frac{M_w}{M_n} \quad \text{Eq 4.3}$$

Où h_i est la hauteur de la courbe HPSEC de l'échantillon élué au volume "i" et M_i est le poids moléculaire de la "i" ème fraction d'un échantillon au volume élué donné "i".

Dans la littérature, les indicateurs Mn et Mw ont par la suite été appliqués pour caractériser des échantillons de matière organique naturelle issue de milieux différents (*e.g.* lacs, rivières, eaux souterraines, effluents de STEU) (Her et al., 2003; Lanxiu et al., 2004; Park, 2009; Huang et al., 2021). Les résultats de ces études ont mis en évidence des différences entre les MOD issues de ces milieux, et que des facteurs environnementaux tels que la photodégradation et l'activité biologique modifiaient le poids moléculaire de la MOD. Ces études ont également montré que la fraction de haut poids moléculaire de la MOD des eaux souterraines était principalement constituée de substances de type polysaccharide (dans la gamme des Mw de plus de 10 kDa) alors que la MOD des rivières et des effluents de STEU était constituée d'un mélange de substances de type polysaccharide et protéine. De plus, les fractions de la MOD ayant un poids moléculaire compris entre 5 et 10 kDa présentent un niveau d'aromaticité plus élevé et ont été associés aux composés de type humique retrouvés dans les eaux naturelles. Enfin, une fraction de la MOD avec un faible Mw apparent (< 600 Da) dans les échantillons prélevés dans les lacs a été associée à composés de type protéine labile. D'autres études se sont basées sur ces indicateurs pour étudier l'efficacité de traitement des eaux résiduaires et des eaux potables (Huang et al., 2016; Brezinski and Gorczyca, 2019).

Cependant, des études récentes ont montré que les chromatogrammes HPSEC peuvent être décomposés en différentes fractions pour fournir davantage d'informations sur le continuum de taille de la MOD. Cette méthode permet d'étudier les propriétés physico-chimiques de la MOD dans des fractions de tailles spécifiques et d'acquérir une compréhension plus approfondie de leur composition et de leur réactivité (Chon et al., 2017; Xue et al., 2022). De plus, dans un contexte d'identification de marqueurs de sources de MOD, l'utilisation du paramètre traditionnel MWD pourrait présenter des limitations, car l'analyse de la MOD par HPSEC montre que les chromatogrammes sont la plupart du temps complexes (Figure 4-1). L'absence de pic bien défini montre la présence d'un continuum de molécules ayant des poids moléculaires variés. Ce continuum peut introduire des biais dans le calcul et rendre difficilement interprétables ces indicateurs Mw , Mn et p .

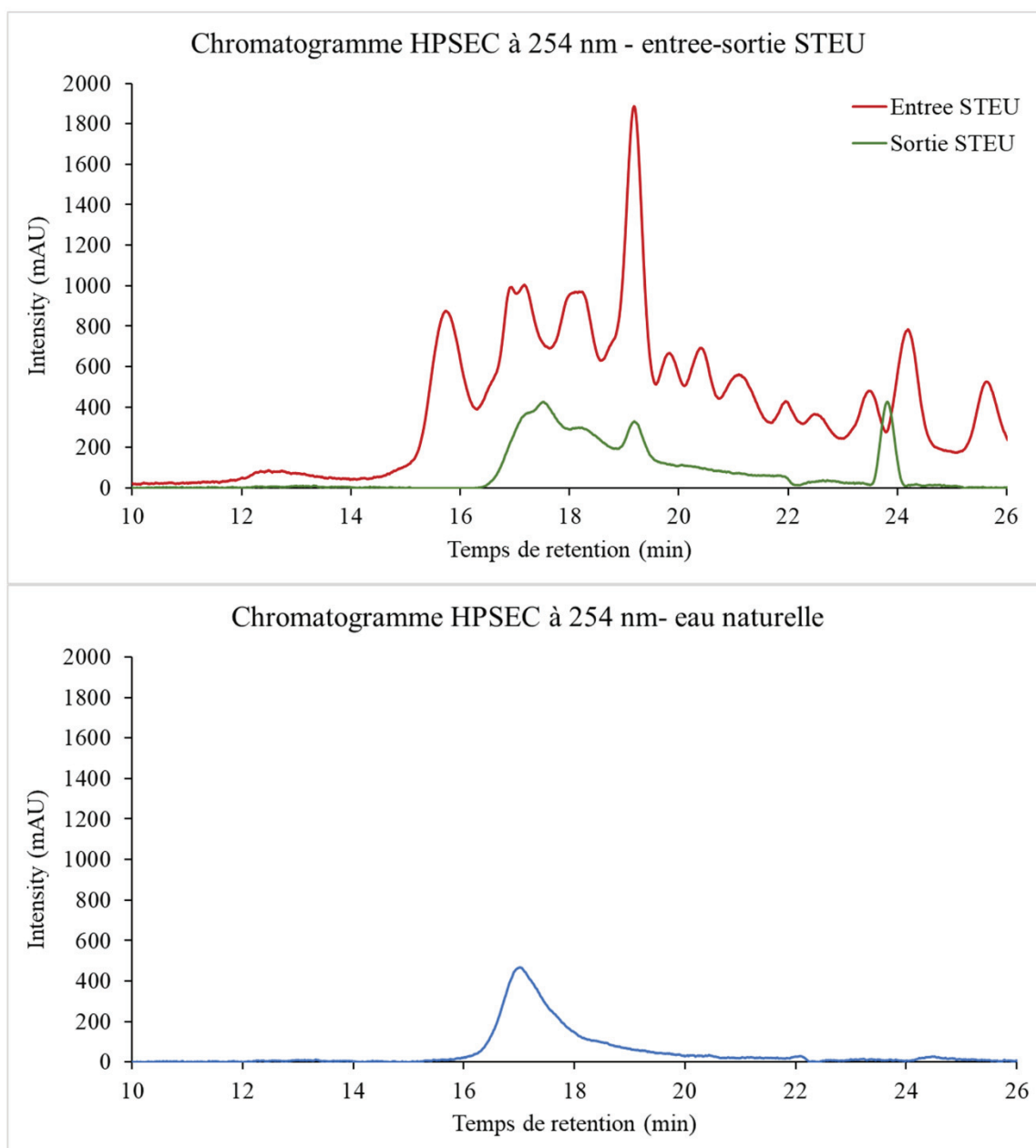


Figure 4-1. Exemple de 2 chromatogrammes HPSEC-UV à 254 nm : haut : une eau de STEU brute et traitée, bas : une d'une eau naturelle prélevée en aval d'un bassin versant forestier.

4.1.2 Nouveaux indicateurs basés sur la décomposition des chromatogrammes HPSEC-UV et fluorescence : LMW, MMW et HMW

4.1.2.1 Définition des indicateurs LMW, MMW et HMW

La décomposition des chromatogrammes HPSEC permet de calculer les contributions relatives des fractions des composés de petit, moyen et haut poids moléculaire dans chaque échantillon. Les détections UV et de fluorescence ajoutent une autre dimension qui permet

d'associer des types de molécules à cette distribution en poids moléculaire. Ainsi, le calcul des contributions peut être envisagé sur un large éventail de longueurs d'ondes et ainsi permettre de mettre en place de nouveaux indicateurs pertinents dans la caractérisation des échantillons de MOD.

Dans le cadre de cette thèse, la décomposition des chromatogrammes a été réalisée avec une décomposition gaussienne préalable, afin d'identifier les principales composantes contenues dans un échantillon (Figure 4-2). Il s'agit d'une technique souvent utilisée pour analyser les chromatogrammes complexes. Elle repose sur l'idée que les chromatogrammes peuvent être décomposés en composants individuels sous la forme de pics qui ont une forme, une largeur et un temps d'élution qui leur sont spécifiques.

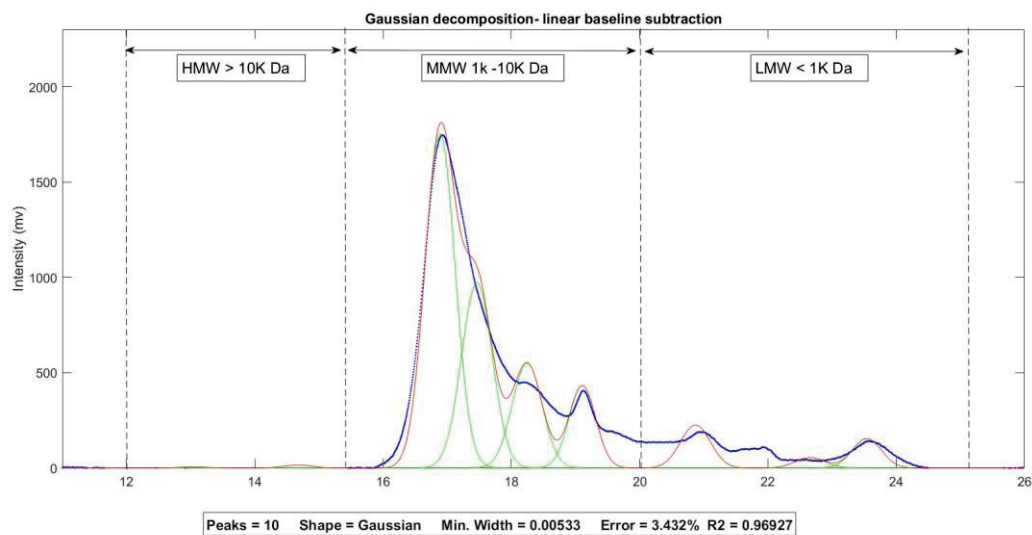


Figure 4-2. Illustration d'une décomposition gaussienne d'un chromatogramme HPSEC

Cette technique est utilisée pour l'analyse d'échantillons à composantes multiples, car elle permet de déterminer les concentrations relatives de chaque composante et d'établir une compréhension globale et complète de l'échantillon. Ces composantes peuvent être associés aux groupes chimiques des colloïdes organométalliques et des résidus biologiques, des composés type humiques, des acides de faible poids moléculaire et des composés aromatiques azotés (Chow et al., 2008). Un modèle de décomposition gaussienne a ainsi été appliqué sur les chromatogrammes en utilisant la *toolbox Peakfit.m* (O'Haver, 2020) sur MatLab R2020a. L'équation du modèle est la suivante :

$$A(x; \mu; \sigma; \varphi) = \sum_{i=0}^{nC} \varphi_i e^{-\frac{(x-\mu_i)^2}{2\sigma^2 i}} + \varepsilon \quad \text{Eq 4.4}$$

Où nC est le nombre de composantes gaussiennes, ε est le résidu qui représente la variabilité non prise en compte par le modèle, x est la surface du pic, μ est le paramètre de position du pic (volume d'élution), σ est la largeur du pic et φ est la proportion de chaque distribution gaussienne ajustée.

Les aires des composantes identifiées ont ensuite été sommées pour calculer le *total sec absorbance* (TSA, Eq 4.5) (Asmala et al., 2021) pour les chromatogrammes UV à 254 et 280 nm et le *total sec fluorescence* (TSF, Eq 4.6) pour les chromatogrammes de fluorescence correspondant aux couples d'excitation-émission 315-443 nm, 280-330 nm et 278-354 nm. Il est en effet courant d'utiliser les longueurs d'onde UV de 254 nm à 280 nm pour caractériser la MOD, car elles correspondent aux maxima d'absorption de deux groupes fonctionnels organiques importants : les doubles liaisons carbone-carbone aromatiques (C=C) et les liaisons peptidiques (C-N). Quant aux longueurs d'ondes d'excitation et d'émission de fluorescence choisies, ce sont les plus classiquement utilisées pour étudier les composés de type humique, de type protéine et de type acide aminé, qui sont les principaux fluorophores de la MOD naturelle.

$$TSA = \sum_{i=1}^{nC} A_C \text{ à } Abs \quad \text{Eq 4.5}$$

$$TSF = \sum_{i=1}^{nC} A_C Ex - Em \quad \text{Eq 4.6}$$

Où C est la composante identifiée par le modèle, A l'aire sous la courbe de la composante, Abs la longueur d'onde d'absorbance UV du chromatogramme, $Ex-Em$ les longueurs d'ondes d'excitation et d'émission de fluorescence du chromatogramme.

Les TSA et TSF ont permis de calculer une série d'indicateurs basés sur les pourcentages de contribution de trois fractions : petites molécules ou *low molecular weight* (LMW), moyennes molécules ou *medium molecular weight* (MMW) et grosses molécules ou *high molecular weight* (HMW). L'indicateur LMW correspond à la somme des aires des composantes dont le

poinds moléculaire est inférieur à 1 kDa rapporté au TSA ou TSF (Eq 4.7). L'indicateur MMW correspond à la somme des aires des composantes compris entre 1 kDa et 10 kDa rapporté au TSA ou TSF (Eq 4.8). L'indicateur HMW correspond à la somme des aires des composantes supérieures à 10 kDa rapporté au TSA ou TSF (Eq 4.9). Le choix des gammes de tailles est principalement basé sur les conditions analytiques choisies, à savoir l'utilisation de deux colonnes dont le domaine d'exclusion est très large entre 0,1 et 100 kDa, et également sur les chromatogrammes des différents échantillons de MOD (*e.g.* STEU, MOD naturelle, etc.)

$$LMW_{TSA} = \frac{\sum_{i=1}^{nC} A_C < 1 \text{ k Da}}{TSA} \quad LMW_{TSF} = \frac{\sum_{i=1}^{nC} A_C < 1 \text{ k Da}}{TSF} \quad \text{Eq 4.7}$$

$$MMW_{TSA} = \frac{\sum_{i=1}^{nC} A_C 1 - 10 \text{ k Da}}{TSA} \quad MMW_{TSF} = \frac{\sum_{i=1}^{nC} A_C 1 - 10 \text{ k Da}}{TSF} \quad \text{Eq 4.8}$$

$$HMW_{TSA} = \frac{\sum_{i=1}^{nC} A_C > 10 \text{ k Da}}{TSA} \quad HMW_{TSF} = \frac{\sum_{i=1}^{nC} A_C > 10 \text{ k Da}}{TSF} \quad \text{Eq 4.9}$$

Où C est la composante identifiée par le modèle, A l'aire sous la courbe de la composante, TSA le *total sec absorbance*, TSF le *total sec fluorescence*.

4.1.2.2 Test du modèle de décomposition gaussienne

Afin d'évaluer l'efficacité du modèle de décomposition gaussienne, un test de comparaison a été réalisé sur les standards analytiques PSS (Chapitre 3, Partie 3.2.4). Dans un premier temps, les standards ont été injectés séparément afin d'obtenir le pic associé à chaque PSS (Figure 4-3. A). Dans un second temps, un mélange des standards PSS a été préparé puis analysé suivant les mêmes conditions analytiques. Le chromatogramme du mélange (Figure 4-3. B) a ensuite été traité en utilisant la *tool box peakfit* afin de procéder à sa décomposition. Les pics obtenus par le modèle (Figure 4-3. C) ont ensuite été comparés avec les pics d'origine des PSS.

L'allure du chromatogramme du mélange de PSS (Figure 4-3. B) représente bien le concept de continuum de poids moléculaire caractéristique des échantillons de matière organique. La décomposition du chromatogramme (PSS mix) a permis d'identifier 12 composantes avec une erreur quadratique moyenne (RMSE) inférieure à 5 %. Cette RMSE est un des indicateurs de performance d'un modèle de régression le plus souvent utilisé. Il s'agit de la différence moyenne entre les valeurs prédites et les valeurs réelles. Elle évalue la précision avec laquelle le modèle peut prédire la valeur de référence. Parmi les composantes identifiées, 7 ont été identifiées dans le domaine d'exclusion de l'appareil (entre 10 et 25 min). Le Tableau 4-1 résume les temps d'éluions des composantes en comparaison avec ceux des PSS injectés séparément.

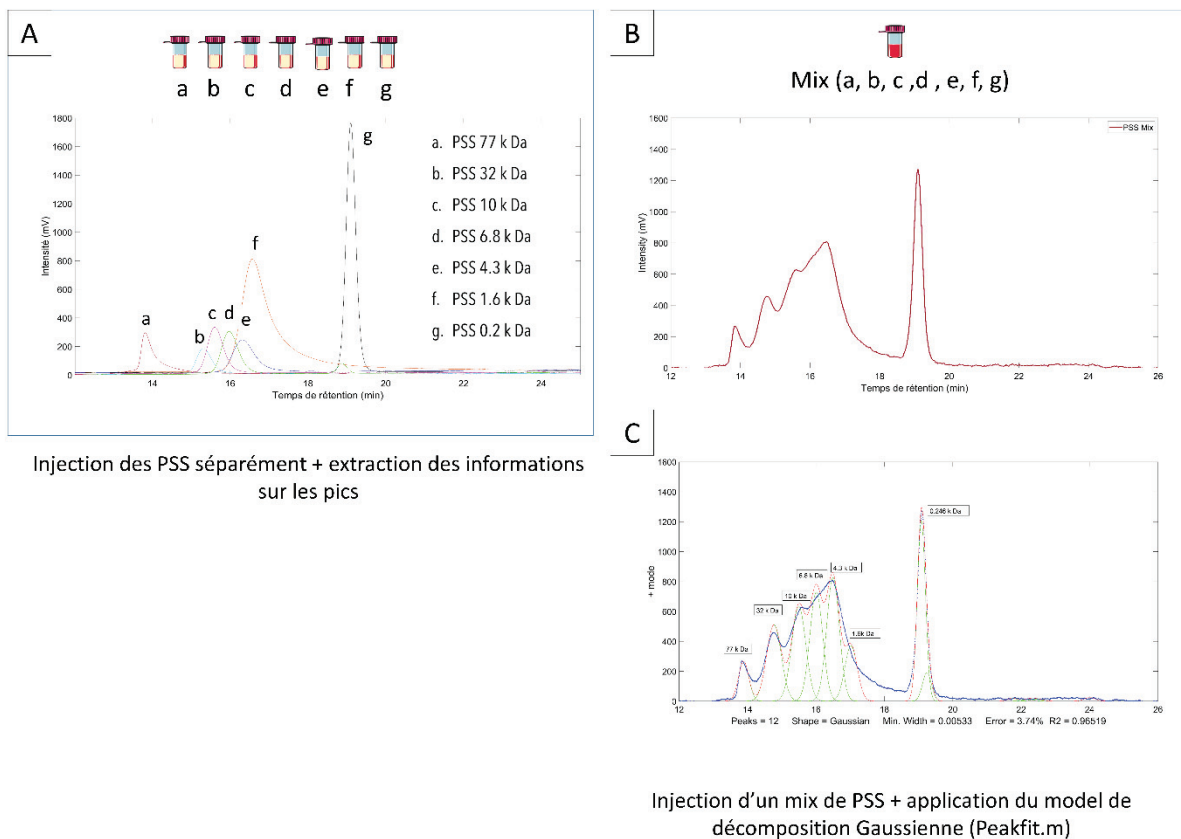


Figure 4-3. Test de comparaison entre injection de PSS séparément et en mélange avec application du modèle de décomposition gaussienne

Tableau 4-1. Temps de rétention (RT) associés aux composantes identifiées par le modèle de décomposition dans le mélange de PSS et les RT des PSS injectés séparément.

Composante	RT (min)	Standard PSS	RT(min)
C1	13.8	77 k Da	13.8
C2	14.7	32 k Da	14.8
C3	15.5	10 k Da	15.6
C4	16.1	6.8 k Da	16
C5	16.6	4.3 k Da	16.3
C6	17.2	1.6 k Da	16.6
C7	19.8	246 Da	19.9

4.1.2.3 Comparaison entre indicateurs classiques et nouveaux indicateurs dans le cadre de l'étude des sources de MOD

Pour évaluer l'efficacité des nouveaux indicateurs dans un contexte d'étude des sources de MOD, une étude comparative entre les indicateurs classiques MWD (M_n , M_w et ρ) et les indicateurs LMW, MMW et HMW a été réalisée sur un jeu de données de 28 échantillons de MOD. Les échantillons ont été prélevés dans des cours d'eau en aval de sous-bassins versants homogènes en occupation de sol (forêts/prairies, terres agricoles, retenues d'eau), durant une période de hautes eaux et une période de basses eaux. Des analyses multivariées de type analyse en composantes principales ont été appliquées pour cette comparaison.

La Figure 4-4.A représente une ACP des échantillons de MOD en fonction des indicateurs LMW, MMW et HMW calculés sur les chromatogrammes générés aux longueurs d'onde UV 254 et 280 nm et fluorescence (Ex315-Em443, Ex280-Em330 et Ex275-Em354 nm). Dans le premier plan de l'ACP (composantes principales 1 et 2) qui explique 48 % de la variance totale, ces indicateurs permettent de séparer certains groupes d'échantillons de MOD en fonction des sources *a priori*. En période de hautes eaux, les échantillons sont groupés en fonction des usages de sols en amont des points de prélèvement, avec un groupe *agriculture HW* différent du groupe des *forest/pastures* et du groupe *water reservoirs*. Cela pourrait s'expliquer par une contribution terrigène marquée par les usages des sols, plus importante en période de hautes eaux. En effet, les écoulements de surface se produisent principalement lors d'évènements pluvieux, plus présents en période de hautes eaux, et favorisent les échanges entre les compartiments sol et eau de surface (Panton et al., 2020). Les eaux prélevées en aval des sous-bassins forestiers sont caractérisées par une plus forte contribution en composés de faibles poids moléculaires, en comparaison avec les eaux agricoles. Différents degrés de maturation

et de dégradation pourraient expliquer la différence de MOD entre ces deux sources (Thieme et al., 2019).

Pour le groupe d'échantillons prélevés en période de basses eaux, les échantillons présentent des signatures relativement proches, contrairement aux échantillons de même source prélevés en période de hautes eaux. Ceci pourrait indiquer une moindre contribution du compartiment terrigène et potentiellement une signature autochtone plus visible en période d'étiage. Les signatures des retenues d'eau (*water reservoirs HW*) sont très hétérogènes, avec une contribution plus importante des composés de poids moléculaires élevés (de type protéine) et une faible contribution des composés de faibles poids moléculaires. Cela pourrait être dû à différents mécanismes de production et de la dégradation de la MOD au niveau de ces retenues d'eau (*e.g.* biodégradation, photodégradation).

La Figure. 4-4.B montre la répartition des mêmes échantillons en fonction des indicateurs classiques de distribution de poids moléculaire MWD, calculés aux mêmes longueurs d'ondes que pour les indicateurs LMW, MMW et HMW. Le résultat de l'ACP ne permet pas de différencier aussi clairement les groupes de MOD prélevés en fonction des usages de sols, ni en fonction des conditions hydrologiques. En conclusion, les indicateurs LMW, MMW et HMW sont d'intérêt pour caractériser des échantillons en fonction des sources de MOD. En effet, les contributions en petites, moyennes et grosses molécules semblent très différentes entre les échantillons et s'expliquent probablement par des origines et des sources différentes. Les petites molécules caractéristiques de certains échantillons ont été principalement associées à des acides aminés et produits de dégradation microbiens (Kawasaki et al., 2011). Les molécules de poids moléculaire moyen regroupent une large diversité de molécules comme les composés de type humique, les composés de type fulvique, les peptides et les lipides (Cai et al., 2020). Enfin, les composés de poids moléculaire élevé correspondent à des complexes de protéines, des glucides et autre biopolymères (Li et al., 2022).

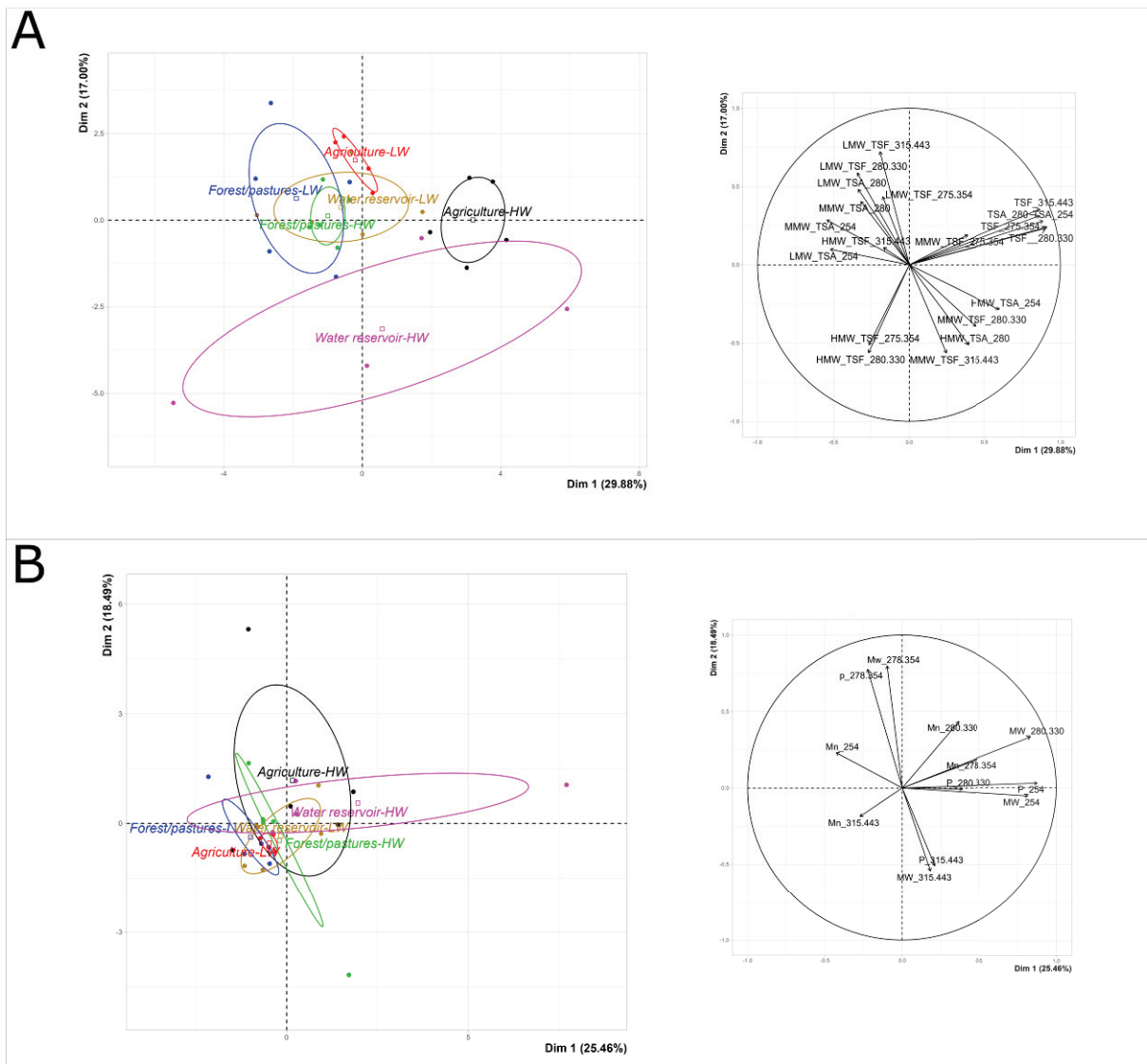


Figure 4-4. Analyse en composantes principales des échantillons de DOM en fonction A. des nouveaux indicateurs (LMW, MMW et HMW) et B. des indicateurs classiques de distribution de poids moléculaire (Mn, Mw et ρ)

Les indicateurs HPSEC/UV-fluorescence permettent d'apporter des informations complémentaires aux indicateurs UV-Visible et EEM fluorescence classiquement utilisés pour caractériser la MOD (Chapitre 3, Tableau 3-5, 3-6). La matrice de corrélation (Figure 4-5) permet de mieux visualiser ces complémentarités. Sur un jeu de données de 135 échantillons de MOD issus de sources naturelles et anthropiques, les corrélations positives et négatives ont été calculées avec un test statistique de Pearson. En cas de corrélation significative ($P < 0.05$), la valeur du coefficient de corrélation est rapportée par une case de couleur bleue pour une corrélation positive et une case de couleur rouge pour une corrélation négative.

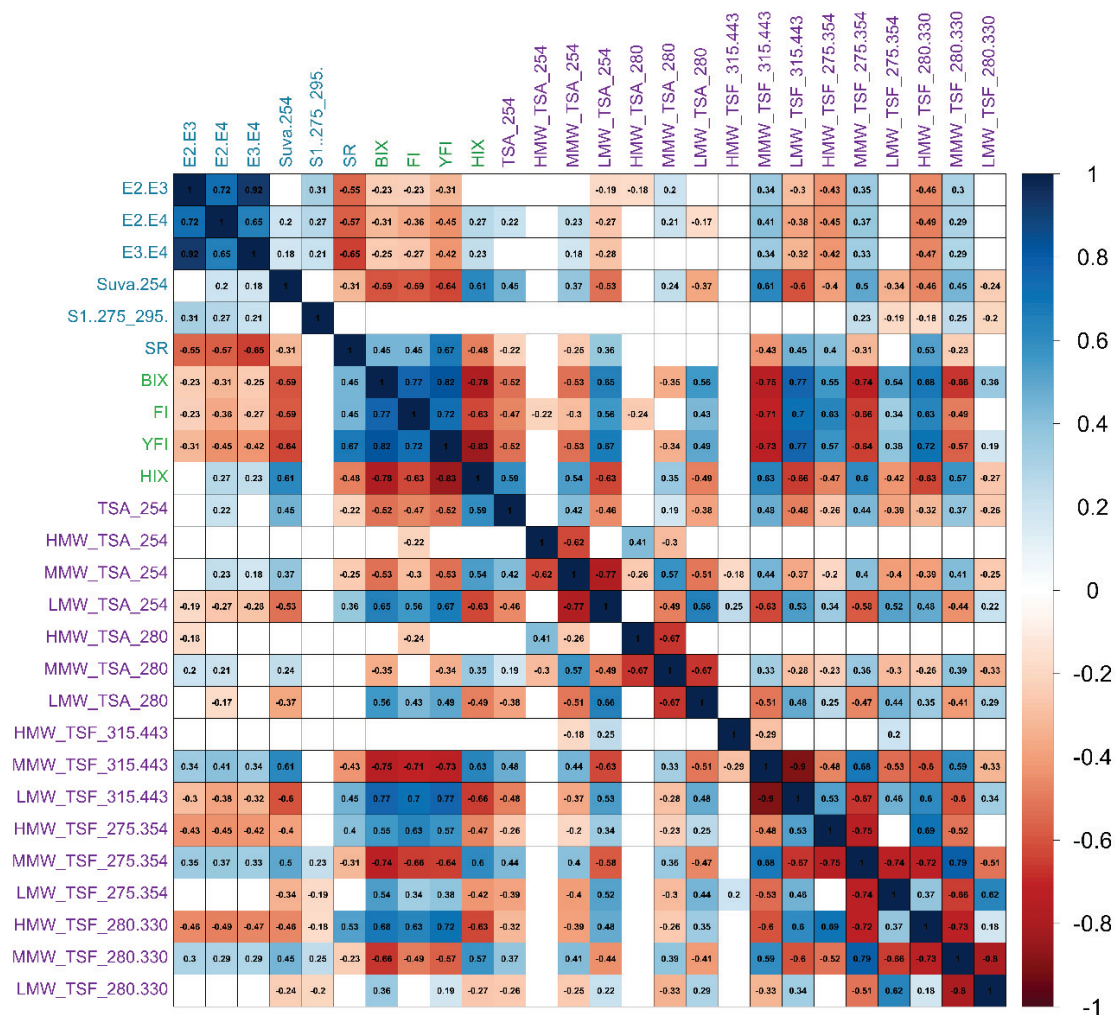


Figure 4-5. Matrice de corrélation des indicateurs UV-Visible, EEM de fluorescence et HPSEC/UV-fluorescence. Les cases blanches indiquent une corrélation Pearson non-significative ($p > 0.05$, $n = 136$ échantillons de DOM).

Pour les indicateurs UV-Visible, le ratio d'absorbance E2.E3, utilisé pour décrire la taille des molécules organiques dans un échantillon d'eau, est inversement corrélé aux indicateurs HMW (> 10 K Da) relatifs aux composés de type protéique (TSF 280-330 et TSF 275-354). En principe, les valeurs de E2.E3 sont faibles en présence de molécules de tailles élevées, ce qui concorde avec le résultat observé dans la matrice de corrélation. En revanche, pour les longueurs d'ondes non spécifiques aux composés de type-protéique, le ratio E2.E3 montre un résultat contradictoire (inversement corrélé aux indicateurs LMW < 1 kDa). Ceci montre que cet indicateur est potentiellement pertinent uniquement pour décrire la taille des composés de nature protéique présents dans un échantillon de MOD. Ce type de contradiction se retrouve également avec le rapport de pentes spectrales SR, souvent utilisé pour décrire le poids moléculaire de la MOD. Le rapport SR est significativement corrélé aux indicateurs LMW relatifs à la longueur d'onde UV 254 nm et aux longueurs d'ondes d'excitation et émission de

fluorescence 315-443 nm (type humique). Ce résultat montre que le rapport SR est pertinent pour décrire le poids moléculaire de certains composés (composés de type humique). Le fait qu'il existe une corrélation négative entre cet indicateur et les indicateurs HMW associés aux composés de type protéique montre qu'il peut lui aussi donner des résultats contradictoires et donc potentiellement fausser l'interprétation dans certains cas, comme dans des eaux très chargées en matériel protéique, telles que les eaux de stations d'épuration. Si l'on s'intéresse à l'indicateur SUVA₂₅₄, qui indique le degré d'aromaticité de la MOD, celui-ci est inversement corrélé aux indicateurs LMW relatifs à toutes les longueurs d'onde analysées. Cela indique que le degré d'aromaticité d'un échantillon de MOD est faible lorsque la contribution en petites molécules est élevée dans un échantillon, et inversement. Le contenu aromatique de la MOD se trouverait principalement dans les composés dont le poids moléculaire est compris entre 1 et 10 kDa (corrélation positive avec l'indicateur MMW).

En ce qui concerne les indicateurs EEM de fluorescence BIX, FI et YFI, qui permettent de différencier des types de matière organique selon une origine autochtone ou allochtone terrestre, ils sont significativement corrélés aux indicateurs LMW de l'ensemble des longueurs d'ondes et aux indicateurs HMW uniquement associés aux composés de type protéique. Ceci montre que l'activité biologique/autochtone traduit la présence d'une MOD constituée à la fois de petites molécules très hétérogènes et de composés de poids moléculaires élevés (biopolymères) de nature protéique. Enfin, l'indicateur HIX, qui reflète le degré d'humification de la MOD dans un échantillon, est inversement corrélé aux indicateurs LMW pour les longueurs d'ondes typiques des protéines. Tout comme pour l'indicateur SUVA₂₅₄, il semble que les composés les plus humifiés soient présents dans la gamme de tailles comprises entre 1 et 10 kDa.

En conclusion, la décomposition gaussienne des chromatogrammes a permis de calculer des indicateurs basés sur le pourcentage de contribution de certaines fractions HPSEC par rapport à l'absorbance ou fluorescence totale. L'indicateur LMW correspond au pourcentage de contribution des petites molécules < 1 kDa, le MMW au pourcentage des molécules moyennes entre 1 et 10 kDa et le HMW au pourcentage des grosses molécules > 10 kDa. Ces indicateurs ont montré leur pertinence pour l'étude des sources de MOD, en particulier pour les sources diffuses (impact de l'usage des sols) qui sont considérées comme les plus difficiles à caractériser car proches en termes de signature. De plus, les nouveaux indicateurs LMW,

MMW et HMW ont montré une complémentarité avec les indicateurs classiques UV-Visible et EEM de fluorescence, non seulement pour caractériser la MOD, mais aussi dans certains cas, pour mettre en évidence les limites d'application de certains indicateurs classiques. Les informations fournies par ces indicateurs pourraient ainsi permettre une meilleure identification des composés d'origine anthropique (*e.g.* pesticides, produits pharmaceutiques) qui sont très difficiles à détecter en utilisant des indicateurs optiques classiques, car la plupart de ces composés ont des propriétés optiques similaires (Hassoun et al., 2017). En outre, l'établissement de bases de données d'indicateurs HPSEC/UV-fluorescence spécifiques aux composés anthropiques d'intérêt pourrait faciliter leur détection à un coût inférieur à celui des techniques actuelles utilisant la chromatographie couplée à la spectrométrie de masse.

4.2 Proposition de nouveaux indicateurs LC-HRMS

4.2.1 Acquisition des données LC-HRMS en approche non-ciblée

4.2.1.1 Introduction – Besoin de *workflows* pour la génération de listes de *features*

Après l'analyse des échantillons en LC-HRMS, la première étape consiste à générer une première liste de *features* (signaux issus du spectromètre de masse et caractérisés par un temps de rétention chromatographique RT, une masse m/z et une intensité), c'est-à-dire extraire l'information sur la composition moléculaire des chromatogrammes des échantillons analysés (Figure 4-6). Il existe pour cela des logiciels fournisseurs comme UNIFI, MassHunter, XCalibur ou des logiciels en libre-service comme MZmine (Pluskal et al., 2010) et XCMS (Smith et al., 2006) qui permettent de générer cette première liste de *features*, d'annoter les adduits et isotopes et de les aligner pour constituer des listes.

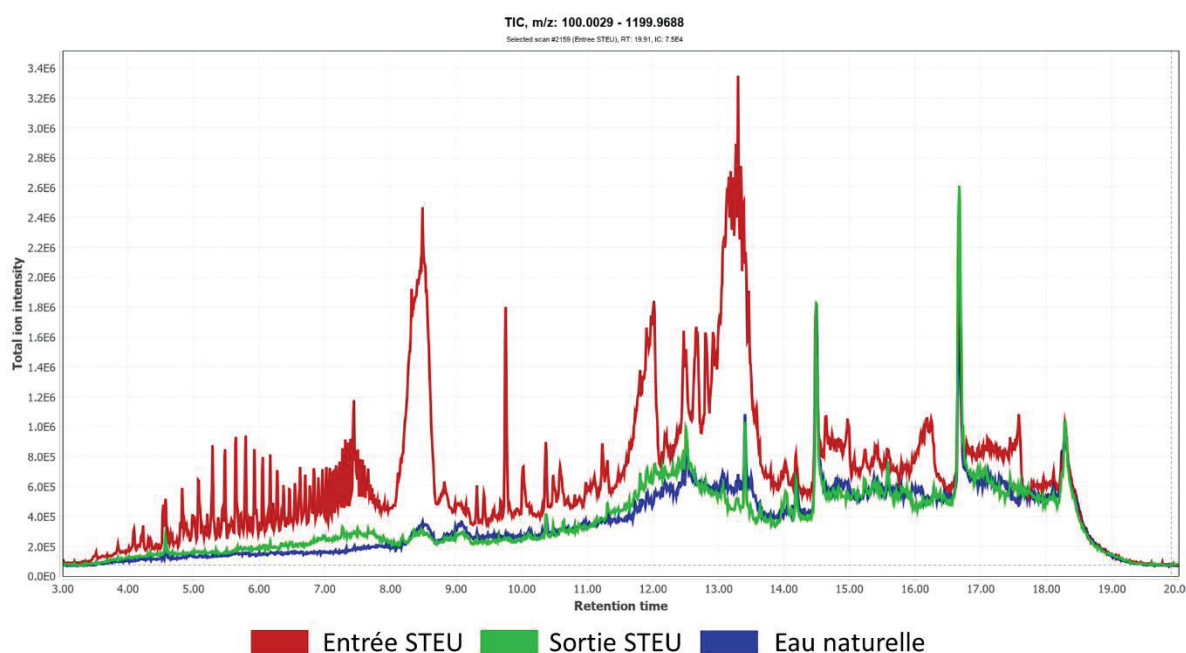


Figure 4-6. Exemple de chromatogrammes TIC (*Total Ion Chromatograms*) générés par LC-HRMS d'eaux de STEU brute et traitée ainsi que d'une eau naturelle prélevée en aval d'un bassin versant forestier. Le TIC représente la somme des signaux générés par tous les ions qui sont élués de la colonne chromatographique au cours de l'analyse.

Parmi l'ensemble des outils mis à disposition pour l'acquisition des données HRMS, MZmine et XCMS sont les plus utilisés à l'heure actuelle. Cela s'explique principalement par leur facilité

d'utilisation, leur disponibilité en libre-service et leur utilisation d'un des algorithmes de détection de pics chromatographique les plus performants, le CWT (*continuous wavelet transform*). Cependant, une précédente comparaison des données générées par les deux logiciels a mis en évidence des différences significatives dans le nombre de *features* détectés, expliquées en partie par la détection conséquente de faux positifs (Myers et al., 2017). Plusieurs améliorations ont par la suite été apportées aux deux logiciels avec les nouvelles versions de MZmine (2.5 et 3.0) et de XCMS (3.5., via la plateforme *workflow for metabolomics-W4M*, <https://workflow4metabolomics.org/>).

La section suivante décrit une comparaison entre les deux logiciels d'acquisition MZmine et XCMS (W4M), en utilisant leurs nouvelles versions respectives 2.5 et 3.5. Pour cela, les mêmes étapes de génération de liste de *features* (Chapitre 3, Partie 3.2.5) ont été appliquées sur trois séquences (*batch*) d'analyses LC-HRMS. Chaque séquence d'analyse était constituée d'échantillons de cours d'eau représentatifs de sources diffuses (n = 28) injectés en triplicats (distribution aléatoire pendant la séquence) et de contrôles qualité (blancs d'injection : n=17, QCP : n=25 et QC mix : n= 3 ; Chapitre 3, Partie 3.2.5.1). Les échantillons de MOD, les blancs d'injection et les QCP ont été dopés au préalable avec une solution de traceurs d'injection (tébuconazole D6 et diuron D6) à 10 µg.L⁻¹. Plus d'informations sur la mise en place des séquences d'analyse sont données dans le Chapitre 3, Partie 3.2.5.1.

4.2.1.2 Comparaison des *workflows* MZmine et XCMS pour la génération d'une première liste de *features*

La comparaison entre les listes de *features* générées par chaque logiciel a ensuite été réalisée à l'aide de l'algorithme d'alignement *MetabCombineR* pour confronter les listes de *features* et détecter ceux en commun ou spécifiques à chaque logiciel d'acquisition. Une analyse des résultats basée sur les contrôles qualité a été réalisée afin de comparer qualitativement l'information produite par les deux logiciels MZmine et XCMS. La présence ou absence des traceurs d'injection et des composés du QC mix (Tableau 3-7, Chapitre 3) dans les listes de *features* a été le premier paramètre de comparaison. Pour cela, un seuil de tolérance en masse inférieur à 10 ppm et à 0,3 min en temps de rétention a été appliqué pour identifier les composés. Par la suite, les contrôles qualité blancs d'injection, QC Pool et triplicats d'injection ont été utilisés comme filtres pour supprimer l'ensemble des *features* dont l'intensité était

instable lors de l'analyse (proche du bruit de fond, variabilité élevée tout au long de l'acquisition). Cela a permis de générer la liste définitive de *features*. Le détail de l'ensemble de ces étapes supplémentaires d'application des filtres est présenté dans la partie 4.2.2. de ce chapitre 4 (article 2).

Le nombre de *features* détectés constitue un premier élément de comparaison. La Figure 4-7 montre le nombre de *features* détectés et alignés à la fin de l'acquisition avec XCMS (W4M) et MZmine, ainsi que les *features* en commun, mis en évidence par l'algorithme MetabCombineR. Les *features* sont considérés comme communs si la différence entre les RT est inférieure à 0,3 min, et celle entre les m/z est inférieure à 10 ppm. Ces seuils de tolérances ont été calculés à partir des variabilités de RT et m/z observées sur les traceurs d'injection diuron d6 et tébuconazole D6 dans 5 séquences différentes analysées de janvier à avril 2021 (Figure 4-8).

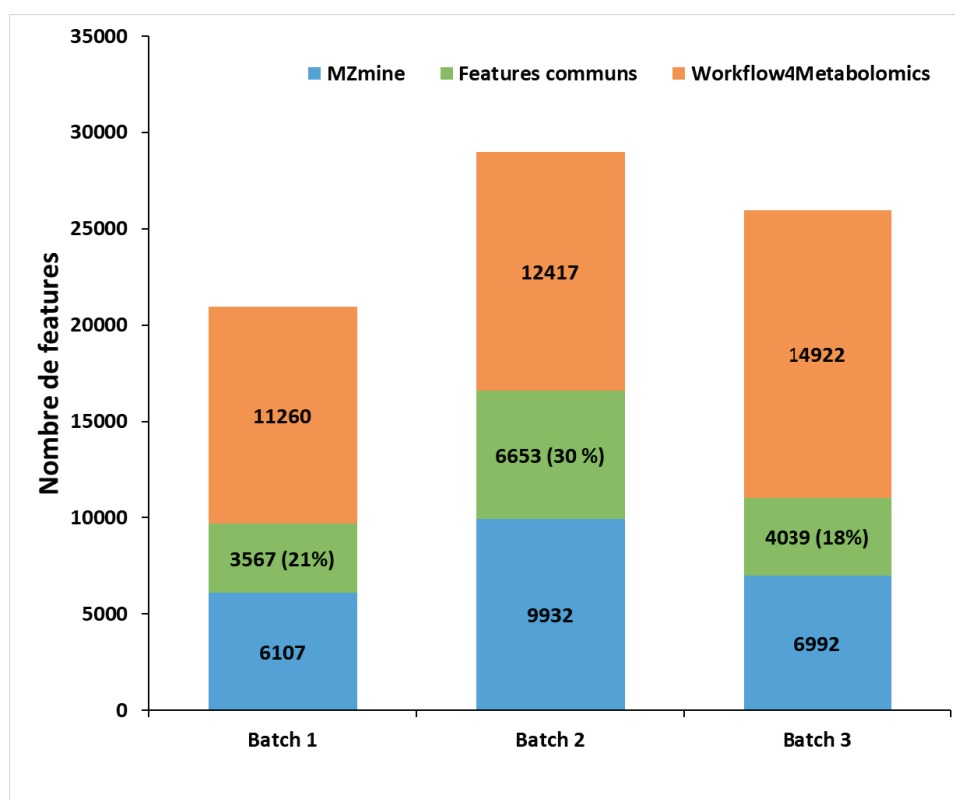


Figure 4-7. Histogrammes empilés pour 3 séquences d'analyse avec le nombre de *features* détectés par XCMS (WorkflowForMetabolomics) et MZmine et les *features* en commun aux deux logiciels.

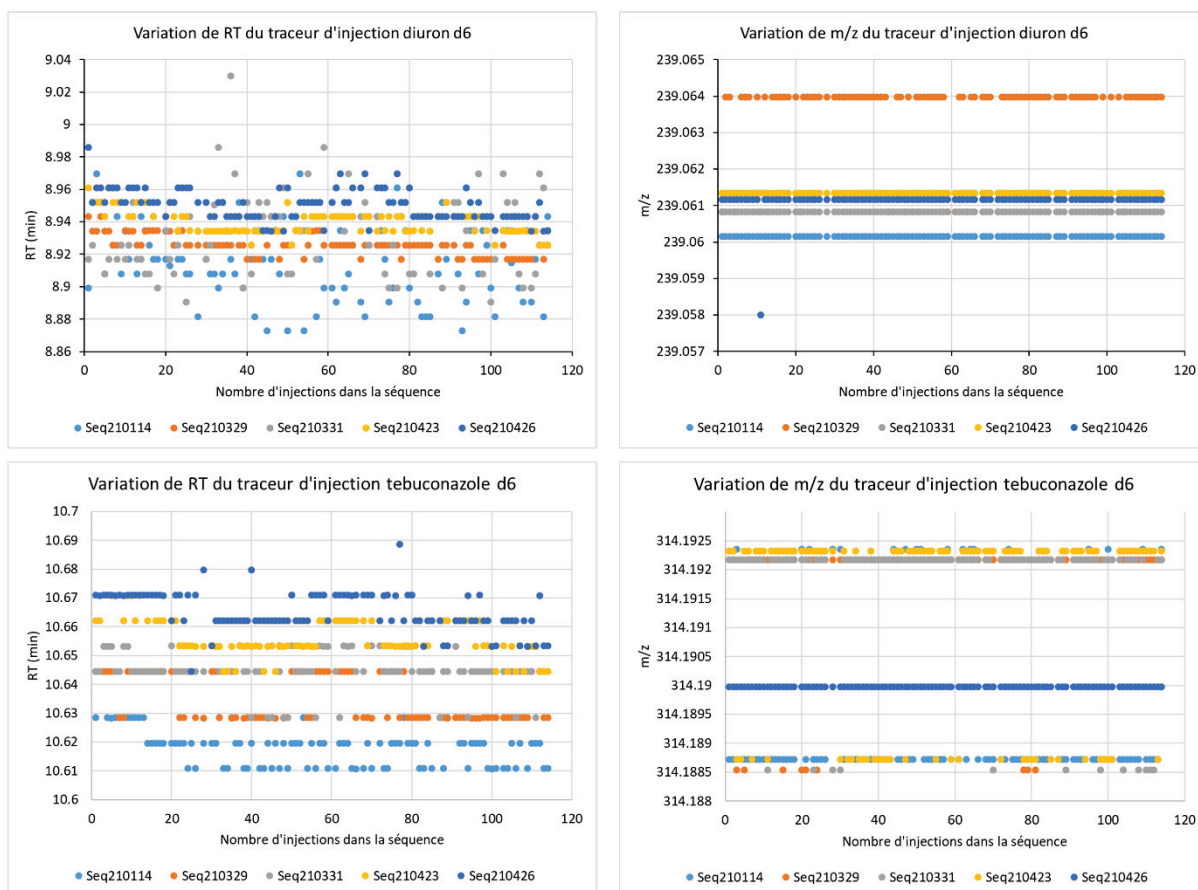


Figure 4-8. Etude de la variation de RT et m/z des traceurs d'injection diuron d6 et tébuconazole D6 dans 5 séquences d'analyses de janvier à avril 2021.

Le résultat montre que le traitement avec XCMS génère plus de *features*, avec respectivement 11260, 12417 et 14922 dans les séquences d'analyse 1, 2 et 3, tandis que pour MZmine, le nombre de *features* est de 6107, 9932 et 6992. Le pourcentage de *features* en commun varie en fonction des séquences d'analyse, et il est compris entre 18 et 30 % du nombre total de *features*. Ces résultats concordent avec ceux de l'étude précédente de Myers et al. (2017). Cette étude a montré qu'un nombre important de pics faux-positifs sont observés par l'algorithme CWT (*i.e.* l'algorithme de détection de pics chromatographiques, *continuous wavelet transform*, inclus dans XCMS et MZmine). D'autres pics d'intérêt (faux-négatifs) sont parfois détectés par un logiciel et pas par l'autre, toujours lors de cette étape réalisée par l'algorithme CWT. Le résultat de notre étude montre que ce problème de faux-positifs et faux-négatifs est toujours présent dans les nouvelles versions de MZmine et XCMS. Dans un contexte d'identification de composés, cela impacte directement le résultat final. Toutefois, dans une approche globale d'établissement d'empreintes de sources, l'impact pourrait être plus faible.

Les traceurs d'injection Tbz d6 et Diu d6 ont été détectés dans les listes de *features* générées par MZmine et XCMS des trois séquences d'analyse. La même procédure a été appliquée pour les composés du QC mix : aténolol, azoxystrobine, bezafibrate, céliprolol, tébuconazole, tiapride, venlafaxine (Figure 4-9).

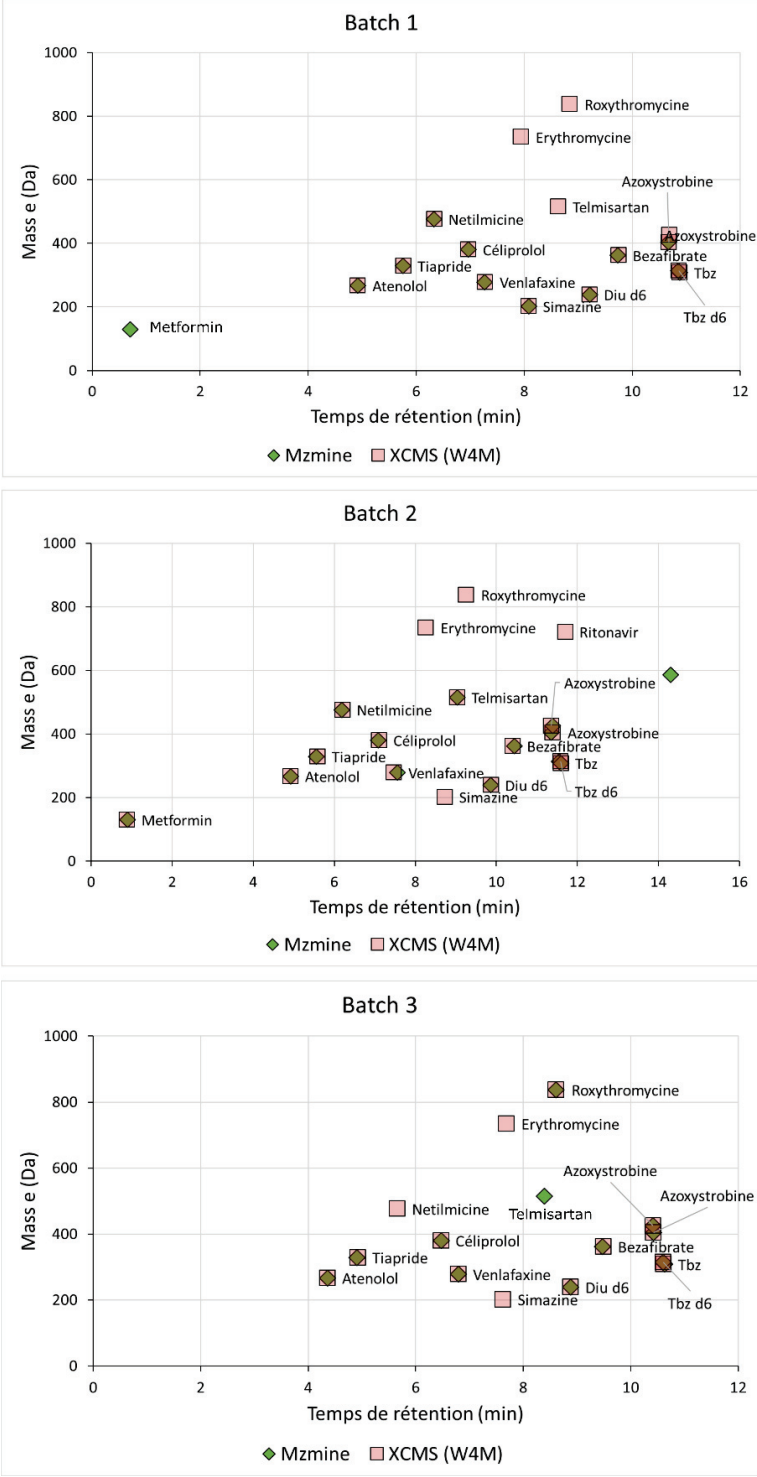


Figure 4-9. Comparaison des traceurs d'injections retrouvés après traitement par MZmine et XCMS.

D'autres composés comme l'érythromycine, la simazine et la roxythromycine n'ont été détectés que dans les listes de *features* générées par XCMS avec des intensités relativement faibles (<1000 cps). La différence de méthode d'établissement du niveau de bruit de fond entre XCMS (égal au rapport signal sur bruit de 3) et MZmine (égal à la valeur absolue de seuil d'intensité de 50 cps) permet d'expliquer le nombre significativement plus élevé de *features* détectés avec XCMS, comparé à MZmine. Le bruit de fond de 50 cps correspond en effet à un niveau plus élevé que qu'un rapport signal sur bruit de 3, ce qui enlève davantage de pics de faibles intensités (Figure 4-7). Les composés ritonavir et montelukaste ont quant à eux été détectés dans une seule séquence générée par XCMS, avec des intensités significativement plus faibles que celles des autres composés (< 300 cps). Enfin, les composés cétirizine et époxiconazole n'ont pas été détectés dans l'ensemble des listes de *features* des trois séquences. Il est possible que ces deux composés aient été dégradés pendant la période de conservation des QC mix (-20° C).

La dernière étape de comparaison des logiciels était principalement portée sur une répartition *a priori* des QC (blancs d'injection : n=17, QC Pool : n=25 et QC mix : n= 3) en utilisant des analyses multivariées. En théorie, les blancs d'injection doivent être regroupés et différents des échantillons des cours d'eau, les QC mix doivent être proches des blancs d'injection et regroupés en triplicat. L'application de cette dernière étape de comparaison nécessite néanmoins un traitement préalable des listes de *features* (suppression des *features* dont l'intensité est fortement variable lors de l'acquisition, cf. partie 4.2.2 pour les explications sur le principe et les performances de ce traitement de données). Les étapes de filtration ont permis de réduire drastiquement le nombre de *features* de 6 107 à 923, de 9 932 à 689 et de 6 992 à 895, respectivement pour les séquences d'analyses 1, 2 et 3 générées par MZmine. Pour les mêmes séquences générées par XCMS, le nombre de *features* est passé de 11 260 à 457, de 12 417 à 568 et de 14 922 à 535. Une analyse en composante principale a ensuite été réalisée avec les listes finales de *features* pour comparer la répartition des contrôles qualités (Figure 4-10). Le premier plan de l'ACP (composantes principales 1 et 2), réalisé avec la liste de *features* générée par MZmine et qui explique 47,4 % de la variance totale, montre une répartition hétérogène des échantillons des cours d'eau (échantillons de MOD) et des QCP. Les blancs d'injection et les QC mix se confondent, avec une grande variabilité entre les triplicats des QC mix. La deuxième ACP, réalisée avec la liste de *features* générée par XCMS et

expliquée par un pourcentage de variance de 23,5 %, montre une meilleure distribution des échantillons de MOD et QC Pool. Les blancs d'injection et QC mix sont proches, avec moins de variabilité observée entre les différentes injections du QC mix, en comparaison avec la première ACP générée avec les données de MZmine.

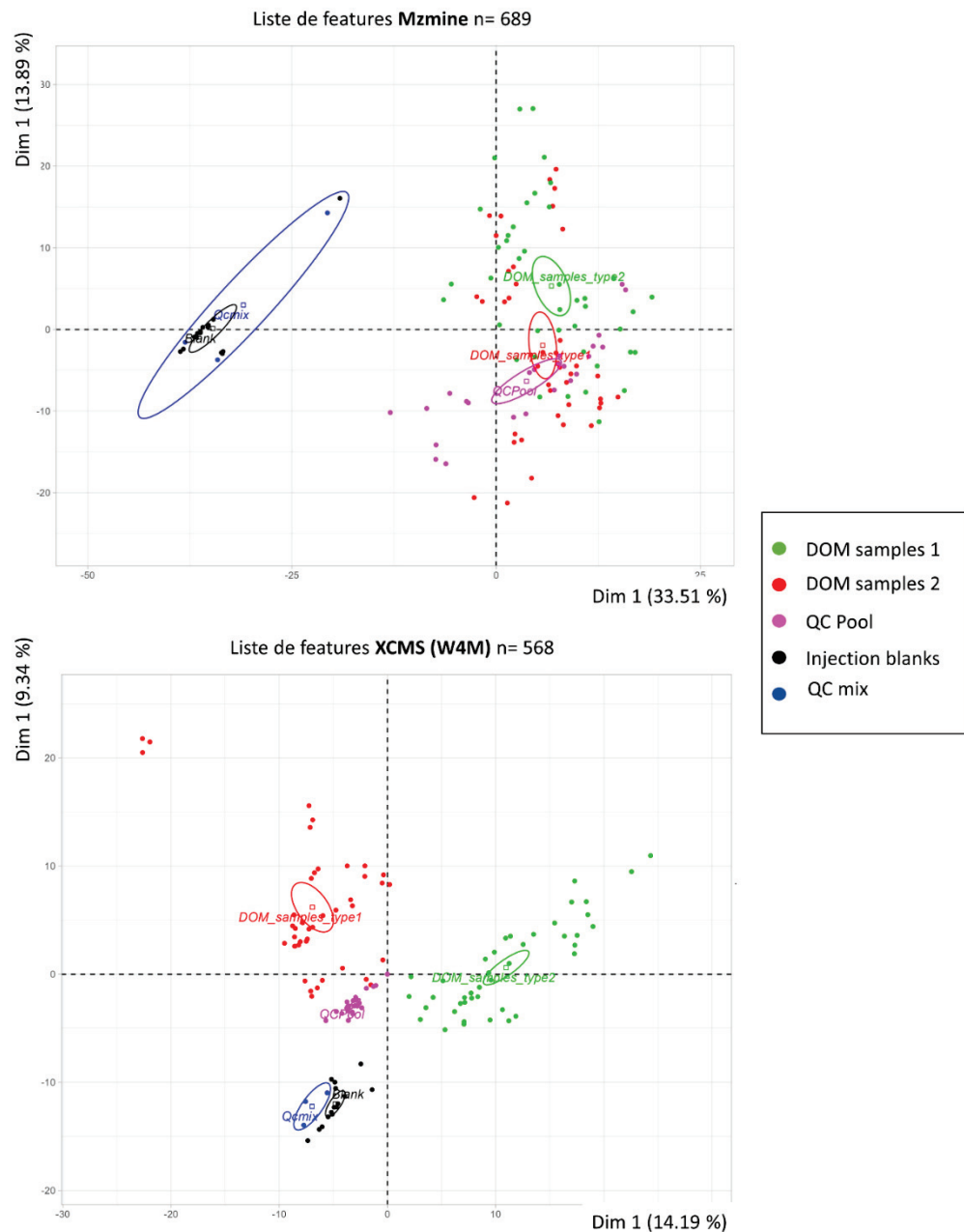


Figure 4-10. Analyses en composantes principales des injections d'échantillons de DOM et contrôles qualités (blancs d'injection : n=17, QCpool : n=25 et QC mix : n= 3) en fonction des intensités des *features* générés par MZmine et XCMS.

De manière globale, la variabilité intra-groupes d'échantillons est plus faible, ce qui constitue un résultat encourageant. De plus, et contrairement à MZmine, les injections de QCP sont bien

regroupées entre eux et sont positionnées entre les deux groupes d'échantillons de MOD. Etant donné que le QCP correspond à un seul échantillon composite de l'ensemble des échantillons de MOD injectés dans la séquence, il doit normalement contenir l'ensemble des *features* à des intensités plus faibles, car diluées et se situer au barycentre de l'ensemble des échantillons dont ils sont constitués. C'est le cas dans cette deuxième ACP générée avec XCMS. A partir de l'ensemble des résultats présentés dans cette partie, nous pouvons donc affirmer que l'acquisition des données par XCMS est plus pertinente qu'avec MZmine dans le cadre de notre étude.

4.2.2 Retraitement des données HRMS pour la génération de la liste finale de *features* en vue de l'étude des sources de MOD (publication n°2)

Avant-propos

Dans ce sous-chapitre, nous proposons une méthodologie de retraitement des données LC-HRMS non-ciblée (*workflow*) principalement basée sur les intensités des composés moléculaires. Le *workflow* reprend des étapes-clés qui permettent de nettoyer la donnée brute, souvent parasitée par des centaines de pics issus des conditions analytiques et des étapes de génération de la liste de *features*. Les résultats de cette partie sont présentés sous la forme d'une publication scientifique.

4.2.2.1 Résumé de la publication scientifique

Cette étude vise à établir une méthodologie de traitement des données (*workflow*) issues de l'analyse LC-HRMS non-ciblée afin d'étudier l'impact de l'usage des sols et des conditions hydrologiques sur la qualité de la MOD dans les cours d'eau. Les données ont été acquises avec le package XCMS via la plateforme WorkflowForMetabolomics grâce aux résultats précédents. Les adduits et les isotopes ont été annotés avec l'algorithme CAMERA et la dérive instrumentale des intensités des composés a été corrigée à l'aide du modèle de régression *Loess*.

Le *workflow* établi dans le cadre de ces travaux a consisté à traiter les adduits et les isotopes, puis à sélectionner les composés moléculaires fiables, c'est à dire caractérisés par des intensités élevées et stables. Pour cela, des valeurs seuils d'intensités et d'écart-type relatif (RSD) ont été établies à l'aide de contrôles qualité (QC) (blancs n=17, QC pool n=24 et échantillons randomisés en triplicats n=28). Les blancs correspondaient à de l'eau ultrapure de qualité Milli-Q (LC-Pack, Millipore), le QCP à un échantillon composite des 28 échantillons représentatifs de différents type d'occupation des sols (sources diffuses agricole et forestière/prairies), prélevés en période de hautes eaux et basses eaux. L'ensemble des contrôles qualités et échantillons a été dopé à l'aide des traceurs d'injection tébuconazole D6 et diuron D6 pour disposer d'un suivi de l'analyse.

La première étape de traitement des adduits et des isotopes annotés pour un même composé a montré que la somme des intensités des adduits ou isotopes associés aux mêmes composés moléculaires fournissait un résultat plus précis que la prise en compte de l'intensité de l'adduit ou de l'isotope principal. L'étape de nettoyage des données, utilisant un seuil basé sur les intensités des blancs d'injection, a permis d'éliminer les composés proches du bruit de fond (représentant 40 % des composés). Ensuite, une seconde étape, basée sur un seuil de variabilité de l'intensité (RSD) considérant les triplicats de QCP et des échantillons de MOD, a permis d'éliminer les composés moléculaires ayant des intensités trop variables (représentant 82 % des composés restants, pour un seuil maximum de variabilité d'intensité RSD de 30 %). La liste finale des composés moléculaires (n = 463) obtenue à l'issue de ces étapes a été considérée pour l'étude des sources de MOD dans les cours d'eau. Sur la base d'analyses statistiques multivariées, l'étude des sources de MOD a mis en évidence que les conditions hydrologiques avaient un impact considérable sur les caractéristiques de la MOD dans les cours d'eau, avec des contributions de MOD terrigène pendant les périodes de hautes eaux cédant la place à une MOD plus autochtone pendant les périodes de basses eaux.

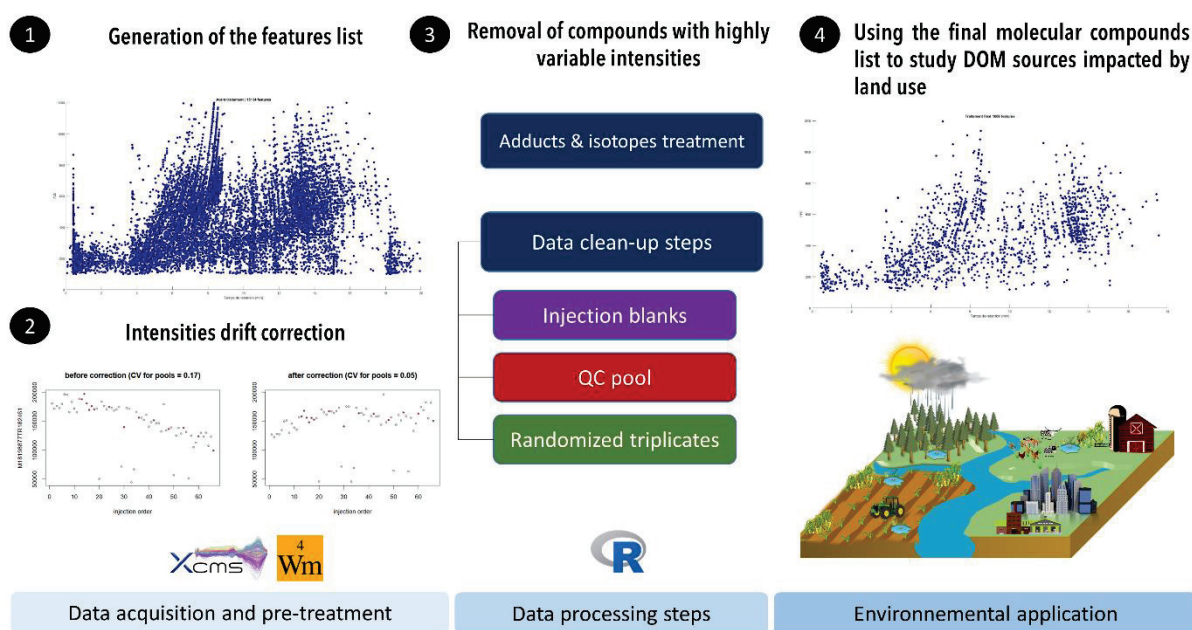


Figure 4-11. Résumé graphique de la publication scientifique: *Processing LC-HRMS non-targeted data to characterize dissolved organic matter in watercourses impacted by land uses*. En cours de soumission dans : *Journal of Environmental Sciences*

Publication n°2. Processing LC-HRMS non-targeted data to characterize dissolved organic matter in watercourses impacted by land uses.

Amine Boukra¹, Matthieu Masson¹, Kevin Rocco¹, Christelle Margoum¹, Cécile Miège¹

¹INRAE, UR RiverLy, 69625 Villeurbanne, France

Correspondence to: matthieu.masson@inrae.fr; +33472208706; RiverLy, INRAE

5 rue de la Doua, Villeurbanne, France, 69625

Highlights:

- LC-HRMS non targeted data cleaning methodology based on compounds intensities
- Quality control threshold sets allowed to remove compounds with too variable intensities
- Molecular fingerprints allow to differentiate between the types of DOM sources
- Sampling in contrasted hydrological conditions allows to differentiate terrigenous and autochthonous DOM

Abstract

This study aims to establish a methodology for processing data from LC-HRMS non-targeted screening (NTS) in order to study the impact of land uses and hydrological conditions on dissolved organic matter (DOM) quality in watercourses. The data acquisition was conducted with XCMS package through WorkflowForMetabolomics platform and CAMERA algorithm was used to annotate adducts and isotopes. Instrumental drift of compounds intensities was corrected using loess regression model. Then, data clean-up consisted of selecting reliable molecular compounds characterized by high and stable intensities. We defined thresholds values of intensities and relative standard deviation (RSD) calculated using various quality controls (QC) (blanks: n=17, QC pools: n=24, and randomized samples triplicates: n=28). At the end, this data clean-up led to reduce the number of molecular compounds by a factor 3 from 13762 to 463. Based on multivariate statistical analyses, the most stable and accurate molecular compounds were selected for a DOM source study. This study highlighted that the hydrological conditions strongly impacted the DOM characteristics in watercourses, with exclusive inputs of terrigenous DOM inputs during high water periods, and a mixture of terrigenous and autochthonous DOM during low water periods. In addition, an impact of various land uses (forest, pastures, vineyard and water reservoir) was detected during high water periods allowing the identification of typical fingerprints from the different DOM sources.

Keywords: High-resolution mass spectrometry, Data processing, Dissolved organic matter, Non-point sources, Land use.

1. Introduction

Dissolved organic matter (DOM) is considered as one of the most complex and heterogeneous set of organic compounds (Catalá et al., 2021) found in watercourses. It is operationally defined as the material obtained after water filtration through 0.7 µm filter (Zsolnay, 2003). At watersheds scale, DOM is ubiquitous and can originate from natural or anthropogenic sources such as the terrigenous or autochthonous sources, industries, urban runoffs or wastewater treatment plants WWTP's (Aiken, 2014). The diversity of these sources is at the origin of different DOM compositions (Parlanti et al., 2002; Mangal et al., 2020). DOM can lead to changes in the physical and chemical properties of the aquatic ecosystems (e.g. impacts on nutrient cycling and availability for aquatic organisms, interactions with pollutants such as heavy metals) (Elovaara et al., 2021; Li and Gong, 2021). Among all the possible types of DOM inputs to rivers, diffuse terrigenous sources related to land uses are the most difficult to identify and characterize. This is mainly due to the fact that inputs from these sources are spread in space and over time. A dilution effect of organic compounds in rivers also makes it difficult to detect certain pollutants (e.g. pesticides, pharmaceuticals and by-products, Polycyclic Aromatic Hydrocarbons) associated with particular land uses (e.g. agriculture, livestock, urban development), regardless of the consequences that can be critical for the aquatic ecosystem (Gormley et al., 2005; Cheng et al., 2022). Still, DOM, which represents a combination of these anthropogenic compounds in addition to those naturally present in the environment (e.g. lipids, proteins, humic-like substances), could constitute a good tool for identifying anthropogenic diffuse discharges but also an indicator of good water quality (Wu et al., 2022).

For many years, the characterization of DOM has attracted the interest of the scientific community and still remains a challenge to this day (Chow et al., 2022). In reality, there is an infinite number of possible compositions of DOM, mainly due to the great variability of its organic constituents. In addition, the high reactivity of DOM (degradability, transformation according to physico-chemical water composition ...) also influence its composition, which makes it very difficult to define DOM precisely at a given time. Optical analyses, such as UV absorbance spectroscopy and excitation emission matrices fluorescence (EEM) have provided major advances in the understanding of the composition and fate of DOM in aquatic environments (Stedmon et al., 2003). These techniques have allowed the development of

indicators giving information on the molecular weight or size, aromaticity or level of humification of DOM (Fellman et al., 2010; Li and Hur, 2017). However, UV-visible and EEM fluorescence can only target specific fractions of DOM such as chromophoric and fluorescent components. Other spectroscopy techniques, such as nuclear magnetic resonance (NMR) have led to investigate DOM on a structural aspect. The NMR technique highlighted the presence of lignin structures in soil organic matter (Simpson, 2002), and also to study interactions between mineral environmental fraction and aliphatic and aromatic DOM components (Mitchell et al., 2018). Although there is potential for further improvement in the future, the application of NMR remains however limited, due to difficulties in sample preparation (tedious extraction steps, use of appropriate solvent) and due to the relatively low resolution of signals not permitting to characterize the complex mixture of molecular compounds constituting the DOM (Mitschke et al., 2022). Instead, high-resolution mass spectrometry (HRMS) techniques, such as ultra-high resolution Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) allowed identifying several thousands of predicted chemical formula in marine and freshwater DOM samples. The mass accuracy (<0.3 ppm) and the very high mass resolution ($>500\,000$) of the FT-ICR-MS coupled with electrospray ionization (ESI) enabled to extract elemental composition of ions. This made it possible to calculate ratios of elements H/C, O/C and N/C to differentiate DOM samples according to chemical structures (e.g. lipid, protein, carbohydrate, lignin or tannin) (Kamjunke et al., 2019; Zhu et al., 2019).

Other alternatives to study DOM molecular characteristics concern HRMS techniques coupled with liquid chromatography (LC). The combination of these two techniques enable to identify and quantify compounds (e.g. molecular features) based on their mass-to-charge (m/z), chromatographic retention time (RT) and intensity. This has opened up new research perspectives on DOM complex composition, as the inclusion of the chromatographic part allows a better separation of the highly heterogeneous organic molecules composing DOM (Brock et al., 2020). However, the lack of DOM analytical standards leads to use analyses strategy based on non-targeted screening (NTS), this means analyzing a sample without any prior knowledge of its chemical composition (Krauss et al., 2010; Hernández et al., 2014). Non-targeted screening strategies based on HRMS have recently demonstrated conveniences for the study of DOM, as it provides an efficient and cost-effective way to analyze a wide range of

organic compounds (Aguilar-Alarcón et al., 2020). For example, in wastewater treatment processes (Verkh et al., 2018a; Schollée et al., 2021), researchers have demonstrated a transformation of DOM throughout the water treatment steps with changes on the number and the chemical nature of molecular features detected. These previous studies have also shown the ability of the non-targeted approach to fingerprint specific sources of organic matter. The basis of NTS lies on the use of post-acquisition processing softwares (e.g. MZmine, xcms) that enable extracting masses from the raw data, to build chromatograms and to align peaks according to the samples (Smith et al., 2006; Pluskal et al., 2010). Each aligned signal is called a molecular feature and is characterized by a m/z , a RT and an intensity. These processing softwares often incorporate annotation algorithms (e.g. CAMERA, cliqueMS, MZadan) to annotate isotopes, adducts and fragments based on mass, RT and peak shape information of each molecular feature (Kuhl et al., 2012; Senan et al., 2019; Stricker et al., 2021). At the end, an aligned list of thousands annotated molecular features is obtained. In order to properly exploit this list, it is necessary to set up a series of quality controls (QC), such as injection blanks, analytical tracers and QC pools (a composite sample of all the samples injected in an analytical batch). These QCs ensure that the analysis runs correctly. They are also used to correct any instrumental drift and to select features among the very complex and dense HRMS datasets. However, several questions remain regarding the reliability of the data generated in non-target analysis. Indeed, given the large amount of information generated by this type of approach, the importance of establishing relevant data treatment is crucial and we still need to develop analytical strategy combining QCs and data processing workflows.

In a context of studying the impact of land use on DOM quality in rivers, this study aims to establish a methodology for processing NTS LC-HRMS data, based mainly on the exploitation of molecular features intensities. For this, DOM samples, as well as quality controls (QCP, injections blanks, randomized triplicates), were analyzed by LC-HRMS coupled with a Quadrupole Time-of-Flight (QToF). The molecular features list was generated using an automated workflow with XCMS package on WorkflowForMetabolomics platform (<https://workflow4metabolomics.org/>). After a correction of the instrumental drift on the features intensities, we tested adducts and isotopes processing and a series of quality control filters (tolerance thresholds based on intensities variability) to select only the most accurate compounds with the least variability along the analysis. In a second step, we used the filtered

data to study molecular fingerprints related to forest/pasture, agriculture and water reservoir sources in watercourses.

2. Materials and methods

2.1. Dissolved organic matter (DOM) sources sampling

The sampling was performed in the Ardières-Morcille watershed (152 km²) (Gouy et al., 2021), in south east of France (Figure 4-12). Five stream samples (FW1 to FW5) were collected manually in the watercourses, downstream homogenous sub-catchment areas influenced by a mix of forest, lands, meadow and fewpasture land uses (called forest/pasture). Five other samples (AW1 to AW5) were collected in rivers downstream sub-catchment areas mostly characterized by vineyard land uses, except for AW4, which drains different proportions of land uses but its proximity to vineyard plots, led to classify it with the vineyard group of samples. Finally, four additional samples (WR1 to WR4) were directly collected in artificial water reservoirs (WR). The land use influence for these WR were defined according to the upstream drained areas: WR1, WR2 and WR4 mainly vineyard and WR3 mainly forest (Figure 4-12). QGIS 3.10 software was used to determine the area of the sub-watersheds drained, as well as the percentage of land uses at each point (Corine Land cover CLC 2018).

The water samples (n=28) were collected during two campaigns, according to the hydrological state of the watercourses. A first sampling campaign, in January 2021, was conducted during a period of high water (3.65 and 0.2 m³.s⁻¹, at the outlet of Ardières and Morcille rivers respectively); and a second campaign, in July 2021, during a period of low water (0.65 and 0.03 m³.s⁻¹, at the outlet of Ardières and Morcille rivers respectively). The purpose of the sampling during these two periods was to highlight a terrigenous input, which could differentiate river DOM during the high water period from a more autochthonous type of river DOM during the low water period. Each spot water sample was collected in a 100 mL borosilicate amber glass bottle, previously washed using an automatic washer at 90°C, then calcined in an oven at 500°C. The samples were filtered in the field using polypropylene syringe and polyvinylidene fluoride (PVDF) filters with a porosity of 0.2 µm. The filtrate was first used to rinse the bottles three times before filling them completely. The samples were then stored in a cooler at 4°C, protected from light during their transport. Once in the laboratory, the samples were stored in amber glass vials of 1.5 mL at -20 °C until analysis.

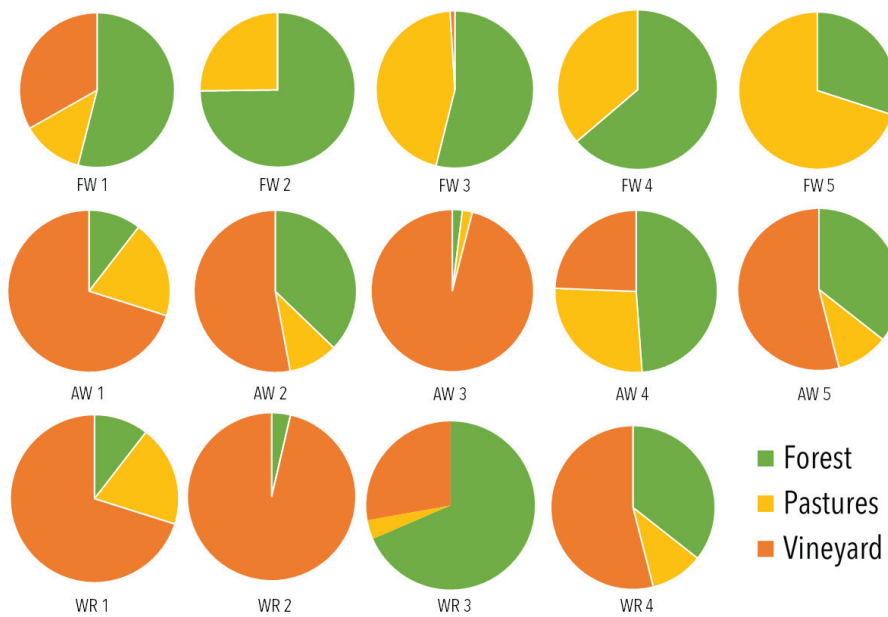
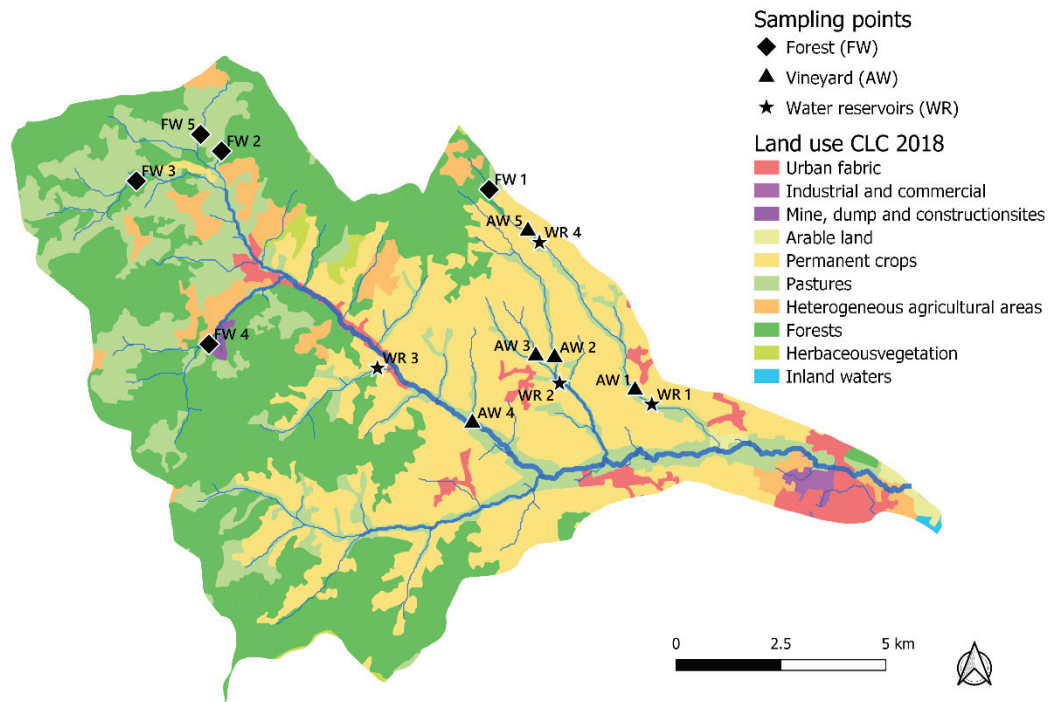


Figure 4-12. Map of Ardieres-Morcille catchment with lands-use boundaries (Corine Land Cover CLC 2018) showing the location of the water-sampling sites. FW: forest water, AW: agricultural water, WR: water reservoir. The percentage of land use contribution for each point was calculated based on the upstream land use area.

2.2. Chemicals and reagents

Acetonitrile and methanol (UPLC-MS grade) were acquired from Biosolve (Dieuze, France). Formic acid and Leucine-Enkephalin (LC-MS grade) were purchased from Waters (Milford, MA, USA). Certified injection tracers: tebuconazole D6 and diuron D6 were purchased from Dr. Ehrenstorfer (Augsburg, Germany). Milli-Q grade water (LC-Pack, Millipore) was used for the analysis and injection blanks preparation.

2.3. LC-HRMS analysis

The samples were analyzed by LC-HRMS composed of an ACQUITY UPLC H-Class system (Waters) coupled to a quadrupole time-of-flight mass spectrometer (Xevo G2-S Q-TOF; Waters). The filtered water samples were injected onto a HSS T3 C18 reversed-phase column (100 mm x 2.1 mm, 1.8 μ m) from Waters. The mobile phase consisted of (A) water + 0.1 % formic acid and (B) acetonitrile + 0.1 % formic acid. The analyses were done at a flow rate of 0.5 mL/min at 30 °C for a total runtime of 20 min, using the following gradient: initial condition with 2% of B for 2 min, followed by a linear increase to 99% of B for 13 min and a hold for 2 min; a return to the initial condition in 1 min to have 2% of B and a hold for 2 min (Bride et al., 2021). For QToF parameters, the ionization was in positive mode ESI+ with the following settings: capillary voltage at 2.3 kV, source temperature at 150 °C, desolvation gas flow at 700 L/Hr, sample and lockSpray infusion flow rates at 20 μ L/ min. Data independent acquisition (DIA) was employed to acquire data in a mass range of 100 to 1500 m/z, with a reference scan frequency of 10 seconds.

2.4. Design of the analytical batch and implementation of quality controls

The analytical batch was designed taking into account all DOM samples (Section 2.1), as well as a series of quality controls. DOM samples were injected in randomized triplicates to eliminate sources of systematic bias in the data and reduce the risk of carryover contamination. In addition to randomized triplicate injections of samples, other QCs were set in order to have a follow-up of the instrumental drifts or particular anomalies during an analytical batch (Schulze et al., 2020). In this study, the following QCs were used:

- QC pool (QCP): corresponds to a composite sample of an equal volume of all samples (final volume 1 mL) injected throughout one analytical batch of samples. QCP was used to correct intensities drifts during the analytical batch.
- Injection blanks: Milli-Q water injected throughout the analytical batch. Injection blanks were used in post processing data in order to remove peaks with intensities close to background noise.
- Injection of deuterated tracers: a mix of diuron-d6 and tebuconazole-d6 were added ($10 \mu\text{g.L}^{-1}$) in each sample, injection blank and QCP to assess RT drift during acquisition.
- QC mix: a mixture of 13 analytical standards (Supplementary Material 1) injected in one randomized triplicate in each analytical batch, in order to assess the intra-batch and the inter-batch drift of RT and mass.

In the end, the analytical batch consisted of 128 injections including 84 injections of DOM samples ($n=28$, in triplicates), 24 injections of QCP, 17 injections of blanks and 3 injections of QC mix.

2.5. LC-HRMS data treatment workflow

2.5.1. Features list generation and intensities correction

The acquired data was processed using the Workflow4Metabolomics (W4M) platform (XCMS package) (Giacomoni et al., 2014). The raw data were previously converted into an mzXML format with MSconvert Proteowizard software. The conversion parameters were as follows: peakPicking with a continuous wavelet transform $\text{cwt snr} = 0.1$, $\text{peakSpace} = 0.1$ and $\text{msLevel} = 1$. Once the data were converted and loaded on the W4M platform, an automated workflow was applied to produce the aligned feature list (all steps are detailed in Supplementary Material 2). The first step of the workflow consisted of extracting the masses and chromatographic peaks. For this, the CentWave algorithm was used with a maximum tolerance in m/z deviation at 8.0 ppm and a peak width set at 5.2 seconds. The signal-to-noise threshold was set at 3 to generate a maximum number of peaks. The second step consisted of grouping the peaks with the algorithm PeakDensity based on a Bandwidth of 10.0, a minimum fraction of samples and a minimum number of samples of 1 and slices of width overlapping m/z of 0.01. Before proceeding areas integration of peaks, we applied a RT correction step with `retcor` function in order to correct drift between retention times of the

peaks detected in all the samples. For this, we applied a loess alignment with 0.5 degree of smoothing and the minimum fraction of samples was set at 0.01. After the RT correction, it was necessary to repeat the grouping of the peaks by reducing the Bandwidth to 5.0. The next step was to integrate areas of missing peaks using fillPeaks function. The number of decimal places for mass values was set at 5. The annotation of adducts and isotopes was performed with CAMERA algorithm (Kuhl et al., 2012), with the following input parameters: a full width half maximum FWHM of 0.6, general ppm error at 5.0 and an absolute error in m/z set at 0.005. For isotope annotation, the maximum ion charge was set at 4 and the number of expected isotopes at 6. The default parameters and extended list of adducts (Supplementary Material 3) were used for the adducts annotation. With these parameters, CAMERA allowed to annotate up to 4 adducts (e.g. $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$) and up to 4 isotopes (e.g. monoisotopic peak $[M]$, $[M+1]$, $[M+2]$, $[M+3]$), as belonging to the molecular compound. The last step of the workflow was the correction of the instrumental drift of intensities performed from the intensities of features in the QCP and using the Batch_correction function of W4M (a loess regression model with a span of 0.7).

2.5.2. Features selection steps using the quality control filters

The features list obtained from W4M was processed by three successive filtering steps (Figure 4-13) in order to make the information more reliable:

(i) The first step consisted of processing the features annotated as adducts and isotopes. The CAMERA algorithm allows to annotate several features, as belonging to the same molecular compound by using the retention time and the peak shape information (Kuhl et al., 2012). It is therefore possible, to group the features belonging to the same molecular compound. In our study, two types of data processing were tested: the first one by taking into account the maximum value of intensity in the set of adducts and isotopes corresponding to the same molecular compound; and the second one by summing up the intensities of the adducts and isotopes corresponding to the same molecular compound.

(ii) The second step consisted of removing the molecular compounds with intensities close to those of the injection blanks. For this, a threshold intensity was calculated for each compound, based on the mean + 3 x standard deviation of the intensities measured in the injection blanks. A molecular compound was kept only if its intensity was higher than the threshold intensity in

at least three random sample injections. At the end of this step, the excluded compounds were studied separately to verify their irrelevance to characterize DOM samples (e.g. high variation between samples in triplicates, no differentiation between QCs and DOM samples).

(iii) The third step consisted of removing the molecular compounds with high variability in their intensities. For each molecular compound, relative standard deviation (RSD) was calculated considering QCP injections (n=24) and randomized samples triplicates (for n=3 x 28 water samples). A first restrictive filter was tested, consisting of removing the compounds with RSD higher than 30 % in QCP injections and then in each samples triplicate. A second data processing was tested, less restrictive, with threshold at 50 % instead of 30 %.

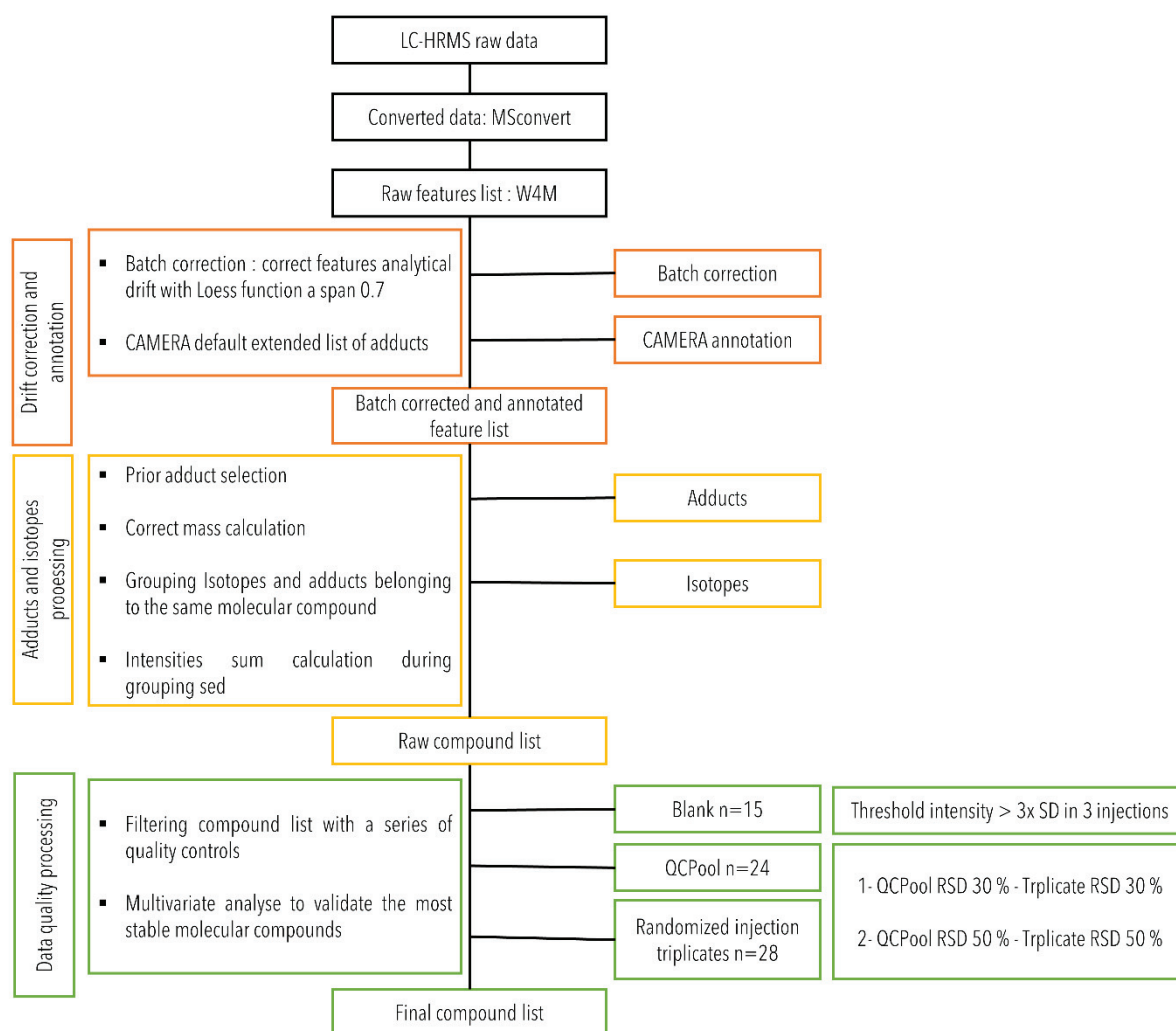


Figure 4-13. LC-HRMS data processing steps.

2.6. Statistical tests for validation of the data quality control treatment

During isotopes and adducts grouping step (Section 2.5.2 – step (i)), univariate statistical test ANOVA analysis coupled with a Tukey test (packages stats and rstatix; RStudio 4.0.2) was used to determine which process was more appropriate for our study. The next filtering steps were validated through multivariate analyses, such as clustering and principal component analysis (package factoextra, RStudio 4.0.2). Indeed, these tests allow to differentiate the samples according to their nature and type (e.g. injection blanks, QCs).

3. Results and discussion

3.1. Optimization of the LC-HRMS data treatment workflow: 3 steps toward an effective identification of DOM source markers

3.1.1. Processing of adducts and isotopes after grouping step:

The features list generated on W4M and annotated by CAMERA algorithm included 18752 molecular features. Within these features, 5858 (31 %) were annotated as isotopes and 2923 (16 %) as adducts. After isotope and adduct grouping step, the number of features was reduced to 13762 molecular compounds. The number of features, initially annotated as isotopes went from 5858 to 2252 molecular compounds and the number of features initially annotated as adducts from 2923 to 1175 molecular compounds. The question remained however on the intensity to be taken into account during the grouping. There are softwares like *MZmine* that perform isotope grouping taking into account either the maximum intensity or the intensity of the lowest m/z isotope (Nürenberg et al., 2015). For adducts, there is no particular grouping available in open source softwares for LC-HRMS data processing. A recent study by Stricker et al.(2021) proposed a non-chromatography-based annotation software that sums the intensity of all peaks related to one or more annotations. From an analytical aspect, the formation of adducts is still a little known phenomenon (Gonzalez-Riano et al., 2021). Mobile phases, ionization parameters (ESI source configuration) and particular matrices such as biological or environmental samples (Schug and McNair, 2003; Erngren et al., 2019; Lu et al., 2020) can affect the adducts formation. With CAMERA algorithm, adducts annotation is based on “features group levels” using peak grouping after retention time and a peak shape correlation (Kuhl et al., 2012). There has to be at least two matching adducts for

the same molecular compound in order to get an annotation. We have therefore assumed that if a given molecular compound has been ionized with a $[M+H]^+$ adduct and for analytical reasons has also been ionized with a different adduct such as $[M+Na]^+$, its intensity would correspond to the sum of the features $[M+H]^+$ and $[M+Na]^+$ intensities.

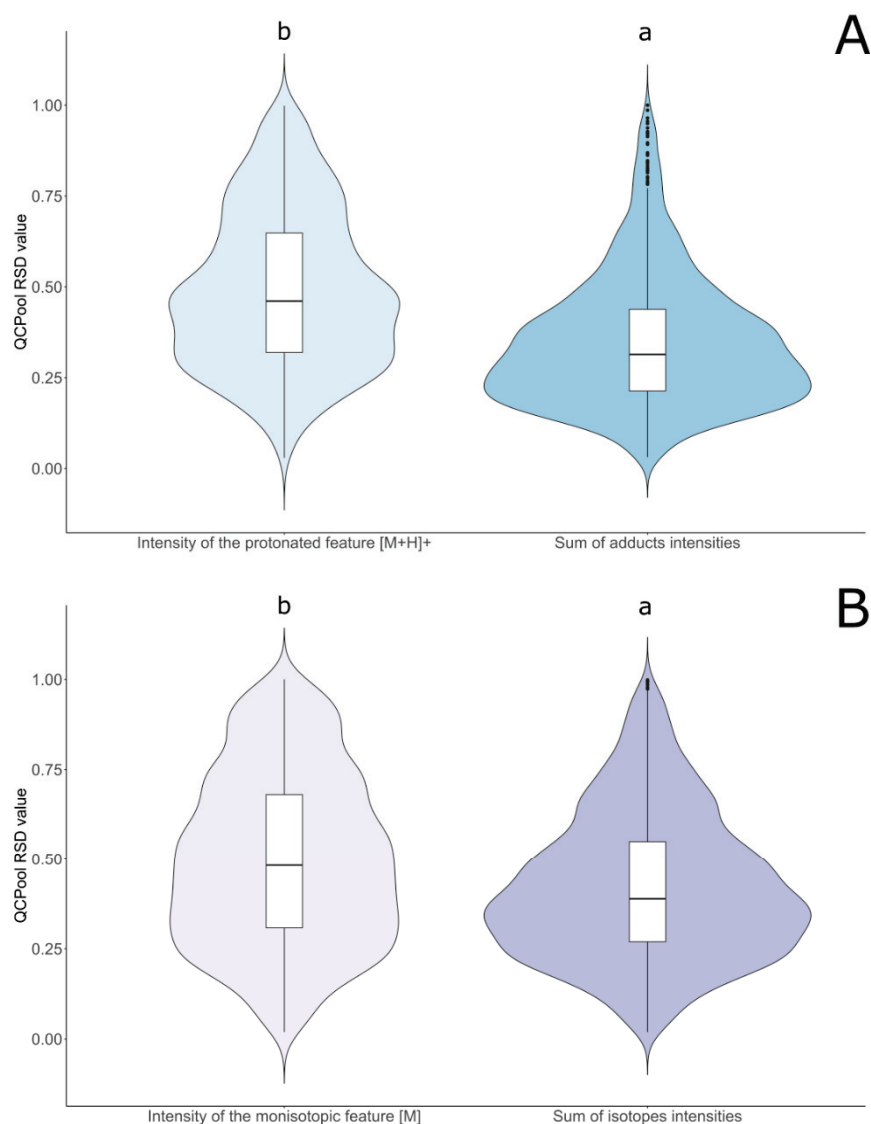


Figure 4-14. Violin plots of A. Relative standard deviation (RSD) of adduct intensities in QCP before (left) and after (right) intensity sum. B. RSD of isotope intensities in QCP before (left) and after (right) isotopes intensity sum.

As explained in part 2.5.2 (step i), for each molecular compound, we compared two strategies related to the intensities: one based on the sum of intensities of adducts or isotopes, versus another one based on adducts or isotopes maximum intensity. We then compared which strategy was the most suitable by calculating the relative standard deviations (RSD) of each molecular compound intensity in the QCP (n=24 injections). Indeed, QCP injections refer to

the same composite sample, collected in the same vial and analyzed under the same conditions throughout the LC-HRMS acquisition batch.

The results (Figure 4-14) show that summing, for a given molecular compound, the intensities of all isotopes or adducts allowed to significantly (Tuckey test, $p < 0.05$) decrease the RSD (from 0.54 to 0.40 for the median of isotopes and from 0.48 to 0.32 for the median of adducts). It means that the intensity of the molecular compounds is higher and less variable when all associated features are summed. For the isotopes, the relative intensity of the monoisotopic peak varied between 78 and 86% (with 100% corresponding to the total intensity summed for all isotopes). The maximum intensity, most of the time associated with the monoisotopic peak [M], is more variable than the sum of isotopes intensities, for a given molecular compound. Then summing all intensities features associated to a given molecular compound provides a more accurate representation of the abundance of that compound and allows to reduce bias and data noise. Since adducts are a compound of interest with chemical functional groups that have been bound to the compound by physical or chemical interactions, and since isotopes are atoms of the same element, then summing the intensities reduced bias and data noise that may have been present initially. For the rest of the data treatment steps, we then processed data by summing intensities.

The mass and RT information for a given molecular compound were processed considering the following conditions:

- For isotope grouping: we considered the mass and RT of the monoisotopic peak only.
- For multiple adducts annotations (e.g. [M+H]⁺, [M+Na]⁺, [M+NH₄]⁺): we considered the mass and RT of the "priority" adduct, i.e. the adduct having the most probability to be formed during the ionization in our analytical conditions (Section 2.3). For this, we considered the list of adducts used by the CAMERA algorithm (Supplementary Material 3). The mass was calculated from the m/z of the compound subtracting the mass of the adduct.
- For molecular features with no annotation: the neutral mass was calculated by removing the mass of [M+H]⁺ adduct (1.007276). Indeed, despite the complexity of the adduct formation phenomenon, in ESI⁺ analysis and for a mobile phase containing

formic acid, the most common adducts formed are protonated adducts (Erngren et al., 2019).

3.1.2. Setting up a minimum intensity threshold based on injection blanks

The second step of data treatment consisted of deleting compounds with intensities close to the injection blanks intensities (e.g. false positives). In the literature, some studies used an intensity threshold corresponding to the sample-to-blank areas ratio (Verkh et al., 2018b; Peter et al., 2019). In our study, we considered the relatively low intensities of compounds contained in natural water samples injected without pre-concentration. This led us to remove 40 % of the molecular compounds, with intensity lower than our threshold (5532 out of 13762 compounds). Among the 5532 deleted compounds, 57 % have an intensity lower than 300 cps, which corresponds to the median intensity of the features in the injection blanks, 24 % have an intensity between 300 and 1000 cps and 19 % have an intensity higher than 1000 cps. The compounds with very high intensities in the injection blanks (> 1000 cps) correspond to interference peaks probably related to compounds found in the vials (plastic cap liners, syringe filters, glassware etc.), in the mobile phase and in the instrumentation (column dumping, residues in the ultrapure water and organic solvents).

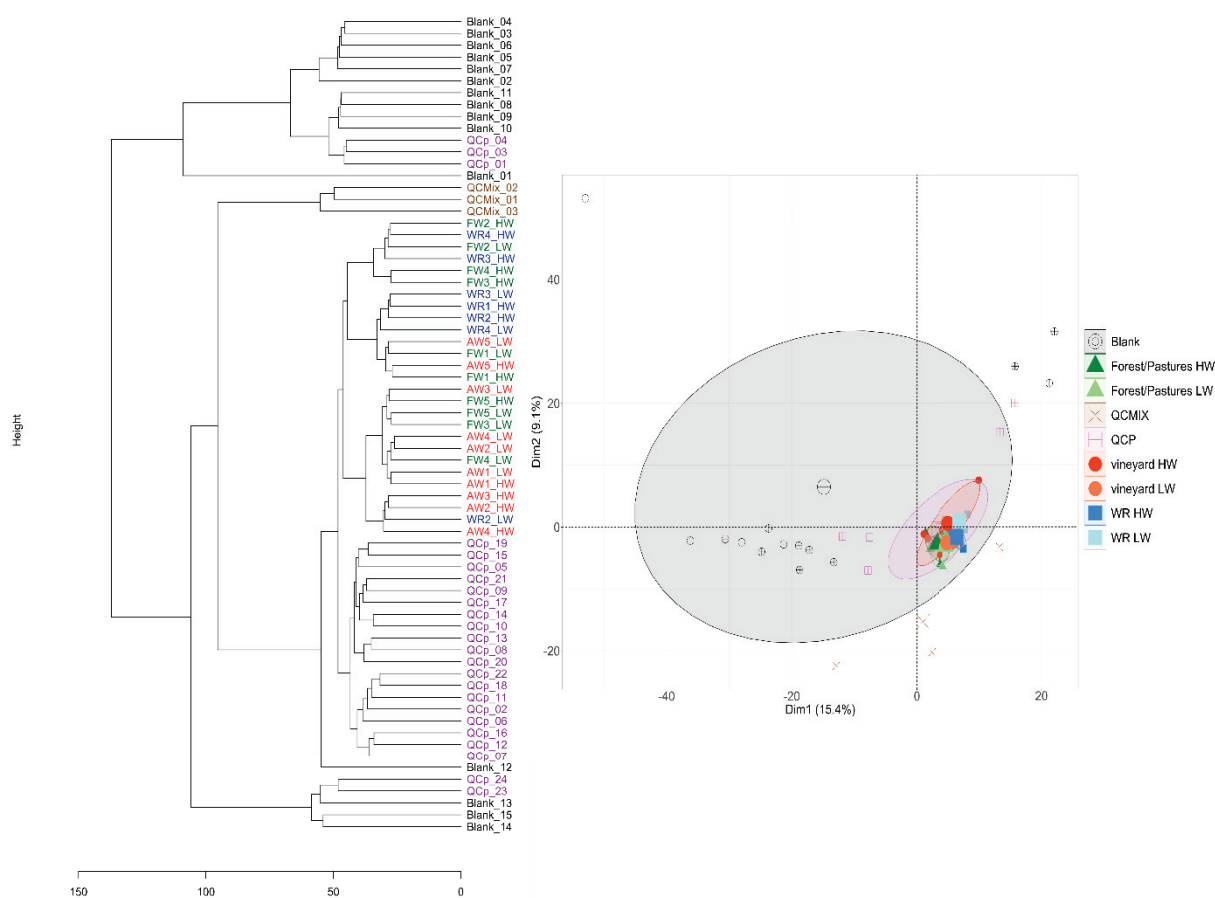


Figure 4-15. First factorial plane of the principal component analysis (PCA) performed on the compounds with intensities lower than the threshold based on injection blanks (n=5532).

A principal component analysis (PCA) was performed to check that the deleted molecular compounds were not of interest for our purpose. Indeed, the first factorial plan of the PCA, explaining 24 % of total data variance (Figure 4-15), highlights that these deleted compounds do not show relevance to discriminate samples according to DOM sources nor to distinguish the specificity of the quality controls. Indeed, the injection blank samples are scattered on both axes and the DOM samples, as well as the other QCP and QC mix quality controls are grouped in the same pool. A hierarchical classification confirms this conclusion, by showing a disparity in the classification of quality controls, with two main clusters formed: a first cluster containing quality controls (blanks, QCP, QC mix) grouped with DOM samples; and a second cluster containing only quality controls (blanks and a single QCP). Based on these results, we confirm that the 5532 molecular compounds can be deleted. The majority of these compounds are present at low intensities, highly variable, which explains the absence of

sample type clusters. Moreover, in a perspective of searching for DOM source markers or establishing specific fingerprints, it is necessary to have molecular compounds present at relatively high intensities.

3.1.3. Quality Control Pool and samples triplicates: useful for a final data clean-up based on an intensity variability threshold

After suppression of compounds with intensities lower than the threshold based on injection blanks, we have 8230 remaining molecular compounds. The final data-filtering step consisted of deleting molecular compounds with highly variable intensities throughout the acquisition. For this, we calculated values of RSD on the intensities of each molecular compounds (i) in the QCP injections (n=24), and subsequently (ii) in each sample randomized triplicates (n= 3 x 28). As explained in part 2.4.2 (step iii), we tested one RSD threshold at 30 % and another one, less restrictive, at 50 %. If the RSD exceeded the threshold in all QCP injections and in at least 75% of the samples triplicates, the compound was removed (Figure 4-16).

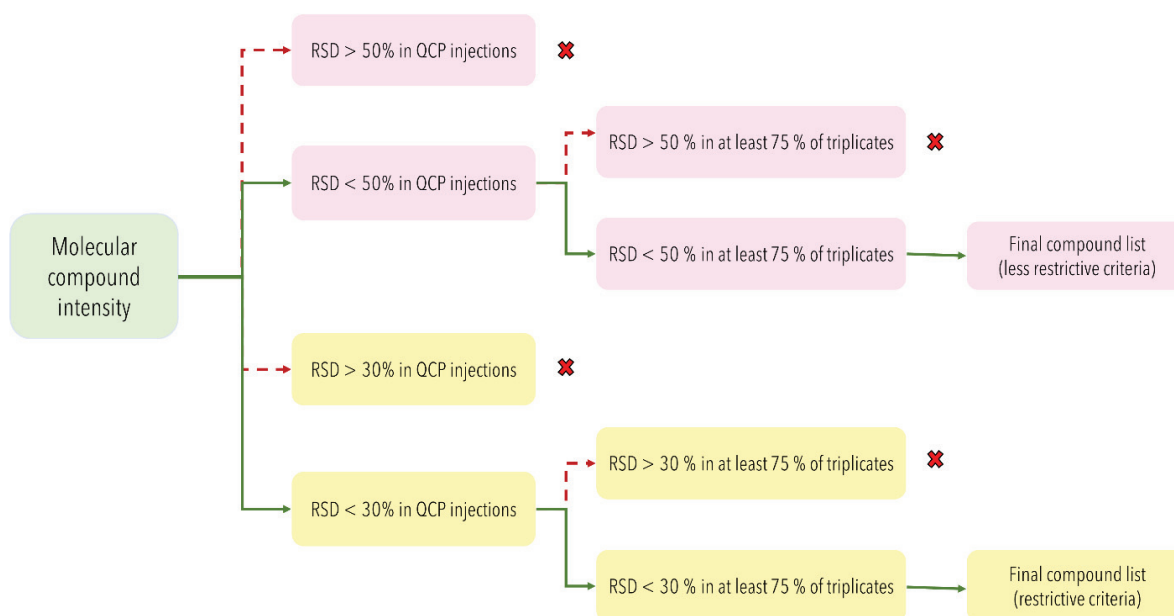


Figure 4-16. Data processing steps of removal of compounds with highly variable intensities in QCP (n=24) and in each triplicate injection (n=28 triplicates). The red cross means that the compounds were deleted at that stage.

Applying the less restrictive threshold (RSD at 50 %) reduced the number of molecular compounds from 8230 to 1478 (removal of 82 % of the compounds). Figure 4-17 shows a PCA

of the DOM samples, as well as the quality controls with the 1478 molecular compounds retained after this process. The groups of samples and quality controls are clearly distinguished between them, with the injection blanks and QC mix close in terms of variability. DOM samples seem separated according to the hydrological conditions, with a first pool of samples taken in high water period and a second pool of samples taken in low water period. The hierarchical classification of this dataset (HAC) (Figure 4-18.A) puts in light 4 clusters. A first cluster made up of injection blanks and QC mix, in which the first blank injection is distinct from the others. All the QC mix injections were clustered in this first cluster. The second cluster consisted of DOM samples taken during high water periods. The third cluster was mostly composed of QCP sample injections and three DOM samples taken during low water period, and the last cluster was formed only of DOM samples taken during low water periods.

Applying a RSD threshold at 30% drastically reduced the number of molecular compounds from 8230 to 457 (removal of 94 % of the compounds). In the PCA analysis (Figure 4-17.B), a better distinction was noticed between the injection blanks and QC mix quality controls, compared to the RSD threshold at 50 %. Even though the number of molecular compounds is 3 times less, the distinction between samples is still observable with distinct groups of DOM samples according the hydrological conditions.

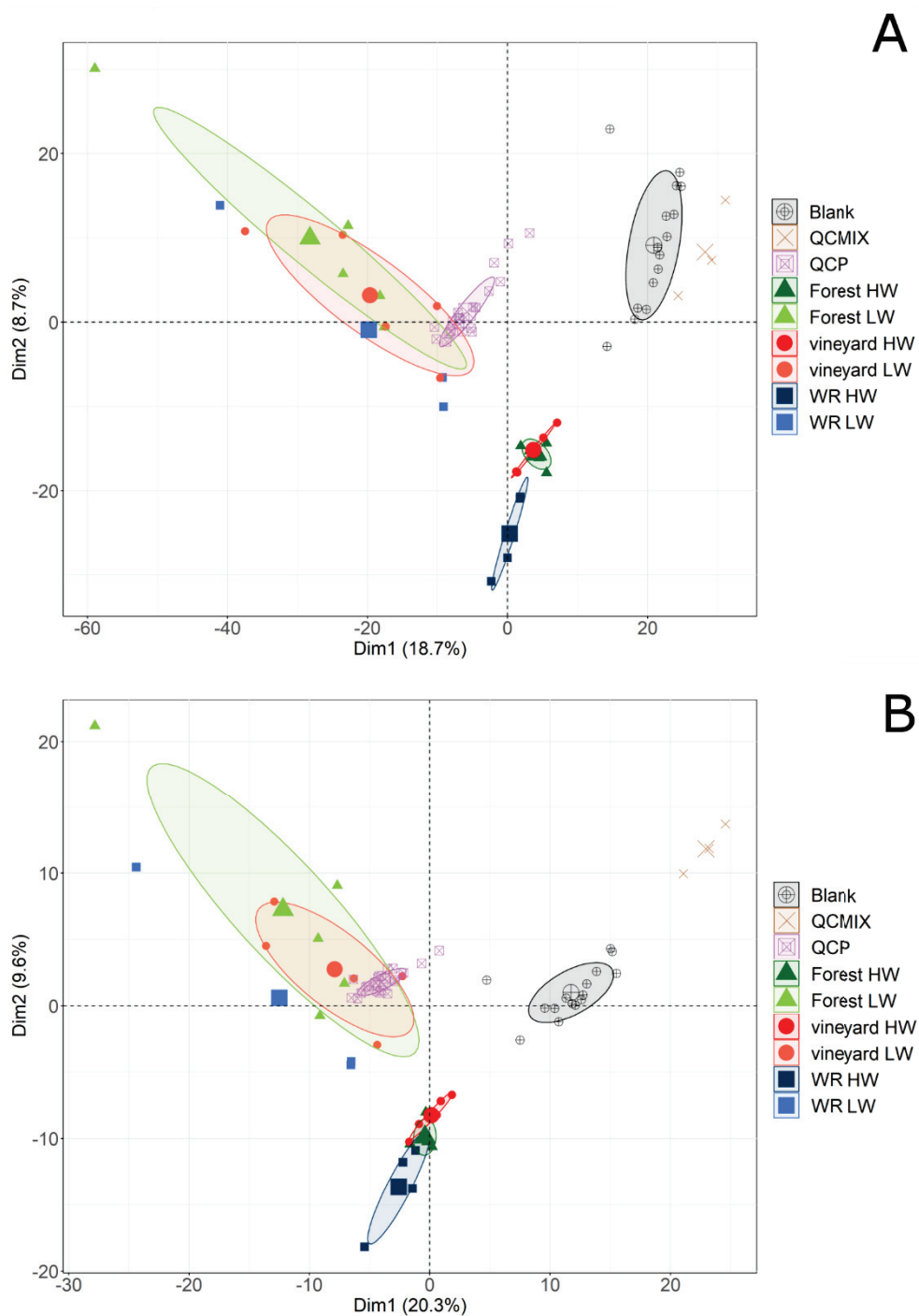


Figure 4-17. First factorial plan of the principal component analysis (PCA) of quality controls and DOM samples. A. Data processing with a maximum threshold at 50% for the relative standard deviation (RSD) on compound intensities in QCP and randomized triplicates samples (number of molecular compounds, $n = 1478$). B. Data processing with a maximum threshold at 30% for the RSD on compound intensities in QCP and randomized triplicates samples ($n = 457$).

The HAC (Figure 4-18.B) puts in light 6 main clusters of samples. A first cluster is formed by the QC mix injections and a second one by the injection blanks. The first blank injection (Blank_01) is still distinct from the others. This could be explained by interfering signals

generated by the LC-HRMS system, as Blank_01 was the first injection of the analytical batch. These interferences can be due to residual contamination from previous analytical batches, despite the stabilization of the instruments before the analysis of our samples. Five molecular compounds show significantly higher intensities in the Blank_01 injection, compared to the other blank injections. These interfering compounds are characterized by a high hydrophobicity, since they are eluted at the end of the chromatographic separation, when the percentage of organic solvent is the higher (98% of acetonitrile). The third cluster is formed mainly by DOM samples taken during high water periods and two samples taken in water reservoirs during low water periods (WR2_LW and WR3_LW). The fourth cluster contains only two DOM samples taken during low water periods, and the fifth cluster contains the rest of DOM samples taken during low water periods. Finally, the sixth group is fully composed of QCP injections, which correspond to the same composite sample injected several times during the whole LC-HRMS acquisition.

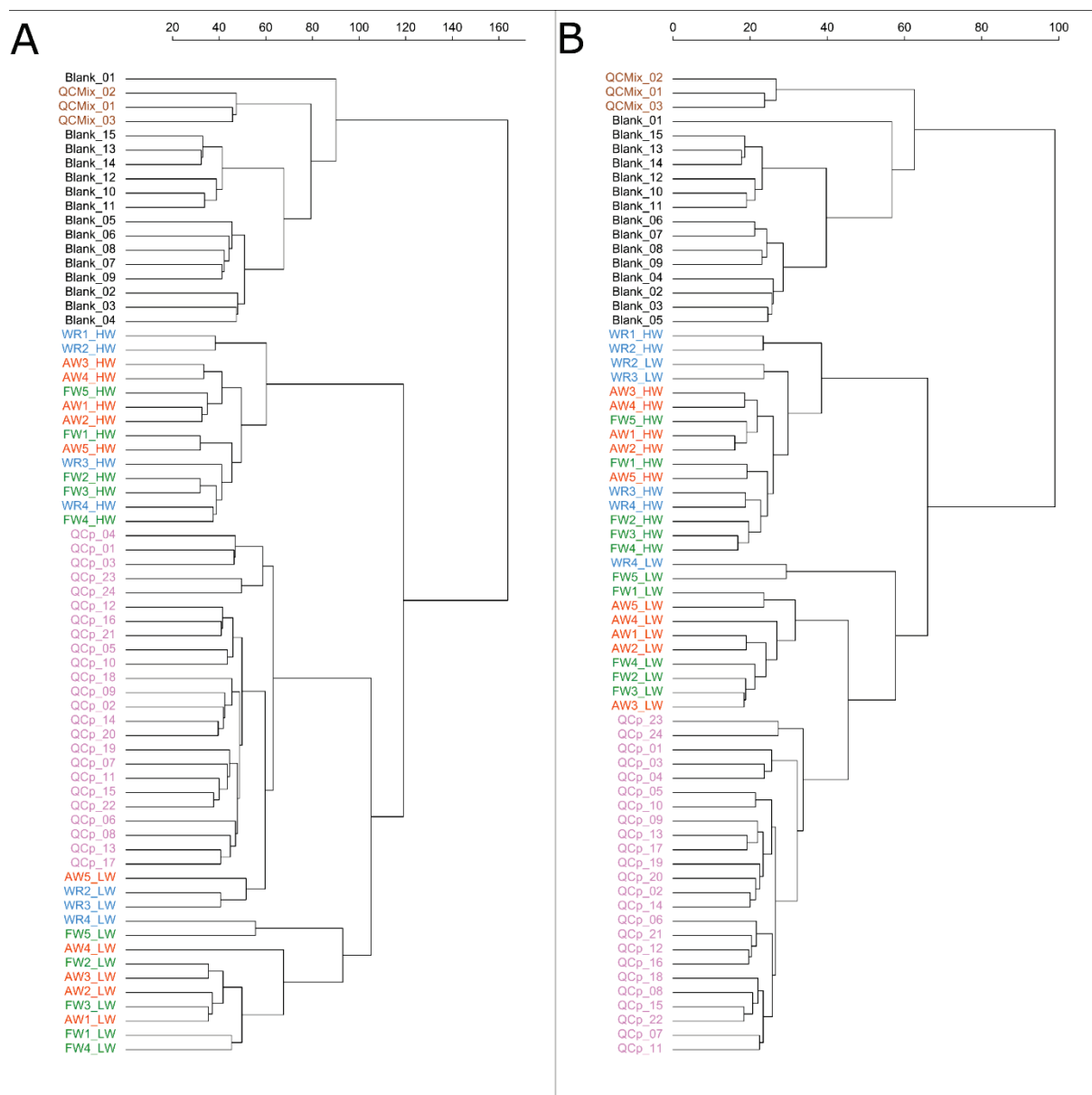


Figure 4-18. Ascendant hierarchical classification (HAC) of quality controls and DOM samples. A. Data processing with a maximum threshold at 50% for the relative standard deviation (RSD) on compound intensities in QCP and randomized triplicates (number of molecular compounds; $n = 1478$). B. Data processing with a maximum threshold at 30% for the RSD on compound intensities in QCP and randomized triplicates ($n = 457$).

To summarize, the compounds removed in this part (RSD thresholds at 30 % or at 50 %) are those whose intensity vary significantly. The batch correction of intensities applied at the very beginning of the workflow (Section 2.5.1) intended to correct the instrumental drift of the intensities during analyses. But, this correction cannot be perfect. Indeed, the instrumental drift can be explained by false positive peaks generated by softwares during the data acquisition steps. These peaks are often generated at the chromatogram construction stage from the extracted ion chromatograms (EIC). Previous studies have shown that the cwt

algorithm, used for mass extraction and detection of chromatographic peaks, was responsible for the detection of a significant number of false positives (Myers et al., 2017). Another explanation could be related to the structure of the compounds themselves, containing multiple functional groups that can lead to different ionization pathways, resulting in different ions fragments during the LC-HRMS acquisition. Finally, the chromatographic separation or the mass detection are characterized by an instrumental variability (during chromatographic separation and ionization) explaining a part of the variability of the compound intensity. Both thresholds, at 30 and 50%, allow to well distinguish between our various samples types, and we chose the most restrictive one (threshold at 30%) for our purpose on DOM sources study.

3.2. What information does the final list of molecular compounds provide on the DOM sources?

To search for markers of diffuse DOM sources in streams, we assumed that both land use and hydrological conditions influence the DOM quality (Guarch-Ribot and Butturini, 2016; Boukra et al., 2023). In periods of high water, streams should be more impacted by land use and terrigenous DOM sources than in periods of low water (Duan et al., 2007). Furthermore, the contribution of autochthonous or even underground sources during low water periods should be more apparent (da Silva et al., 2021; Hosen et al., 2021). The molecular compound list obtained after our data processing (463 molecular compounds) shows different signatures of DOM samples when grouped according to the hydrological conditions (Figure 4-19.A). As expected and given the very steep slopes that favors surface runoffs in the Ardières-Morcille watershed, the differences observed between high and low water periods can be mainly explained by surface runoffs of soil organic matter during high waters. This terrigenous contribution decreases during low water periods, giving way to a more autochthonous DOM produced locally, in watercourses. Previous studies have demonstrated a difference in the molecular and optical properties of DOM driven from terrigenous inputs versus autochthonous river DOM (Murphy et al., 2008; Liu et al., 2020).

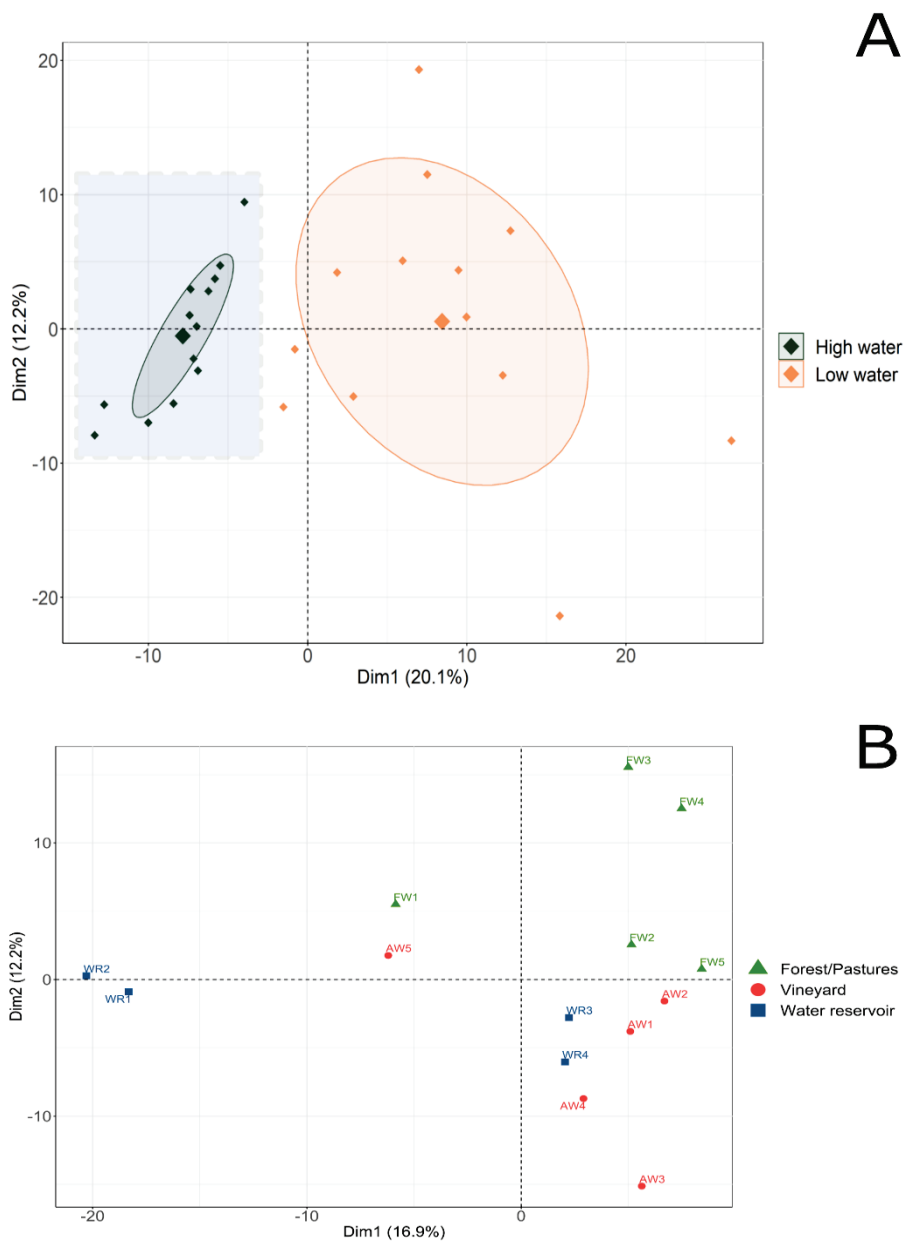


Figure 4-19. First factorial plan of principal component analysis (PCA) performed on compounds intensities ($n=463$ molecular compounds). A. The data correspond to DOM samples collected during high and low water periods. B. The data correspond to DOM samples collected only during high water periods

The difference is partly due to the presence of compounds less aliphatic, more oxidized with low lipidic and high lignin content in terrigenous DOM compared to the autochthonous DOM (Liu et al., 2020). Other molecular markers such as ratios of *n*-alkanes, fatty acids and steroids are referenced in the literature to differentiate between terrigenous DOM sources and DOM locally produced in rivers (Derrien et al., 2017). The molecular information provided by the LC-HRMS in our study, besides differentiating the types of DOM sources, can provide additional information to characterize the nature of DOM based on molecular mass and polarity

(inversely correlated to the chromatographic retention time in our analytical conditions). Overall, samples collected from streams during high water periods are characterized by non-polar molecular compounds of relatively low mass compared to samples collected during low water periods.

By studying only the samples taken during high water periods, the next question was to establish whether these compounds could mark differences in terms of land uses. Figure 4-19.B shows a PCA of high water samples plotted according to molecular compounds intensities (n=463). In this Figure 4-19.B, the samples are grouped according to the land use upstream the sampling points (forests/pastures, vineyards, water reservoirs). The distribution of samples on the PCA shows difference between samples taken downstream of forests/pastures vs. vineyards. In addition to the intensity of the compounds that allowed plotting the PCA, the mass and polarity (using retention time) information can help to understand the difference between the sample groups. Among the molecular compounds in the final list, a higher contribution of non-polar compounds of relatively low mass is observed in forests/pastures samples compared to vineyards ones. These non-polar compounds could partly result from the degradation of the plants and the process of DOM humification in the soils. Indeed, a previous study reported DOM from forestry sources as mostly composed of polyphenolics, condensed aromatics and highly unsaturated compounds (Spencer et al., 2019). Downstream of the vineyard land use, we hypothesize the presence of different organic contaminants related to the cultivation of the soils (as pesticides and their degradation/transformation products) (Moschet et al., 2014). Some of the pesticides and transformation products are persistent during their transfer towards the watercourse and thus can mark the land uses (Belles et al., 2019).

Moreover, the classification of point AW4 with the vineyard group can be explained by its immediate proximity with intense vineyard activities. The PCA also shows that samples FW1 and AW5 have relatively close molecular signatures. This can be explained by the geographical proximity between these two points and then a relatively similar contribution in terms of land uses. The water samples collected in the water reservoirs have rather heterogeneous signatures: Samples WR3 and WR4 have signatures close to the vineyard ones, contrary to WR1 and WR2. Land uses do not allow explaining DOM fingerprints in WR1 and WR2 samples; indeed, they are located in sub-watersheds that drain significant vineyard contributions, while

they are different from the vineyard sample group. Various specific processes of biodegradation, photodegradation, primary production and maturation of organic matter in these reservoirs can explain this result (Liu et al., 2020). This shows that the strategy implemented in this work to evaluate the contribution of water reservoirs, is not relevant. Indeed, the multiplicity of inputs, the long residence time, as well as the numerous biogeochemical processes in such stagnant environments, induce such variability in the composition of the DOM, that it is not possible to simply explain their molecular fingerprints according to the land use around.

4. Conclusion

In this study, we proposed a data treatment workflow to ensure the reliability of the data generated by LC-HRMS analysis in order to study molecular fingerprints of diffuse terrigenous and autochthonous DOM sources in watercourses. This workflow includes a correction of the features intensities drift occurring during LC-HRMS acquisition, and a strategy to annotate and consider the isotopes and adducts of molecular compounds. Quality controls, such as injection blanks, QC Pool and sample triplicates, are then used to identify reliable molecular compounds characterized by high and stable intensities. More precisely, a first data-filtering step using a threshold based on injection blanks intensities, allowed to remove the background noise compounds (representing 40 % of the compounds). Then, a second step, based on a threshold of intensity RSD considering QCP and samples triplicates, allowed to remove molecular compounds with too variable intensities (representing 82 % of the remaining compounds, for an intensity RSD threshold at 30%). The final list of molecular compounds can be considered as stable and reliable for the study of DOM sources in watercourses.

Using our LC-HRMS data treatment workflow, we were able to put in light that the molecular characteristics of DOM were mainly influenced by hydrological conditions: terrigenous contribution during high water periods and autochthonous production (with underground contribution) during low water periods. An influence of the land use was also observed on the samples taken in high water periods, with the presence of compounds specific to the vineyard and forest land uses. The establishment of fingerprints specific to DOM sources (the identification of some of the compounds composing these fingerprints), would allow to better understand the impact of anthropogenic sources in watercourses at the scale of watersheds

(e.g. agricultural activities, urban runoffs, storm overflows ...). However, further work is required to fully utilize LC-HRMS non-targeted analysis for environmental applications, especially regarding matrices effects and inter-batch variations of features intensity and mass. Moreover, a possible future approach to study DOM sources could be to propose markers combining information on mass, retention time, and intensity.

5. References

- Aguilar-Alarcón, P., Gonzalez, S.V., Simonsen, M.A., Borrero-Santiago, A.R., Sanchís, J., Meriac, A., Kolarevic, J., Asimakopoulos, A.G., Mikkelsen, Ø., 2020. Characterizing changes of dissolved organic matter composition with the use of distinct feeds in recirculating aquaculture systems via high-resolution mass spectrometry. *Science of The Total Environment* 749, 142326. <https://doi.org/10.1016/j.scitotenv.2020.142326>
- Aiken, G.R., 2014. 1.11 - Dissolved Organic Matter in Aquatic Systems, in: Ahuja, S. (Ed.), *Comprehensive Water Quality and Purification*. Elsevier, Waltham, pp. 205–220. <https://doi.org/10.1016/B978-0-12-382182-9.00014-1>
- Asmala, E., Massicotte, P., Carstensen, J., 2021. Identification of dissolved organic matter size components in freshwater and marine environments. *Limnology and Oceanography* 66, 1381–1393. <https://doi.org/10.1002/lno.11692>
- Belles, A., Alary, C., Rivière, A., Guillon, S., Patault, E., Flipo, N., Franke, C., 2019. Transfer Pathways and Fluxes of Water-Soluble Pesticides in Various Compartments of the Agricultural Catchment of the Canche River (Northern France). *Water* 11, 1428. <https://doi.org/10.3390/w11071428>
- Boukra, A., Masson, M., Brosse, C., Sourzac, M., Parlanti, E., Miège, C., 2023. Sampling terrigenous diffuse sources in watercourse: Influence of land use and hydrological conditions on dissolved organic matter characteristics. *Science of The Total Environment* 162104. <https://doi.org/10.1016/j.scitotenv.2023.162104>
- Brezinski, K., Gorczyca, B., 2019. An overview of the uses of high performance size exclusion chromatography (HPSEC) in the characterization of natural organic matter (NOM) in potable water, and ion-exchange applications. *Chemosphere* 217, 122–139. <https://doi.org/10.1016/j.chemosphere.2018.10.028>
- Bride, E., Heinisch, S., Bonnefille, B., Guillemain, C., Margoum, C., 2021. Suspect screening of environmental contaminants by UHPLC-HRMS and transposable Quantitative Structure-Retention Relationship modelling. *Journal of Hazardous Materials* 409, 124652. <https://doi.org/10.1016/j.jhazmat.2020.124652>
- Brock, O., Helmus, R., Kalbitz, K., Jansen, B., 2020. Non-target screening of leaf litter-derived dissolved organic matter using liquid chromatography coupled to high-resolution mass spectrometry (LC-QTOF-MS). *European Journal of Soil Science* 71, 420–432. <https://doi.org/10.1111/ejss.12894>
- Cai, M.-H., Wu, Y.-P., Ji, W.-X., Han, Y.-Z., Li, Y., Wu, J.-C., Shuang, C.-D., Korshin, G.V., Li, A.-M., Li, W.-T., 2020. Characterizing property and treatability of dissolved effluent organic matter using size exclusion chromatography with an array of absorbance, fluorescence, organic nitrogen and organic carbon detectors. *Chemosphere* 243, 125321. <https://doi.org/10.1016/j.chemosphere.2019.125321>
- Catalá, T.S., Shorte, S., Dittmar, T., 2021. Marine dissolved organic matter: a vast and unexplored molecular space. *Appl Microbiol Biotechnol* 105, 7225–7239. <https://doi.org/10.1007/s00253-021-11489-3>

- Cheng, C., Zhang, F., Shi, J., Kung, H.-T., 2022. What is the relationship between land use and surface water quality? A review and prospects from remote sensing perspective. *Environ Sci Pollut Res* 29, 56887–56907. <https://doi.org/10.1007/s11356-022-21348-x>
- Chon, Kangmin, Chon, Kyongmi, Cho, J., 2017. Characterization of size fractionated dissolved organic matter from river water and wastewater effluent using preparative high performance size exclusion chromatography. *Organic Geochemistry* 103, 105–112. <https://doi.org/10.1016/j.orggeochem.2016.11.003>
- Chow, A.T.-S., Ulus, Y., Huang, G., Kline, M.A., Cheah, W.-Y., 2022. Challenges in quantifying and characterizing dissolved organic carbon: Sampling, isolation, storage, and analysis. *Journal of Environmental Quality* 51, 837–871. <https://doi.org/10.1002/jeq2.20392>
- Chow, C.W.K., Fabris, R., Leeuwen, J. van, Wang, D., Drikas, M., 2008. Assessing Natural Organic Matter Treatability Using High Performance Size Exclusion Chromatography. *Environ. Sci. Technol.* 42, 6683–6689. <https://doi.org/10.1021/es800794r>
- da Silva, M.P., Blaurock, K., Beudert, B., Fleckenstein, J.H., Hopp, L., Peiffer, S., Reemtsma, T., Lechtenfeld, O.J., 2021. Delineating Source Contributions to Stream Dissolved Organic Matter Composition Under Baseflow Conditions in Forested Headwater Catchments. *Journal of Geophysical Research: Biogeosciences* 126, e2021JG006425. <https://doi.org/10.1029/2021JG006425>
- Derrien, M., Yang, L., Hur, J., 2017. Lipid biomarkers and spectroscopic indices for identifying organic matter sources in aquatic environments: A review. *Water Research* 112, 58–71. <https://doi.org/10.1016/j.watres.2017.01.023>
- Duan, S., Bianchi, T.S., Sampere, T.P., 2007. Temporal variability in the composition and abundance of terrestrially-derived dissolved organic matter in the lower Mississippi and Pearl Rivers. *Marine Chemistry* 103, 172–184. <https://doi.org/10.1016/j.marchem.2006.07.003>
- Elovaara, S., Eronen-Rasimus, E., Asmala, E., Tamelander, T., Kaartokallio, H., 2021. Contrasting patterns of carbon cycling and dissolved organic matter processing in two phytoplankton–bacteria communities. *Biogeosciences* 18, 6589–6616. <https://doi.org/10.5194/bg-18-6589-2021>
- Erngren, I., Haglöf, J., Engskog, M.K.R., Nestor, M., Hedeland, M., Arvidsson, T., Pettersson, C., 2019. Adduct formation in electrospray ionisation-mass spectrometry with hydrophilic interaction liquid chromatography is strongly affected by the inorganic ion concentration of the samples. *Journal of Chromatography A* 1600, 174–182. <https://doi.org/10.1016/j.chroma.2019.04.049>
- Fellman, J.B., Hood, E., Spencer, R.G.M., 2010. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnology and Oceanography* 55, 2452–2462. <https://doi.org/10.4319/lo.2010.55.6.2452>
- Gonzalez-Riano, C., Gradillas, A., Barbas, C., 2021. Exploiting the formation of adducts in mobile phases with ammonium fluoride for the enhancement of annotation in liquid chromatography-high resolution mass spectrometry based lipidomics. *Journal of Chromatography Open* 1, 100018. <https://doi.org/10.1016/j.jcoa.2021.100018>
- GORMLEY, K.L., TEATHER, K.L., GUIGNION, D.L., 2005. Changes in Salmonid Communities Associated with Pesticide Runoff Events. *Ecotoxicology* 14, 671–678. <https://doi.org/10.1007/s10646-005-0017-5>
- Gouy, V., Liger, L., Ahrouch, S., Bonnineau, C., Carluer, N., Chaumot, A., Coquery, M., Dabrin, A., Margoum, C., Pesce, S., 2021. Ardières-Morcille in the Beaujolais, France: A research catchment dedicated to study of the transport and impacts of diffuse agricultural pollution in rivers. *Hydrological Processes* 35, e14384. <https://doi.org/10.1002/hyp.14384>
- Guarch-Ribot, A., Butturini, A., 2016. Hydrological conditions regulate dissolved organic matter quality in an intermittent headwater stream. From drought to storm analysis. *Science of The Total Environment* 571, 1358–1369. <https://doi.org/10.1016/j.scitotenv.2016.07.060>

- Hassoun, H., Lamhasni, T., Foudeil, S., El Bakkali, A., Ait Lyazidi, S., Haddad, M., Choukrad, M., Hnach, M., 2017. Total Fluorescence Fingerprinting of Pesticides: A Reliable Approach for Continuous Monitoring of Soils and Waters. *J Fluoresc* 27, 1633–1642. <https://doi.org/10.1007/s10895-017-2100-8>
- Her, N., Amy, G., McKnight, D., Sohn, J., Yoon, Y., 2003. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. *Water Research* 37, 4295–4303. [https://doi.org/10.1016/S0043-1354\(03\)00317-8](https://doi.org/10.1016/S0043-1354(03)00317-8)
- Hernández, F., Ibáñez, M., Bade, R., Bijlsma, L., Sancho, J.V., 2014. Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry. *TrAC Trends in Analytical Chemistry, UHPLC: Where Are We 10 Years After Its Commercial Introduction?* 63, 140–157. <https://doi.org/10.1016/j.trac.2014.08.003>
- Hosen, J.D., Aho, K.S., Fair, J.H., Kyzivat, E.D., Matt, S., Morrison, J., Stubbins, A., Weber, L.C., Yoon, B., Raymond, P.A., 2021. Source Switching Maintains Dissolved Organic Matter Chemostasis Across Discharge Levels in a Large Temperate River Network. *Ecosystems* 24, 227–247. <https://doi.org/10.1007/s10021-020-00514-7>
- Huang, H., Sawade, E., Cook, D., Chow, C.W.K., Drikas, M., Jin, B., 2016. High-performance size exclusion chromatography with a multi-wavelength absorbance detector study on dissolved organic matter characterisation along a water distribution system. *Journal of Environmental Sciences* 44, 235–243. <https://doi.org/10.1016/j.jes.2015.12.011>
- Huang, Y., Du, Y., Ma, T., Deng, Y., Tao, Y., Xu, Y., Leng, Z., 2021. Dissolved organic matter characterization in high and low ammonium groundwater of Dongting Plain, central China. *Ecotoxicology and Environmental Safety* 208, 111779. <https://doi.org/10.1016/j.ecoenv.2020.111779>
- Kamjunke, N., Hertkorn, N., Harir, M., Schmitt-Kopplin, P., Griebler, C., Brauns, M., von Tümpling, W., Weitere, M., Herzsprung, P., 2019. Molecular change of dissolved organic matter and patterns of bacterial activity in a stream along a land-use gradient. *Water Research* 164, 114919. <https://doi.org/10.1016/j.watres.2019.114919>
- Kawasaki, N., Matsushige, K., Komatsu, K., Kohzu, A., Nara, F.W., Ogishi, F., Yahata, M., Mikami, H., Goto, T., Imai, A., 2011. Fast and precise method for HPLC–size exclusion chromatography with UV and TOC (NDIR) detection: Importance of multiple detectors to evaluate the characteristics of dissolved organic matter. *Water Research* 45, 6240–6248. <https://doi.org/10.1016/j.watres.2011.09.021>
- Kim, M.-S., Lim, B.R., Jeon, P., Hong, Seoyeon, Jeon, D., Park, S.Y., Hong, Sunhwa, Yoo, E.J., Kim, H.S., Shin, S., Yoon, J. ki, 2023. Innovative approach to reveal source contribution of dissolved organic matter in a complex river watershed using end-member mixing analysis based on spectroscopic proxies and multi-isotopes. *Water Research* 230, 119470. <https://doi.org/10.1016/j.watres.2022.119470>
- Krauss, M., Singer, H., Hollender, J., 2010. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal Bioanal Chem* 397, 943–951. <https://doi.org/10.1007/s00216-010-3608-9>
- Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T.R., Neumann, S., 2012. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* 84, 283–289. <https://doi.org/10.1021/ac202450g>
- Lanxiu, Y., Fengchang, W., Congqiang, L., Wen, L., Jing, W., Yi, M., 2004. Molecular weight distribution of dissolved organic matter in Lake Hongfeng determined by high performance size exclusion chromatography (HPSEC) with on-line UV-vis absorbance and fluorescence detection. *Chin. J. Geochem.* 23, 275–283. <https://doi.org/10.1007/BF02842075>
- Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., Pfister, H., 2014. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Visual. Comput. Graphics* 20, 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>

- Li, P., Hur, J., 2017. Utilization of UV-Vis spectroscopy and related data analyses for dissolved organic matter (DOM) studies: A review. *Critical Reviews in Environmental Science and Technology* 47. <https://doi.org/10.1080/10643389.2017.1309186>
- Li, Y., Gong, X., 2021. Effects of Dissolved Organic Matter on the Bioavailability of Heavy Metals During Microbial Dissimilatory Iron Reduction: A Review. *Rev Environ Contam Toxicol* 257, 69–92. https://doi.org/10.1007/398_2020_63
- Li, Y., Gong, X., Sun, Y., Shu, Y., Niu, D., Ye, H., 2022. High molecular weight fractions of dissolved organic matter (DOM) determined the adsorption and electron transfer capacity of DOM on iron minerals. *Chemical Geology* 604, 120907. <https://doi.org/10.1016/j.chemgeo.2022.120907>
- Liu, S., He, Z., Tang, Z., Liu, L., Hou, J., Li, T., Zhang, Y., Shi, Q., Giesy, J.P., Wu, F., 2020. Linking the molecular composition of autochthonous dissolved organic matter to source identification for freshwater lake ecosystems by combination of optical spectroscopy and FT-ICR-MS analysis. *Science of The Total Environment* 703, 134764. <https://doi.org/10.1016/j.scitotenv.2019.134764>
- Lu, W., Xing, X., Wang, L., Chen, L., Zhang, S., McReynolds, M.R., Rabinowitz, J.D., 2020. Improved Annotation of Untargeted Metabolomics Data through Buffer Modifications That Shift Adduct Mass and Intensity. *Anal. Chem.* 92, 11573–11581. <https://doi.org/10.1021/acs.analchem.0c00985>
- Mangal, V., DeGasparro, S., Beresford, D.V., Guéguen, C., 2020. Linking molecular and optical properties of dissolved organic matter across a soil-water interface on Akimiski Island (Nunavut, Canada). *Science of The Total Environment* 704, 135415. <https://doi.org/10.1016/j.scitotenv.2019.135415>
- Mitchell, P.J., Simpson, A.J., Soong, R., Simpson, M.J., 2018. Nuclear Magnetic Resonance Analysis of Changes in Dissolved Organic Matter Composition with Successive Layering on Clay Mineral Surfaces. *Soil Systems* 2, 8. <https://doi.org/10.3390/soils2010008>
- Mitschke, N., Vemulapalli, S.P.B., Dittmar, T., 2022. NMR spectroscopy of dissolved organic matter: a review. *Environ Chem Lett.* <https://doi.org/10.1007/s10311-022-01528-4>
- Moschet, C., Wittmer, I., Simovic, J., Junghans, M., Piazzoli, A., Singer, H., Stamm, C., Leu, C., Hollender, J., 2014. How a Complete Pesticide Screening Changes the Assessment of Surface Water Quality. *Environ. Sci. Technol.* 48, 5423–5432. <https://doi.org/10.1021/es500371t>
- Murphy, K.R., Stedmon, C.A., Waite, T.D., Ruiz, G.M., 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Marine Chemistry* 108, 40–58. <https://doi.org/10.1016/j.marchem.2007.10.003>
- Myers, O.D., Sumner, S.J., Li, S., Barnes, S., Du, X., 2017. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Anal. Chem.* 89, 8689–8695. <https://doi.org/10.1021/acs.analchem.7b01069>
- 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>
- Panton, A., Couceiro, F., Fones, G.R., Purdie, D.A., 2020. The impact of rainfall events, catchment characteristics and estuarine processes on the export of dissolved organic matter from two lowland rivers and their shared estuary. *Science of The Total Environment* 735, 139481. <https://doi.org/10.1016/j.scitotenv.2020.139481>
- Park, J.-H., 2009. Spectroscopic characterization of dissolved organic matter and its interactions with metals in surface waters using size exclusion chromatography. *Chemosphere* 77, 485–494. <https://doi.org/10.1016/j.chemosphere.2009.07.054>

- Parlanti, E., Morin, B., Vacher, L., 2002. Combined 3D-spectrofluorometry, high performance liquid chromatography and capillary electrophoresis for the characterization of dissolved organic matter in natural waters. *Organic Geochemistry* 33, 221–236. [https://doi.org/10.1016/S0146-6380\(01\)00154-1](https://doi.org/10.1016/S0146-6380(01)00154-1)
- Peter, K.T., Wu, C., Tian, Z., Kolodziej, E.P., 2019. Application of Nontarget High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. *Environ. Sci. Technol.* 53, 12257–12268. <https://doi.org/10.1021/acs.est.9b04481>
- Pluskal, T., Castillo, S., Villar-Briones, A., Oresic, 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>
- Schollée, J.E., Hollender, J., McArdell, C.S., 2021. Characterization of advanced wastewater treatment with ozone and activated carbon using LC-HRMS based non-target screening with automated trend assignment. *Water Research* 200, 117209. <https://doi.org/10.1016/j.watres.2021.117209>
- Schug, K., McNair, H.M., 2003. Adduct formation in electrospray ionization mass spectrometry: II. Benzoic acid derivatives. *Journal of Chromatography A, 25th International Symposium on Capillary Chromatography* 985, 531–539. [https://doi.org/10.1016/S0021-9673\(02\)01732-6](https://doi.org/10.1016/S0021-9673(02)01732-6)
- 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>
- Senan, O., Aguilar-Mogas, A., Navarro, M., Capellades, J., Noon, L., Burks, D., Yanes, O., Guimerà, R., Sales-Pardo, M., 2019. CliqueMS: a computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* 35, 4089–4097. <https://doi.org/10.1093/bioinformatics/btz207>
- Simpson, A.J., 2002. Determining the molecular weight, aggregation, structures and interactions of natural organic matter using diffusion ordered spectroscopy. *Magnetic Resonance in Chemistry* 40, S72–S82. <https://doi.org/10.1002/mrc.1106>
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78, 779–787. <https://doi.org/10.1021/ac051437y>
- Spencer, R.G.M., Kellerman, A.M., Podgorski, D.C., Macedo, M.N., Jankowski, K., Nunes, D., Neill, C., 2019. Identifying the Molecular Signatures of Agricultural Expansion in Amazonian Headwater Streams. *J. Geophys. Res. Biogeosci.* 124, 1637–1650. <https://doi.org/10.1029/2018JG004910>
- Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry* 82, 239–254. [https://doi.org/10.1016/S0304-4203\(03\)00072-0](https://doi.org/10.1016/S0304-4203(03)00072-0)
- Stricker, T., Bonner, R., Lisacek, F., Hopfgartner, G., 2021. Adduct annotation in liquid chromatography/high-resolution mass spectrometry to enhance compound identification. *Anal Bioanal Chem* 413, 503–517. <https://doi.org/10.1007/s00216-020-03019-3>
- Thieme, L., Graeber, D., Hofmann, D., Bischoff, S., Schwarz, M.T., Steffen, B., Meyer, U.-N., Kaupenjohann, M., Wilcke, W., Michalzik, B., Siemens, J., 2019. Dissolved organic matter characteristics of deciduous and coniferous forests with variable management: different at the source, aligned in the soil. *Biogeosciences* 16, 1411–1432. <https://doi.org/10.5194/bg-16-1411-2019>
- Verkh, Y., Rozman, M., Petrovic, M., 2018a. 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>
- Verkh, Y., Rozman, M., Petrovic, M., 2018b. Extraction and cleansing of data for a non-targeted analysis of high-resolution mass spectrometry data of wastewater. *MethodsX* 5, 395–402. <https://doi.org/10.1016/j.mex.2018.04.008>

- Wu, Y., Koch, B.P., Wang, Xiaona, Witt, M., Wang, Xiaolu, Bao, H., Gan, S., Kattner, G., Zhang, J., 2022. Molecular Composition of Dissolved Organic Matter in the Changjiang (Yangtze River) – Imprints of Anthropogenic Impact. *Frontiers in Marine Science* 9.
- Xue, Z., Lv, Z., Li, L., 2022. Combination of chromatographic and spectroscopic characterization based on primitive ultraviolet absorbance detection to fulfill advanced monitoring of dissolved organic matter in municipal wastewater treatment plant. *Journal of Environmental Chemical Engineering* 10, 107538. <https://doi.org/10.1016/j.jece.2022.107538>
- Zhu, Y., Vieth-Hillebrand, A., Noah, M., Poetz, S., 2019. Molecular characterization of extracted dissolved organic matter from New Zealand coals identified by ultrahigh resolution mass spectrometry. *International Journal of Coal Geology* 203, 74–86. <https://doi.org/10.1016/j.coal.2019.01.007>
- Zsolnay, Á., 2003. Dissolved organic matter: artefacts, definitions, and functions. *Geoderma, Ecological aspects of dissolved organic matter in soils* 113, 187–209. [https://doi.org/10.1016/S0016-7061\(02\)00361-0](https://doi.org/10.1016/S0016-7061(02)00361-0)

6. Supplementary material

Supplementary material 1. List of the QC mix compounds. The expected m/z and RT (min) were defined based on (Bride et al., 2021)

Component	Neutral mass (Da)	Expected m/z	Observed m/z	Mass diff (ppm)	Expected RT (min)	Observed RT (min)
Simazine	201.07812	202.085	202.0850	0.5	7.78	8.08
Diuron d6	238.05468	239.0616	239.0620	2	8.99	9.21
Atenolol	266.16304	267.1703	267.1688	5	4.57	4.92
Venlafaxine	277.20418	278.2113	278.2114	0.5	6.72	7.27
Tebuconazole	307.14514	308.1529	308.1540	4	10.76	10.88
Tiapride	328.14568	329.1526	329.1526	0.1	5.09	5.76
Bezafibrate	361.10809	362.1156	362.1136	5	9.64	10.85
Celiprolol	379.24711	380.2544	380.2546	2	6.44	6.96
Azoxystrobine	403.11682	404.1235	404.1243	2	10.6	10.68
Netilmicine	475.3006	476.3062	476.3047	3	5.8	6.33
Telmisartan	514.23688	515.2456	515.2424	6	8.21	8.63
Erythromycine	733.46124	734.4681	734.4686	0.7	7.47	7.93
Roxythromycine	836.52457	837.531	837.5310	0.04	8.35	8.83

Supplementary material 2. Applied workflow for the features list generation on W4M platform.

Step 1: xcms findChromPeaks (xcmsSet)

▪ **Spectra Filters:**

Filter on Acquisition Numbers: Empty.

Filter on Retention Time: Empty.

Filter on Ms: Empty.

Extraction method for peaks detection: CentWave - chromatographic peak detection using the centWave method

Max tolerated ppm m/z deviation in consecutive scans in ppm: 8.0

Min,Max peak width in seconds: 5,20

▪ **Advanced Options:**

Signal to Noise ratio cutoff: 3

Prefilter step for for the first analysis step (ROI detection): 3,100

Name of the function to calculate the m/z center of the chromatographic peak: intensity weighted mean of the peak's m/z values

Integration method: peak limits are found through descent on the mexican hat filtered data (more robust, but less exact)

Minimum difference in m/z for peaks with overlapping retention times: -0.001

Fitgauss : False

Noise filter : 0

verbose Columns : False

Get a list of found chromatographic peaks :

False

List of regions-of-interest (ROI):

List of regions-of-interest (ROI) representing detected mass traces: *select at runtime*

Is continuous data within regions is checked to be above the first baseline: True

Numeric vector defining the scale for each region of interest in 'roiList': Empty.

Resubmit your dataset or your zip file: no need

Step 2: xcms findChromPeaks Merger

RData file: Output dataset 'xsetRData' from previous step

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file: no need

Step 3: xcms groupChromPeaks (group)

RData file :Output dataset 'xsetRData' from previous step

Method to use for grouping: PeakDensity - peak grouping based on time dimension peak densities

Bandwidth: 10.0

Minimum fraction of samples: 0.01

Minimum number of samples: 1

Width of overlapping m/z slices: 0.01

Advanced Options:

Maximum number of groups to identify in a single m/z slice: 50

Get the Peak List: True

Convert retention time (seconds) into minutes: False

Number of decimal places for mass values reported in ions' identifiers: 5

Number of decimal places for retention time values reported in ions' identifiers: 0

Reported intensity values: into

If NA values remain, replace them by 0 in the dataMatrix: True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file: no need

Step 4: xcms adjustRtime (retcor)

RData file: Output dataset 'xsetRData' from step 6

Method to use for retention time correction : PeakGroups - retention time correction based on alignment of features (peak groups) present in most/all samples.

Minimum required fraction of samples in which peaks for the peak group were identified: 0.01

Maximal number of additional peaks for all samples to be assigned to a peak group for retention time correction: 1

Smooth method: loess - non-linear alignment

Advanced Options:

Degree of smoothing for the loess fitting: 0.5

Family: gaussian

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file: no need

Step 5: xcms groupChromPeaks (group)

RData file: Output dataset 'xsetRData' from previous step

Method to use for grouping: PeakDensity - peak grouping based on time dimension peak densities

Bandwidth: 5.0

Minimum fraction of samples: 0.01

Minimum number of samples: 1

Width of overlapping m/z slices: 0.01

Advanced Options:

Maximum number of groups to identify in a single m/z slice: 50

Get the Peak List: True

Convert retention time (seconds) into minutes: False

Number of decimal places for mass values reported in ions' identifiers: 5

Number of decimal places for retention time values reported in ions' identifiers: 0

Reported intensity values: into

If NA values remain, replace them by 0 in the dataMatrix: True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file: no need

Step 6: xcms fillChromPeaks (fillPeaks)

RData file: Output dataset 'xsetRData' from previous step

Advanced Options:

Value by which the m/z width of peaks should be expanded

0

Value by which the RT width of peaks should be expanded

0

Specifying a ppm by which the m/z width of the peak region should be expanded

0

Peak List:

None: True

Convert retention time (seconds) into minutes: False

Number of decimal places for mass values reported in ions' identifiers: 5

Number of decimal places for retention time values reported in ions' identifiers: 0

Reported intensity values: into

If NA values remain, replace them by 0 in the dataMatrix: True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file: no need

Step 7: CAMERA.annotate

RData file :Output dataset 'xsetRData' from previous step

Group co-eluted peaks based on RT [groupFWHM]:

Multiplier of the standard deviation :6

Percentage of FWHM width: 0.6

Annotation general options:

General ppm error: 5

General absolut error in m/z: 0.005

Annotate Isotopes [findIsotopes]:

Max. ion charge: 4

Max. number of expected isotopes: 6

The percentage number of samples, which must satisfy the C12/C13 rule for isotope annotation: 0.5

Mode: All functions

Verifying grouping co-eluted peaks [groupCorr]:

groupCorr: correlation threshold (0..1):0.75

groupCorr: Method selection for grouping peaks after correlation analysis into pseudospectra: hcs

groupCorr: significant correlation threshold: 0.05

groupCorr: Use correlation inside samples for peak grouping: True

groupCorr: Use isotopic relationship for peak grouping: False

groupCorr: Use correlation across samples for peak grouping: False

Annotate Adducts [findAdducts]:

Which polarity mode was used for measuring of the ms sample: positive

How much peaks will be calculated in every thread using the parallel mode: 100

Use a personal ruleset file: FALSE

If no ruleset is provided, calculate ruleset with max. number n of [nM+x] clusterions: 3

Statistics and results export: [diffreport]:

Number of condition: One condition

Export options:

Convert retention time (seconds) into minutes: False

Number of decimal places for mass values reported in ions' identifiers: 4

Number of decimal places for retention time values reported in ions' identifiers: 0

General used intensity value: into

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file: no need

Step 8: Batch correction

Data matrix file: Output dataset 'dataMatrix' from previous step

Variable metadata file: Output dataset 'variableMetadata' from previous step

Sample metadata file coding parameters:

Batch column name: batch

Injection order column name: injectionOrder

Sample type column name: sampleType

Type of regression model: loess

Span: 0.7

Supplementary material 3. The list of adducts used by CAMERA algorithm to annotate molecular features.

Name	nmol	charge	massdiff	oidscore	quasi	ips	
[M+H] ⁺	1	1	1.007276	1	1	1	1
[2M+H] ⁺	2	1	1.007276	1	0	0	0.5
[3M+H] ⁺	3	1	1.007276	1	0	0	0.25
[M+2H] ²⁺	1	2	2.014552	2	0	0	0.75
[2M+2H] ²⁺	2	2	2.014552	2	0	0	0.5
[3M+2H] ²⁺	3	2	2.014552	2	0	0	0.25
[M+3H] ³⁺	1	3	3.021828	3	0	0	0.75
[2M+3H] ³⁺	2	3	3.021828	3	0	0	0.5
[3M+3H] ³⁺	3	3	3.021828	3	0	0	0.25
[M+NH ₄] ⁺	1	1	18.033823	16	1	1	1
[2M+NH ₄] ⁺	2	1	18.033823	16	0	0	0.5
[3M+NH ₄] ⁺	3	1	18.033823	16	0	0	0.25
[M+H+NH ₄] ²⁺	1	2	19.041099	7	0	0	0.5
[2M+H+NH ₄] ²⁺	2	2	19.041099	7	0	0	0.5
[3M+H+NH ₄] ²⁺	3	2	19.041099	7	0	0	0.25
[M+Na] ⁺	1	1	22.989218	8	1	1	1
[2M+Na] ⁺	2	1	22.989218	8	0	0	0.5
[3M+Na] ⁺	3	1	22.989218	8	0	0	0.25
[M+H+Na] ²⁺	1	2	23.996494	4	0	0	0.5
[2M+H+Na] ²⁺	2	2	23.996494	4	0	0	0.5
[3M+H+Na] ²⁺	3	2	23.996494	4	0	0	0.25
[M+K] ⁺	1	1	38.963158	10	1	1	1
[2M+K] ⁺	2	1	38.963158	10	0	0	0.5
[3M+K] ⁺	3	1	38.963158	10	0	0	0.25
[M+H+K] ²⁺	1	2	39.970434	6	0	0	0.5
[2M+H+K] ²⁺	2	2	39.970434	6	0	0	0.5
[3M+H+K] ²⁺	3	2	39.970434	6	0	0	0.25
[M+2Na] ²⁺	1	2	45.978436	9	0	0	0.5
[2M+2Na] ²⁺	2	2	45.978436	9	0	0	0.5
[3M+2Na] ²⁺	3	2	45.978436	9	0	0	0.25
[M+Na+K] ²⁺	1	2	61.952376	11	0	0	0.5
[2M+Na+K] ²⁺	2	2	61.952376	11	0	0	0.5
[3M+Na+K] ²⁺	3	2	61.952376	11	0	0	0.25
[M+NH ₄ +HCOOH] ⁺	1	1	64.039293	89	0	0	0.25
[M+Na+HCOOH] ⁺	1	1	68.994688	87	0	0	0.25
[M+2K] ²⁺	1	2	77.926316	13	0	0	0.5
[2M+2K] ²⁺	2	2	77.926316	13	0	0	0.5
[3M+2K] ²⁺	3	2	77.926316	13	0	0	0.25

4.2.3 Exploitation des données LC-HRMS dans un contexte d'étude des sources de MOD dans les cours d'eau récepteurs

4.2.3.1 Indicateurs globaux basés sur la répartition des composés en masse et polarité

Après avoir acquis les données LC-HRMS non-ciblées et procédé au traitement et tri des données à travers les filtres qualités (cf. partie 4.2.2 de ce chapitre), la dernière étape consiste ici en l'exploitation des données dans un contexte de traçage de sources de MOD. Dans la partie précédente, la liste finale de composés moléculaires a directement été intégrée dans des analyses multivariées pour mettre en évidence des différences de caractéristiques entre les sources de MOD (Figure 4-20).

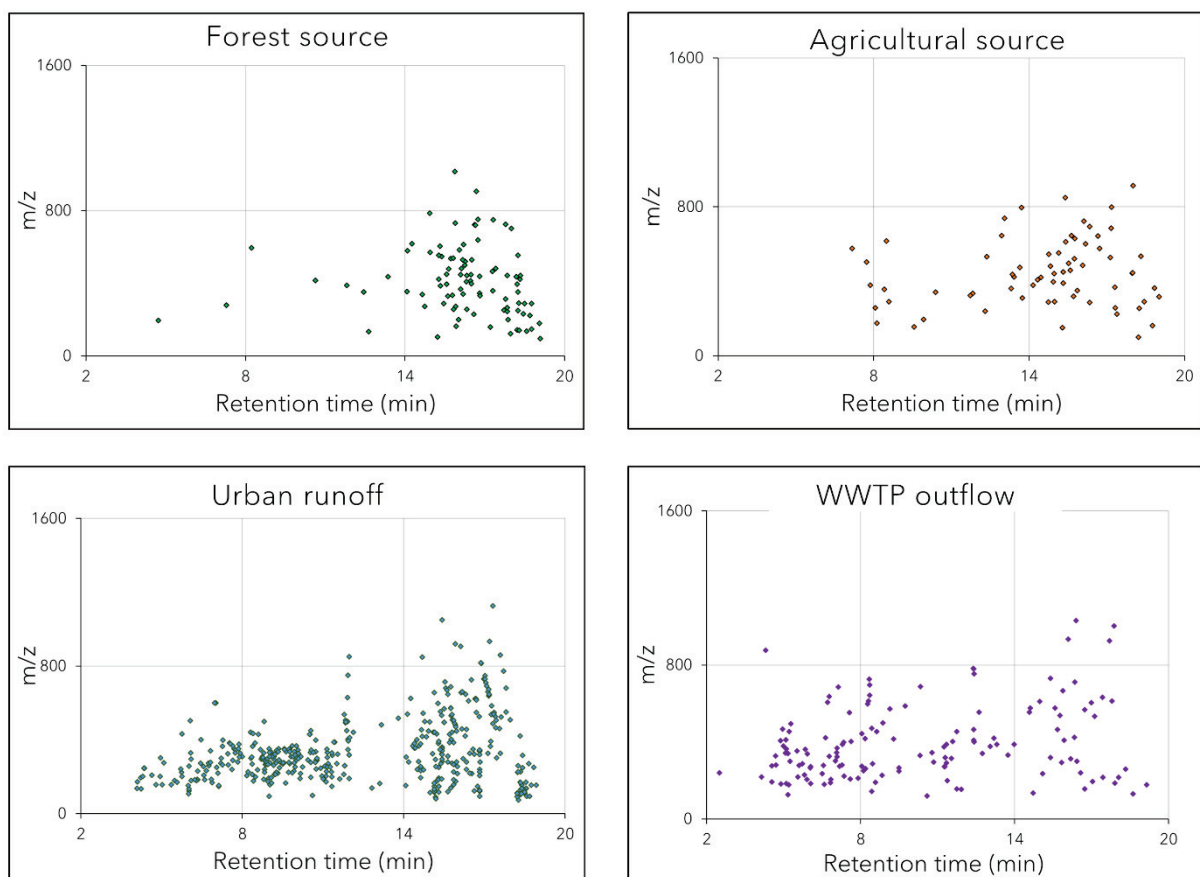


Figure 4-20. Exemples de la répartition en RT et m/z de composés moléculaires dans des échantillons de DOM issus de sources différentes

Dans cette partie, nous proposons une première piste d'exploitation de ces données HRMS, basée sur le calcul d'indicateurs globaux qui permettent d'intégrer les masses (m/z), les temps

de rétention chromatographique (RT, indicateur de la polarité, et les intensités (Figure 4-21) de composés moléculaires issues de plusieurs séquences analytiques. En effet, dans les conditions de séparation chromatographique en phase inverse choisies, les composés polaires éluent en début d'acquisition, contrairement aux composés apolaires qui éluent en fin d'analyse.

Les problématiques liées à la dérive instrumentale (dérive des temps de rétention RT et masse m/z), existant déjà à l'intérieur d'une même séquence analytique, sont beaucoup plus fortes entre différentes séquences. Cette dérive inter-séquence est un verrou qui doit être levé pour permettre l'exploitation de données sur un nombre important d'échantillons qui ne peuvent pas être analysés dans une seule séquence. La démarche inédite de traitement des données présentée dans cette partie, utilisant des indicateurs globalisés, peut y contribuer.

Pour mettre en place ces indicateurs, nous avons fractionné la matrice de répartition des composés en 9 zones délimitées en masse (notée MW pour la suite, pour *molecular weight*) et en RT (Figure 4-21). Dans chaque zone (notée RT_MW), la somme des intensités des composés a été calculée puis normalisée par la teneur en carbone organique dissous (DOC) de l'échantillon (Eq. 4.10). Cette normalisation permet de proposer des indicateurs qualitatifs de MOD, sur le même principe que pour les indicateurs UV-Visible ou EEM de fluorescence qui sont normalisés soit par une longueur d'onde d'absorbance UV, de fluorescence ou une concentration de DOC (e.g. SUVA, HIX, E2.E3).

$$RT_MW = \frac{\sum_{i=1}^{nF} intensity(cps)}{[DOC]} \quad \text{Eq 4.10}$$

Où RT et MW correspondent à des intervalles de temps de rétention et de masse et nF correspond au nombre de composés moléculaires contenus dans ces intervalles.

Un premier exemple de calcul de ces indicateurs (Figure 4-21) indique que les échantillons d'eaux naturelles prélevés en aval de sous-bassins versants forestiers semblent être principalement caractérisés par des composés dont la masse est très hétérogène, avec de fortes intensités pour les composés qui éluent en milieu et en fin d'acquisition (composés apolaires). Le même résultat est obtenu pour les eaux de ruissellement urbain. Cependant, les échantillons de ruissellement urbain des eaux naturelles présente une différence majeure au niveau de l'indicateur RT0.6_MW500-1000, qui correspond à la somme des composés dans la

zone de RT entre 0 et 6 min, et de masse entre 500 et 1000 Da. Ces composés sont en effet totalement absents des échantillons d'eaux naturelles, au contraire des entrées de STEU dans lesquelles ils sont présents à de très fortes intensités. Ces composés semblent être labiles et dégradables, car absents dans les eaux de sorties de STEU (après traitement des eaux).

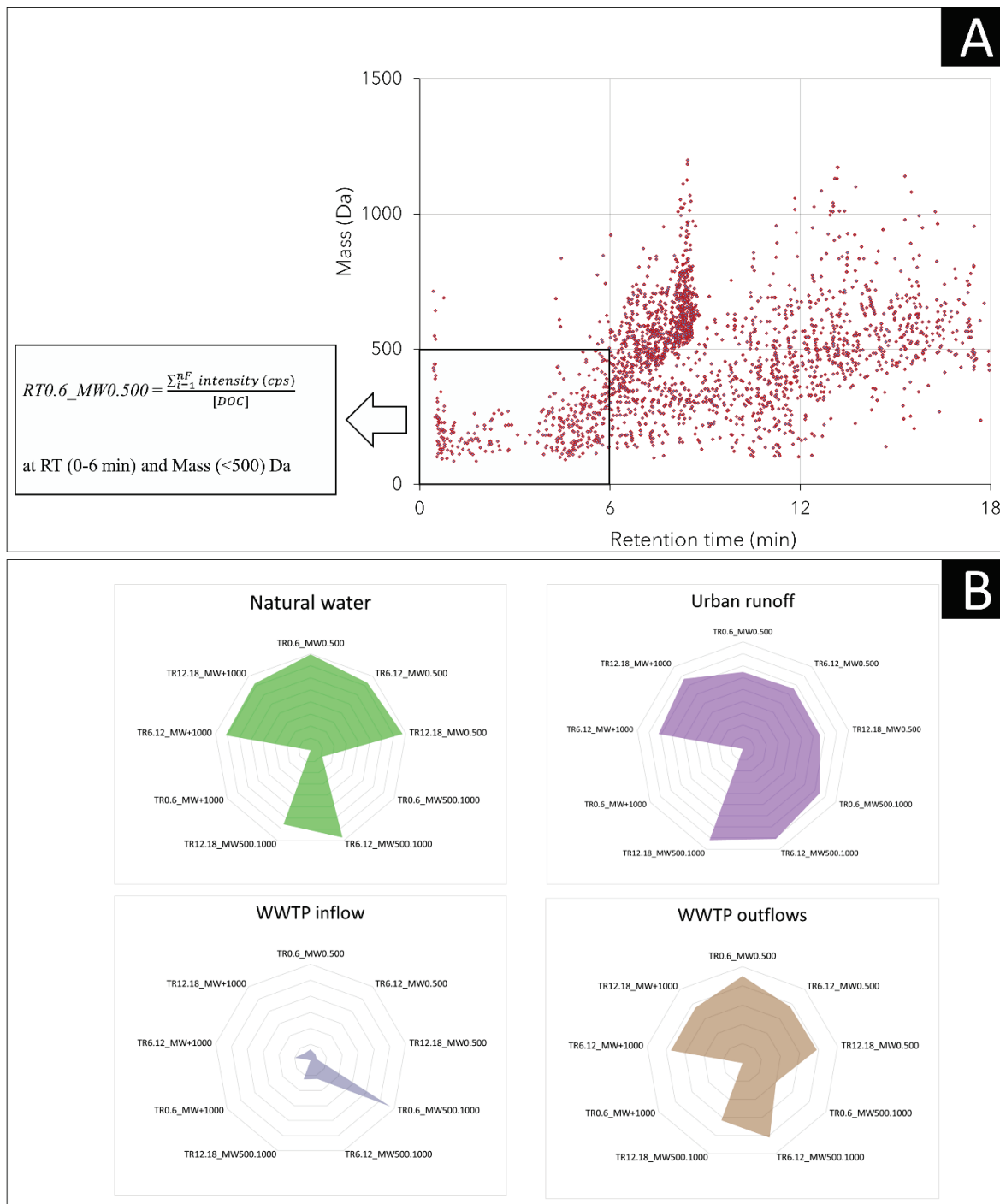


Figure 4-21. A. Illustration des indicateurs globaux LC-HRMS basés sur la répartition des composés moléculaires en fonction des temps de rétention, des masses et des intensités des composés. B. Exemple d'application sur des échantillons de DOM issues de sources différentes (e.g. eau naturelle en sortie de bassin forestier, ruissellement urbain, entrées et sorties de STEU)

4.2.3.2 Confrontation avec des indicateurs classiques optiques

Afin de mieux comprendre la portée de ces indicateurs, une matrice de corrélation avec des indicateurs classiques optiques a été réalisée sur un jeu de données correspondant à 135 échantillons de MOD représentatifs de l'ensemble des sources considérées dans cette thèse. Les indicateurs LC-HRMS ont été classés en fonction de la masse, avec un ordre décroissant en polarité. Les principaux enseignements sont les suivants :

- Les intensités des composés dont la masse est inférieure à 500 Da sont principalement corrélées aux rapports d'absorbances (E2.E3, E2.E4 et E3.E4), qui donnent une indication sur la taille des molécules organiques. Les échantillons de MOD dont les rapports d'absorbances sont élevés (présence de petites molécules peu humifiées) sont ceux qui contiennent des molécules inférieures à 500 Da avec des intensités élevées.
- L'indice SUVA, qui est un indicateur de l'aromaticité de la MOD, est inversement corrélé aux indicateurs LC-HRMS pour les fractions comprises entre 0 et 12 min et de masse supérieure à 500 Da. Cela indique que les échantillons dont la MOD est faiblement aromatique sont caractérisés par la présence de molécules polaires de masse relativement élevée et à de fortes intensités.
- La fraction RT_12.18 de l'ensemble des classes de masses (les molécules apolaires de masse variable) est négativement corrélée à l'indicateur de fluorescence FI. Pour rappel, de fortes valeurs de FI indiquent la présence d'une matière organique récemment produite et peu humifiée. Ceci est particulièrement le cas dans les eaux de STEU où des études ont rapporté la forte présence de matériel organique microbien pour expliquer les faibles valeurs de FI en entrée de STEU. De plus, les valeurs de FI élevées dans les eaux naturelles pourraient être attribuées à des composés fortement aromatiques et d'origine terrestre (Kim et al., 2023). Cela pourrait correspondre à des composés polycondensés, aromatiques et donc apolaires.
- La corrélation positive entre l'indicateur RT0.6_MW1000 et le FI confirme que les échantillons qui sont caractérisés par une MOD peu aromatique issue d'activité microbienne sont principalement composés de molécules polaires de poids supérieur à 1 kDa.

- Des corrélations positives entre le BIX et les indicateurs LC-HRMS représentatifs des masses supérieures à 500 Da indiquent qu'il s'agit probablement de composés de type protéique issus de l'activité microbienne. Les corrélations négatives entre le HIX et ces mêmes indicateurs LC-HRMS montrent que les indicateurs de fluorescence HIX ne permettent pas de caractériser les petites molécules inférieures à 500 Da. Il pourrait s'agir de molécules ne présentant pas de propriété particulière d'absorbance ni de fluorescence.

De manière générale, les indicateurs proposés dans cette partie sont complémentaires aux indicateurs optiques classiques. Ils permettent une interprétation approfondie de ces derniers en apportant des informations supplémentaires sur la composition moléculaire complexe des MOD.

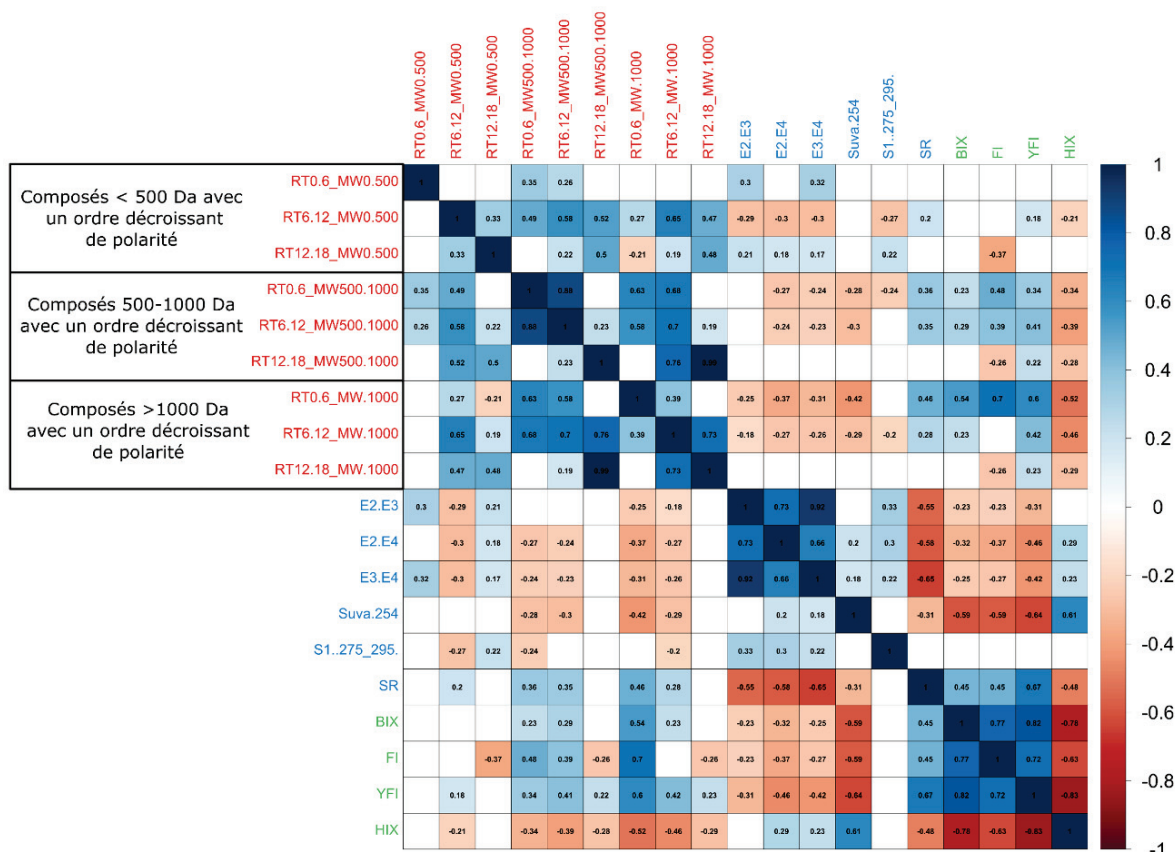


Figure 4-22. Matrice de corrélation entre les indicateurs globaux LC-HRMS (en rouge), les indicateurs UV-Visible (en bleu) et les indicateurs EEM de fluorescence (en vert)

4.2.3.3 Recherche de marqueurs moléculaires spécifiques à des types de sources de MOD

Une autre possibilité offerte avec l'exploitation des données LC-HRMS, dans le contexte d'étude de sources de MOD, est l'identification des composés moléculaires spécifiques à des types de sources. Pour cela, des diagrammes d'intersection (Figure 4-24) ont été réalisés sur la base de la présence/absence des composés dans des *pools* d'échantillons identiques (*e.g.* ruissellement urbain, sortie de STEU, etc.). Ces diagrammes d'intersection ont été réalisés à l'aide du package R (*upsetR*) (Lex et al., 2014). Dans un second temps, les composés ont été listés en exploitant plusieurs séquences analytiques à l'aide du package *metabcombineR*. Avec cette approche, il a été possible d'établir des listes de composés moléculaires spécifiques de types de sources de MOD.

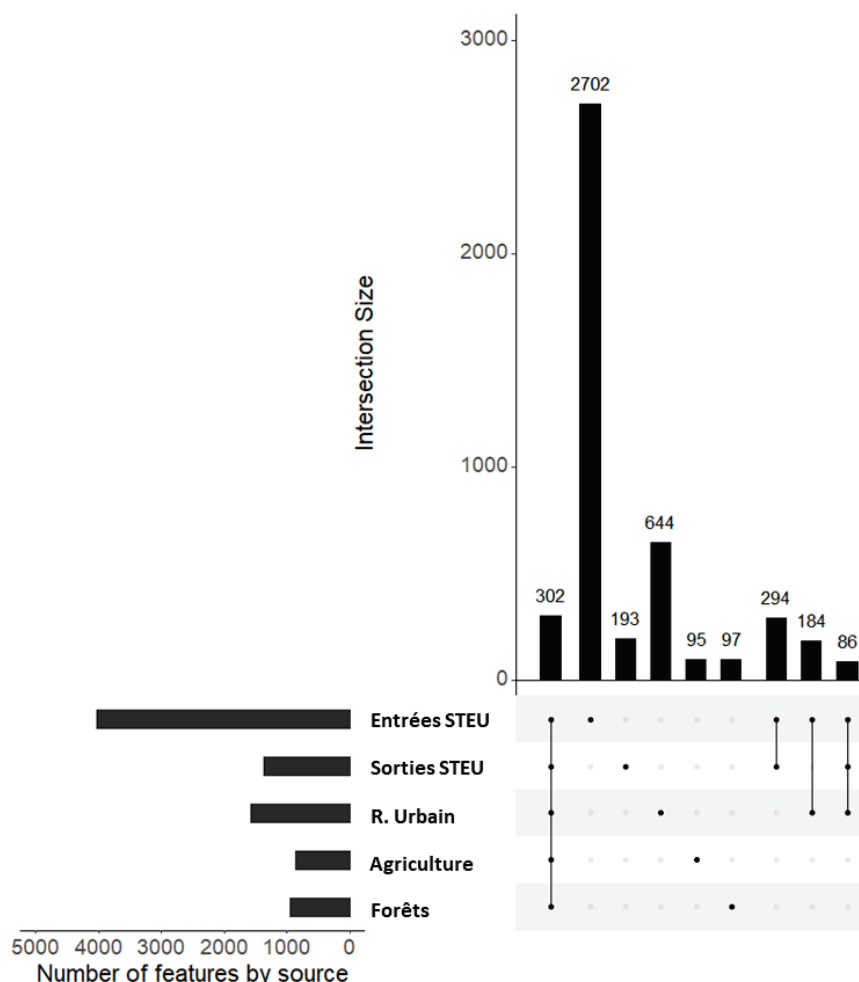


Figure 4-23. Diagramme illustrant le nombre d'intersections et de non-intersections entre les composés moléculaires détectés dans différents types de sources de DOM. Les points reliés par un trait représentent les intersections, *i.e.* les composés détectés dans deux ou plusieurs sources. Les points non reliés indiquent le nombre de composés spécifiques à une seule source donnée.

Le diagramme se lit dans un premier temps horizontalement, avec les intersections (histogrammes du haut avec le nombre de composés pour chaque intersection) entre les échantillons. Dans l'exemple présenté ici, on s'intéresse aux intersections entre des échantillons de MOD issus d'entrées et sorties de STEU, de ruissellement urbain, de sources agricoles et forestières (10 échantillons par source). Pour expliciter les intersections, des points ou des points reliés par un trait noir permettent une lecture verticale, avec indication du nombre de composés moléculaires par type d'échantillon. Si le point est relié à un autre point par un trait, cela signifie que les composés dénombrés sont communs aux échantillons reliés ; sinon, il s'agit de composés spécifiques à un seul échantillon. Ainsi, dans l'exemple de la Figure 4-23, le nombre de composés moléculaires spécifiques à une source unique est de 2 702 pour les eaux d'entrées de STEU et de 193 pour les sorties de STEU. Ces 193 composés n'ont donc pas été détectés dans les eaux d'entrée. Cela pourrait par exemple correspondre à des sous-produits de transformation générés lors des étapes de traitement. Les eaux de ruissellement urbain comprennent quant à elles 644 composés moléculaires spécifiques. Enfin, le nombre de composés spécifiques aux sources agricoles et forestières est de 95 et 97 respectivement. Dans un contexte de traçage, ces composés spécifiques à un type de source peuvent constituer de bons marqueurs de ces sources dans les milieux récepteurs. Les intersections entre deux ou plusieurs sources peuvent également aider à identifier des composés moléculaires d'intérêt. Par exemple, si l'on considère l'intersection entre les entrées et les sorties de STEU, avec un nombre de composés communs de 294, ces composés peuvent correspondre à des molécules réfractaires au traitement appliqué dans la STEU. Enfin, 302 composés moléculaires non-spécifiques (composés dits « universels », car communs à tous les types de sources) ont été identifiés.

4.2.4 Conclusion générale du chapitre 4

Ce chapitre avait pour but de proposer des indicateurs HPSEC/UV-fluorescence et LC-HRMS pertinents dans l'étude des sources de MOD. Dans un premier temps, une décomposition gaussienne des chromatogrammes HPSEC a permis de mettre en place des indicateurs basés sur 3 fractions en poids moléculaire de la MOD. Ces indicateurs correspondent à des pourcentages de contribution en petites, moyennes et grosses molécules (LMW, MMW et HMW), calculés à deux longueurs d'ondes UV (254 et 280 nm) et trois longueurs d'ondes d'excitation-émission de fluorescence (314-443, 280-330 et 278-354 nm). Ces indicateurs ont permis de différencier des types de MOD issues de sources diffuses. Une étude de corrélation a ensuite montré une complémentarité avec des indicateurs classiques UV-Visible et EEM de fluorescence, mais également la limite d'interprétation de certains indicateurs classiques. Le même principe de calcul pourrait être appliqué à d'autres longueurs d'onde d'intérêt pour mieux explorer le continuum de taille de la MOD (avec les 3 fractions de taille proposées) et améliorer les connaissances sur la composition et transformation de cette MOD à l'échelle des bassins versants.

La mise en place d'indicateurs LC-HRMS pour l'étude des sources a nécessité plusieurs étapes de traitement de données préalables, à commencer par la génération de la liste de *features*. Les logiciels MZmine et XCMS (versions 2.5 et 3.5) ont été comparés par rapport au nombre de *features* générés, à la détection des traceurs d'injections et des composés du QC mix et enfin par rapport à la cohérence de répartition des contrôles qualités après analyses multivariées. Les résultats ont montré que XCMS permet de générer un plus grand nombre de *features*, de retrouver un plus grand nombre de composés du QC mix et d'obtenir une répartition des contrôles qualités plus cohérente. Suite à ces résultats, le package XCMS a été utilisé pour générer les listes de *features* de l'ensemble des séquences de la thèse.

Après avoir généré une première liste de *features*, la deuxième étape a été d'appliquer un *workflow* de retraitement pour réduire le nombre de pics parasites ou faux positifs, en se basant principalement sur les intensités (niveau et variabilité) des composés. A travers une série de filtres qualité (seuils d'intensités et de variabilité RSD dans les blancs d'injections, QC pools et triplicats randomisés), il a été possible de réduire drastiquement le nombre de composés tout en gardant une information fiable et pertinente pour une étude de

construction d’empreinte de MOD issues de sources différentes. En effet, la liste finale de composés moléculaires a permis de proposer deux types d’exploitations des données HRMS. La première utilise une série de neuf indicateurs globaux basés sur la répartition en masse, polarité et intensité des composés moléculaires. Une étude de corrélation avec des indicateurs optiques classiques a montré une complémentarité des indicateurs HRMS avec les indicateurs des fractions chromophoriques et fluorescentes de la MOD. La deuxième exploitation des données a été d’identifier des composés moléculaires spécifiques à des types de sources de MOD. Pour un ensemble d’échantillons analysés dans une séquence et sur la base de la présence/absence des composés, il est ainsi possible d’explorer les spécificités et non-spécificités moléculaires par type de source.

Chapitre 5.

Etude des sources diffuses de matière organique dissoute.

Chapitre 5. Etude des sources diffuses de matière organique dissoute

Avant-propos

L'étude bibliographique (Chapitre 1) a montré la diversité des sources potentielles de MOD à l'échelle des bassins versants. Parmi elles, les sources diffuses représentent une part non négligeable de MOD dans les cours d'eau, notamment via les écoulements des eaux de pluie sur des sous-bassins présentant des types d'occupation des sol variés (*e.g.* agricoles, d'élevage, forestier, prairies...) et via les apports par les retenues d'eau. Ces sources peuvent apporter dans le cours d'eau un large éventail de composés organiques, y compris des pesticides, des produits pharmaceutiques et d'autres polluants qui peuvent avoir un impact sur la composition de la MOD et, au-delà, sur l'ensemble de l'écosystème aquatique. L'identification de ces sources reste cependant très difficile du fait de la multiplicité des apports dans les cours d'eau, de la complexité de la MOD ainsi que de la survenue d'évènements non contrôlés à l'origine de ces sources (*e.g.* précipitations atmosphériques, nature des sols, topographie du bassin versant). Dans ce chapitre, une première partie a pour objectif d'étudier la représentativité de l'échantillonnage directement dans les cours d'eau des sources diffuses de MOD liées à des occupations de sols typiques. Pour cela, une étude comparative entre la MOD prélevée dans les cours d'eau et celle extraite de sols adjacents aux cours d'eau a été réalisée. La deuxième partie du chapitre est une perspective d'application de la méthodologie proposée dans la Partie 1 dans deux autres bassins versants.

5.1 Échantillonnage des sources diffuses terrigènes dans les cours d'eau : influence de l'utilisation des sols et des conditions hydrologiques sur les caractéristiques de la matière organique dissoute (publication n°1, publiée).

5.1.1 Résumé de la publication scientifique

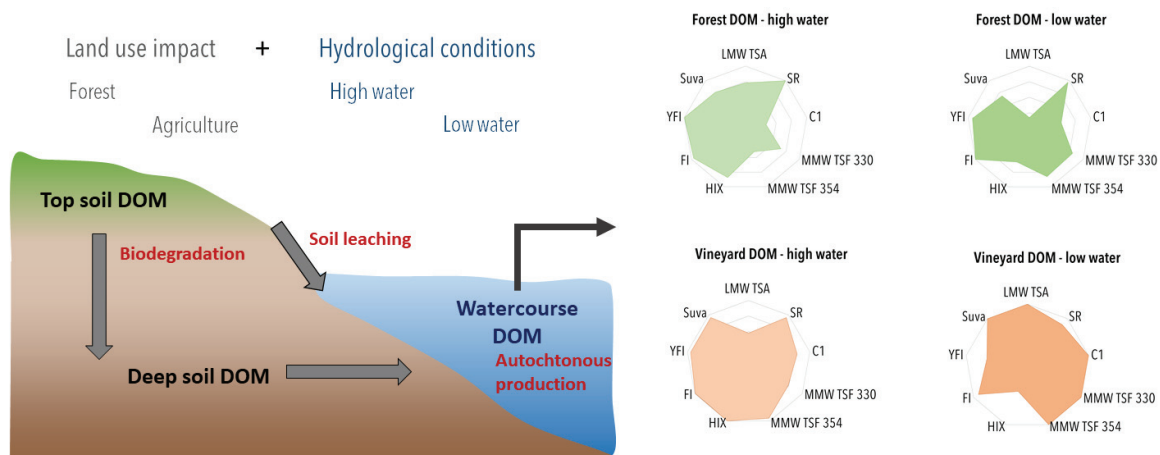
Pour déterminer la représentativité de l'échantillonnage d'une source diffuse terrigène de MOD dans les cours d'eau, nous avons étudié le lien entre la MOD extraite du sol, caractérisée par une occupation des sols donnée (forestiers/prairies et viticultures) et la MOD collectée dans le cours d'eau adjacent au sol considéré. La matière organique des sols a été extraite à

l'eau enrichie en CaCl_2 , puis biodégradée, suivant un protocole basé sur l'ajout d'inoculum naturel de sols (McDowell et al., 2006). La MOD des cours d'eau a été échantillonnée sur deux périodes en fonction des conditions hydrologiques du milieu, avec l'hypothèse d'un apport terrigène marqué en période de hautes eaux et une production autochtone plus présente en période de basses eaux durant le printemps et l'été. L'expérience de dégradation de la MOD extraite des sols a été réalisée sur une période de 28 jours et avait pour but de s'affranchir d'un potentiel apport de MOD labile. En effet, dans le milieu naturel, la matière organique issue d'apports terrigènes subit une multitude de processus qui induisent une série de transformations avant même son rejet dans les cours d'eau. En fonction des écoulements (ruissellement de surface, sub-surface ou profond), il existe un gradient vertical selon lequel la MOD transportée vers les cours d'eau peut être plus ou moins dégradée (Sanderman et al., 2009). Parmi les processus de transformation de la MOD, la dégradation microbienne est le phénomène le plus important (Fujii et al., 2010).

La comparaison entre les caractéristiques de la MOD extraite des sols et celles des cours d'eau a été réalisée sur la base des propriétés optiques et les distributions en poids moléculaire de la MOD (UV-Visible, Fluorescence 3D, HPSEC/UV-fluorescence). Les principaux résultats ont mis en évidence des différences de nature entre la MOD extraite des sols forestiers et celle des sols viticoles. Ces différences étaient d'autant plus marquées après la période de dégradation de la MOD (après 28 jours d'incubation). L'évolution de la MOD se produit principalement au cours des trois premiers jours d'incubation, ce qui entraîne la dégradation de molécules de type protéique de poids moléculaire élevé (>10 kDa) et de type acides aminés de faible poids moléculaire (<1 kDa) (fraction labile de la MOD), conduisant à une augmentation de l'aromaticité et de l'humification, observée pour les composés résiduels principalement de poids moléculaire moyen (de 1 à 10 kDa). La fraction labile ne présente en revanche pas de différence en fonction des usages de sols.

Les échantillons de cours d'eau adjacents présentent des empreintes optiques distinctes, principalement influencées par le régime hydrologique. En effet, la MOD prélevée pendant les périodes de hautes eaux présente une contribution plus importante de composés LMW (<1 kDa) et HMW (>10 kDa) que la MOD prélevée pendant les périodes de basses eaux. De plus, un lien a été établi entre la MOD dégradée extraite des sols et celle des cours d'eau adjacents pour les mêmes occupations des sols. Une analyse linéaire discriminante a mis en évidence

une série d'indicateurs qui ont montré que les sources diffuses ont un impact sur la qualité de la MOD, indépendamment des conditions hydrologiques. Ces indicateurs sont : le rapport de pentes spectrales (SR), le $SUVA_{254}$, les indices de fluorescence et PARAFAC (HIX, FI, YFI et composante PARAFAC C1(Ex320-Em440)) et les indicateurs HPSEC/UV-fluorescence pour les fractions LMW (< 1 kDa) et MMW (1-10 kDa). Ces résultats ont montré que les empreintes de MOD dans le cours d'eau étaient influencées à la fois par les usages de sols et par les périodes de prélèvements (hautes eaux et basses eaux). La contribution d'une source autochtone était ainsi visible en période de basses eaux et différente en fonction des usages de sols. Ainsi, l'échantillonnage représentatif d'une source diffuse de MOD dans les cours d'eau doit tenir compte à la fois de l'homogénéité (usage des sols) et des conditions hydrologiques des sous-bassins versants situés en amont.



Sampling terrigenous diffuse sources in watercourse: influence of land use and hydrological conditions on dissolved organic matter characteristics.

Amine Boukra¹, Matthieu Masson¹, Corinne Brosse¹, Mahaut Sourzac², Edith Parlanti², Cécile Miège¹

¹ INRAE, RiverLy, F-69625, Villeurbanne, France

²Univ. Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, F-33600 Pessac, France

Highlights:

- This study addresses the sampling of non-point sources of DOM in watercourses
- Land use markers combination discriminated DOM terrigenous signature in streams
- Hydrological conditions impacted the signature of watercourse DOM
- Relevant indicators from HPSEC/UV-fluorescence improved the study of DOM composition

Abstract

Diffuse and point sources of dissolved organic matter (DOM) in streams influence its composition, interactions and fate in the aquatic ecosystem. These inputs can be very numerous at the scale of a watershed, and their identification remains a challenge especially in the case of diffuse sources related to land use. The complexity of the transfer mechanisms and the reactivity of DOM throughout the soil-water column continuum raise questions about the sampling of diffuse sources in watercourses. To answer this question, we compared the characteristics of soil-extracted DOM influenced by a particular land use (homogenous sub-catchment of forest and vineyard) and DOM collected from the watercourse adjacent to the soil samples. A 28-day incubation experiment of soil extracts was designed to remove the labile fraction of DOM. During the first 3 days, between 40 and 70 % of the DOC mass was lost for both types of soils. A set of optical indicators (UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence) showed that the labile fraction was mostly composed by low (<1 kDa) and high (>10 kDa) protein-like molecules. At the end of the incubation, soil-extracted DOM was mainly composed of medium molecules (1-10 kDa) associated to terrigenous humic-like compounds. Its optical and size molecular signature tended towards that in the adjacent watercourses and was specific to land use. However, the characteristics of DOM in watercourses was also influenced by the hydrological conditions, probably due to a transfer of top soil DOM during high water periods and both deep soil and autochthonous DOM during low water periods. These results were obtained by a set of indicators including novel ones derived from high-pressure size exclusion chromatography (HPSEC/UV-fluorescence). Finally, this study demonstrated that it is possible to sample the DOM representative of a land use directly in the river downstream of the homogeneous sub-basin by multiplying the samples during contrasting hydrological conditions.

Keywords: Dissolved organic matter; non-point sources; land-use; optical fingerprint

1. Introduction

Dissolved organic matter (DOM) is a complex set of organic compounds operationally defined as material obtained after water filtration (commonly ranging from 0.2 to 0.7 μm) (Zsolnay, 2003; Zark and Dittmar, 2018) with an average molecular size distribution that ranges from <100 Da to >100000 Da (Nebbioso and Piccolo, 2013; Hawkes et al., 2019). DOM plays a major role in environmental biogeochemistry, particularly as the main driver of terrestrial carbon to the aquatic environment, as a substrate for the growth of microorganisms (Bolan et al., 2011; Mladenov et al., 2022), but also in the fate and bioavailability of pollutants (Reuter and Perdue, 1977; Aiken, 2014). At the scale of a watershed, DOM inputs to the rivers can originate from a wide range of natural and anthropogenic sources (Nebbioso and Piccolo, 2013). These sources can be either point sources (e.g. wastewater treatment plants and storm overflows) or diffuse discharges (e.g. terrigenous inputs related to land use, underground inputs, autochthonous production; Carstea et al., 2020; Williams et al., 2016). Point discharges of DOM can be easily located and their sampling is often representative of these sources. For example, DOM from wastewater treatment plant discharges, constitutes a mixture of organic molecules including microbial degradation products, surfactants, pharmaceutical products, phytosanitary products (herbicides, fungicides, insecticides) and disinfection by-products (Michael-Kordatou et al., 2015; Verkh et al., 2018).

Diffuse sources of DOM are more difficult to locate and study due to the complexity of geochemical interactions, degradation processes and the uncontrolled meteorological events at the origin of these DOM sources. In the case of terrigenous contributions of diffuse DOM sources, studies have shown that the soil type governs the amount of DOM transferred to the aquatic downstream environment (Bolan et al., 2011). Organic type soils discharge more DOM than mineral type soils. Land use as well as soil biomass influence the composition of DOM transferred to the watercourse (Kalbitz et al., 2003). This could be explained by the fact that DOM is considered to be the most active and labile part of soil organic matter (Guigue et al., 2014). Low molecular weight labile compounds make up the organic matter in the surface layers of the soil. Their proportion decreases in depth soils giving way to higher fraction of refractory compounds of high molecular weight. In addition, the hydrology of the watershed, or more specifically, the direction of water flow along the soil, determines the type of organic matter transferred to rivers. Indeed, Sanderman et al., 2009 demonstrated that vertical flow would favor the transfer of the more refractory denatured organic matter, whereas lateral

flow would transfer labile organic matter characteristic of the surface soil layers. However, establishing DOM fingerprints in rivers related to land uses and occupations, driven hydrology, is a complex task. The high reactivity of DOM labile fraction coupled with environmental conditions and the nature of the soils make difficult the establishment of reliable DOM fingerprints. Here we define a DOM fingerprint as a set of physico-chemical markers calculated from one or several analytical techniques (e.g. optical characteristics, molecular size distribution), which are relevant to discriminate a given source. At present, classically used techniques such as UV-Visible spectroscopy and excitation emission matrix (EEM) fluorescence have brought major advances in the study of chromophoric and fluorescent DOM. The application of these spectroscopic techniques has allowed, for example, to differentiate between types of DOM depending on the sources and the state of evolution (Li and Hur, 2017). Another analytical technique that showed relevance to characterize DOM composition, is the size exclusion chromatography (HPSEC). The principle of HPSEC is based on the separation of molecules according to their size or hydrodynamic volume. Previous studies have shown that HPSEC can be used to determine the distribution of molecular weight for DOM by calculating indicators used to characterize polymers such as average molecular weight (Mw), number average molecular weight (Mn) and polydispersity (p) (Huang et al., 2021). Detection by diode arrays and fluorescence (HPSEC/UV-fluorescence) allowed to specifically target absorbing and fluorescent compounds in organic matter, providing further qualitative information on the size continuum of DOM molecules (Asmala et al., 2021; Brezinski and Gorczyca, 2019). However, the combined information of molecule size and type, provided by HPSEC/UV-fluorescence, could be further explored to propose relevant indicators improving DOM fingerprints. In this context, the aim of this paper was to determine the representativeness of sampling a terrigenous diffuse source of DOM in watercourses. More precisely, we studied the link between soil-extracted DOM, characterized by a particular land use, and DOM collected in the soil-adjacent watercourse. A second objective was to improve the study of DOM optical fingerprints by introducing new indicators from size exclusion chromatography HPSEC/UV-fluorescence.

To achieve this, the characteristics of i) DOM extracted from soils defined by distinct land use (forest and vineyard), and ii) DOM from adjacent watercourses collected during high and low water periods were compared. The starting hypothesis is that during the high water period,

there would be a greater contribution of the terrigenous source. This comparison was performed by characterizing DOM properties using a combination of optical analysis techniques including UV-Visible spectrophotometry, EEM fluorescence spectroscopy and HPSEC/UV-fluorescence.

2. Materials and methods

2.1. Site description

All sampling areas were located in Ardière-Morcille watershed (152 km²) in eastern France. The catchment area, in its downstream part, is characterized by vineyard activities representing more than 32 % of total landscape (Peyrard et al., 2016). Other land uses such as forests (37 %) and pastures (21 %), are also present in the upstream watershed (Figure 5-2).

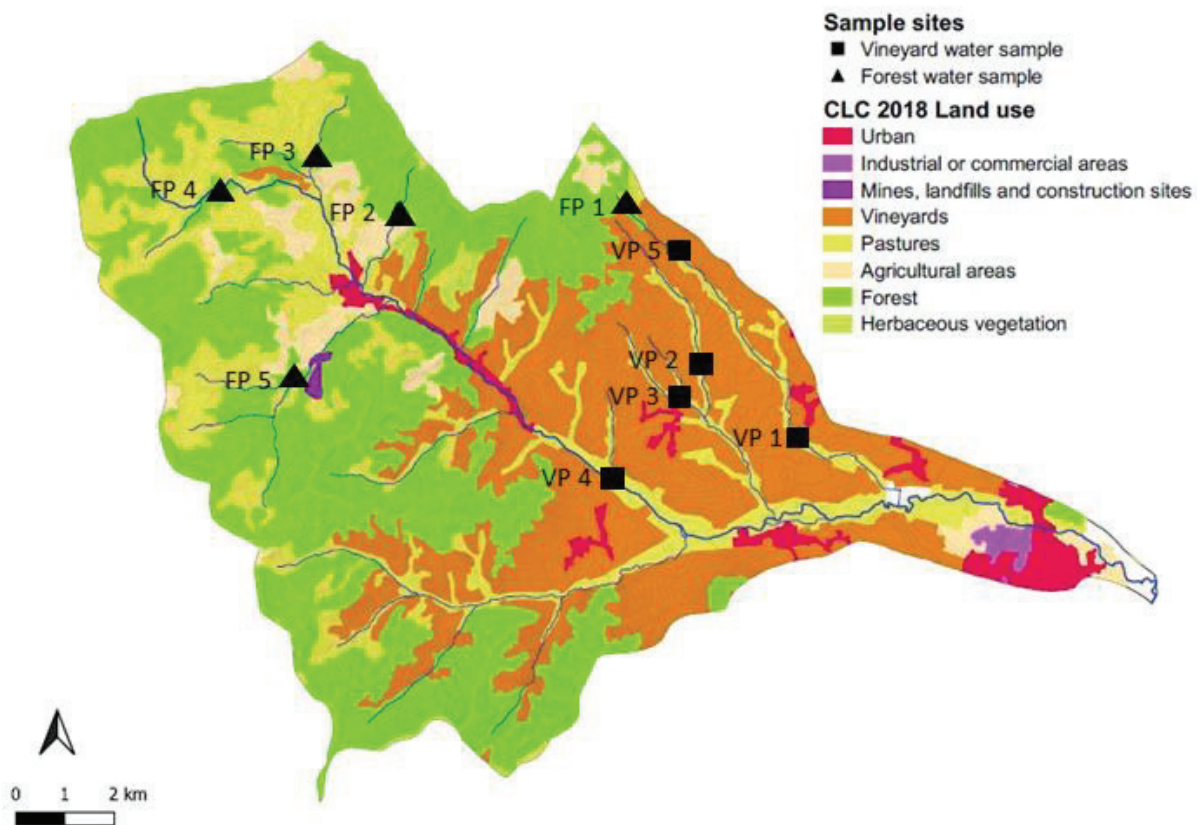


Figure 5-2. Map of Ardières-Morcille catchment showing the location of the water-sampling sites (FP: forest sampling points, VP: vineyard sampling points) and the land-use boundaries (Corine Land Cover CLC 2018). The soil samples were collected upstream the water sampling points.

The region's climate is semi-continental with a mean annual rainfall of 940 mm and episodic snowfalls during winter (Gouy et al., 2021). The soils are mostly sandy and erodible and the geological characteristics are an altered hercynian crystalline basement of mostly porphyric

biotite granite. The watershed is characterized by very steep slopes, ranging between 5 and 30-35 % favoring surface runoff. In addition, numerous and fast hydrological events make the riverbed, rarely dry (Rabiet et al., 2015).

2.2. Sampling design and samples preparation before experimentations in laboratory

Water samples were collected manually from rivers, downstream of 10 homogenous sub-catchment areas characterized by 5 forest land occupations and 5 vineyard land uses (Figure 5-2). Sub-catchment land uses were determined based on Corine Land Cover 2018 using Qgis 3.14 software. Sampling was carried out during two campaigns: A first sampling campaign in January 2021 during a period of high water (3.65 and $0.2 \text{ m}^3 \cdot \text{s}^{-1}$ respectively for Ardières and Morcille Rivers) and a second sampling campaign in July 2021 during the period of low water (0.65 and $0.03 \text{ m}^3 \cdot \text{s}^{-1}$ respectively for Ardières and Morcille Rivers). The flow data was obtained at Villié-Morgon and Beaujeau stations located downstream of the Ardières and Morcille rivers. The water samples were collected in 250 mL borosilicate glass bottles previously cleaned by an automatic washer at 90°C , then calcined in an oven at 500°C . The samples were filtered in the field using polypropylene syringe and polyvinylidene fluoride (PVDF) filters with a porosity of $0.45 \mu\text{m}$. The filtrate was first used to rinse the 250 mL borosilicate glass bottle three times before filling them completely. The samples were then stored in a cooler (at $\pm 4^\circ\text{C}$) protected from light during their transport to the laboratory, then stored at 4°C in the dark until analysis. Experimental blanks using ultrapure water (resistivity at $18.2 \text{ M}\Omega \cdot \text{cm}$) were prepared in laboratory using the same protocol as in the field, in order to identify possible contamination due to bottles, syringes or filters.

Soils were sampled in the forest and vineyard sub-catchment areas upstream of water samples (radius of 30 to 50 m) during the high water sampling campaign. Detailed geographic coordinates are available in supplementary material 1. In each site, 20 cm of top soils were taken with an auger; four subsamples were randomly selected to collect approximately 1 kg of soil. Each subsample was placed in a polypropylene bag. Once all 4 subsamples were collected, the bulk sample was mixed in the bag. The 10 soil samples were stored at 4°C protected from light until preparation in the laboratory where the soils were sieved with a

2 mm stainless steel sieve, placed in aluminum trays (40x60 cm) and dried for 96 h at 40 °C in an oven.

To summarize, for each sub-catchment area, one water sample was collected during high water period as well as one soil sample (composite of 4 randomly collected subsamples). During low water periods, an additional water sample was collected.

2.3. Dissolved organic matter extracted from soils (soil extracts)

Dissolved organic matter was extracted from soil samples by stirring 20 g (dry weight) of soil in 200 mL of ultrapure water with 10 mM CaCl₂ (Supplementary Material 2). Stirring was performed at 180 rotations per minute in a Lovibond thermoregulator cabinet at 20 °C in dark condition. After 6 h of agitation, the samples were centrifuged at 4500 g during 15 min and the supernatant was filtered through 0.45 µm (PVDF filter). An initial 60 mL sample was collected before the incubation experiment.

2.4. Incubation experiment

Two inoculums of forest and vineyard soils were prepared 48 h before the incubation experiment. To obtain better microbial diversity, both forest and vineyard inoculums were prepared from, respectively, a mixture of the 5 forest soils and the 5 vineyard soils (Kalbitz et al., 2003; Guigue et al., 2014). For each inoculum, 5 g of the 5 corresponding soil samples were added to 200 mL of ultrapure water. These preparations were then shaken at 20° for 48h in the dark. A nutrient solution of (NH₄)₂SO₄ and KH₂PO₄ (C:N:P:S:K molar ratio of <5:1:1:1:1) was also prepared 1h prior to the incubation. The addition of nutrients allows optimizing the activity of the microorganisms during the following incubation.

Incubation of DOM extracted from soils was performed by adding 35 µL of inoculum and 1 mL of nutrient solution to the samples immediately after the soil DOM extraction step (Guigue et al., 2014) (Supplementary Material 2). Controls without inoculum nor nutrient solution (inoculum blanks) and controls without soil extracted DOM (extraction blanks) were performed. All samples and controls were incubated at 20°C in the dark, with continuous shaking at 180 rotations per minute during 28 days. The incubation period of 28 days was chosen based on previous work that studied the time required to degrade the labile fraction of DOM from soil samples (Guigue et al., 2014). Sub-samples at 3, 7, and 14 days of incubation were collected to monitor changes in DOM quantity and quality (optical properties).

2.6. Analytical strategy

2.6.1. Dissolved organic Carbon

Dissolved organic carbon (DOC) analyses were performed by high temperature catalytic combustion according to the standard NF EN 1484 and using an Analytik Jena multi N/C® 3100 TOC analyzer. The calibration, rinsing, quality controls solutions, as well as samples, were introduced in precombusted (at 500°C during 2 h) glass vials, where they were acidified to pH=1 with hydrochloric acid (HCl 37% RPE). The detection limit was 0.2 mgC.L⁻¹ and the analytical uncertainty was 12%. The accuracy of the analyses (105%) was routinely checked using a certified standard of potassium hydrogen phthalate (5 mg.L⁻¹, Fluka®) different from that used for the calibration (0.2–25 mgL⁻¹, Sigma-Aldrich®).

2.6.2. UV-Visible spectroscopic analysis

UV-Visible spectroscopy analyses were performed using a Shimadzu UV 1900 spectrophotometer with a 1 cm path length quartz cell (Hellma Analytics®). Absorbance spectra were acquired between 200 and 800 nm with a resolution of 1 nm and a scan speed of 0.2 s.nm⁻¹. A baseline correction was carried out using an ultrapure water blank over the same wavelength range. A series of classical descriptive indicators (SUVA₂₅₄, absorbance ratios and spectral slopes) were calculated from the generated absorbance spectra (Table 1).

Table 1. Classical UV-Visible and fluorescence indicators to characterize dissolved organic matter.

Indicator	Calculation	Interpretation	References
SUVA ₂₅₄	$\frac{Abs\ 254\ nm}{[DOC]}$	Correlated with the aromaticity of DOM.	(Weishaar et al., 2003)
E2.E3	$\frac{Abs\ 250\ nm}{Abs\ 365\ nm}$	Negatively correlated to organic molecules size	Helms et al., 2008
E2.E4	$\frac{Abs\ 254\ nm}{Abs\ 436\ nm}$	Autochthonous versus terrestrial DOM	Jaffé et al., 2004; Li and Hur, 2017
E3.E4	$\frac{Abs\ 300\ nm}{Abs\ 400\ nm}$	Negatively correlated to humification degree	Claret et al., 2003; Li and Hur, 2017
S ₁ (275-295 nm)	$\alpha_{\lambda} = \alpha_{\lambda_0} e^{S(\lambda_0 - \lambda)}$ α_{λ} : absorption coefficient at specific wavelength λ_0 : reference wavelength S : exponential slope	Negatively correlated to aromatic carbon content	Stedmon et al., 2000 Helms et al., 2008
S ₂ (350-400 nm)			
SR	$\frac{S_1}{S_2}$	Negatively correlated to molecular weight	Helms et al., 2008
HIX	$\frac{\sum IF\ 435-480}{\sum IF\ 300-445}$ at 255 nm	Correlated to the aromatic carbon content/ humification degree of DOM	Zsolnay et al., 1999
BIX	$\frac{IF\ 380}{IF\ 430}$ at 310 nm	Indicator of recent autochthonous production of DOM due to biological activity	(Huguet et al., 2009)
FI	$\frac{IF\ 450}{IF\ 500}$ at 370 nm	Terrigenous vs microbial DOM	Mcknight et al., 2001
YFI	$\frac{FEI\ 350-400}{FEI\ 400-450}$ at 280 nm FEI : average fluorescence intensities	Increases the resolution of FI indicator	Heo et al., 2016

2.6.3. EEM fluorescence spectroscopy and PARAFAC decomposition

Excitation-emission matrix (EEM) fluorescence analyses were performed using an Aqualog (Horiba Scientific) spectrofluorometer in a 1 cm path length quartz cuvette (Hellma Analytics®) thermostated at 20°C. The Aqualog instrument was equipped with a 150 W Xenon arc lamp and excitation wavelengths were scanned using a double-grating monochromator from 240 to 800 nm at 5 nm intervals, while the emission spectra were obtained using a high-gain CCD detector at intervals of approximately 0.58-nm (1-pixel) for wavelengths from 240 to 800 nm. All EEM spectra were corrected for instrumental biases and normalized to the area under the Raman peak of water, at 350 nm excitation, acquired daily. Fluorescence intensities are therefore expressed in Raman units (R.U). In order to avoid inner filter effects, all samples with maximum absorbance > 0.1 were diluted. Raman and Rayleigh scattering peaks were removed by subtracting blank spectrum. The generated spectra were treated using TreatEEM-V1 software developed by Dario Omanović (Ruđer Bošković Institute, Zagreb, Croatia; <https://sites.google.com/site/daromasoft/home/treateem>). The fluorescence indices HIX (Zsolnay), BIX, FI, YFI (Table 1) were calculated for an excitation wavelength of 250, 310, 370 and 280 nm respectively (Heo et al., 2016; Huguet et al., 2009; Mcknight et al., 2001; Zsolnay et al., 1999)

The multi-way parallel factor analysis (PARAFAC) was applied to a set of 103 EEM spectra (corresponding to 83 soil extracts and 20 water samples collected from the adjacent streams). The PARAFAC model was run, using the drEEM toolbox under MatLab R2020a, for 4 to 8 components with non-negativity constraints (Murphy et al., 2013). Four components were validated based on split-half analysis and compared to previously reported components using the OpenFluor fluorescence database.

2.6.4. Size exclusion chromatography HPSEC/UV-fluorescence

High performance size exclusion chromatography (HPSEC; LC-2030C 3D Plus, Shimadzu) coupled with photodiode array (Deterium D2 Lamp PDA) and fluorescence (RF-20A xs) detectors was used to determine the molecular size distribution of DOM. The injection volume was 10 µL with a flow rate at 0.8 mL.mn⁻¹ of phosphate buffer mobile phase (NaCl, KH₂PO₄ and K₂HPO₄). Two columns (Shodex OHpak SB-803 HQ, Shodex OHpak SB-802 HQ) coupled in series were used for the chromatographic separation. The first column covered an exclusion

range between 1 kDa and 100 kDa and the second one between 0.2 kDa and 1 kDa. Analytical standards PSS (polysulfonate) were used as references of average molecular weights (0.246, 1.6, 4.3, 6.8, 10, 17, 32, 77 kDa). Other standards were also used to cover a wider range of molecule types, such as nitrates (0.063 kDa), Bovin Serum Albumin BSA (66 kDa) and Polyacrylic Acid PAA standards (1, 2.6 and 3.6 kDa). A calibration curve of molecular weight (from 0.2 - to 77 kDa) according to elution volume was calculated using these analytical standards (Brezinski and Gorczyca, 2019).

The UV-DAD detector allowed the acquisition of chromatograms on an absorbance range between 200 and 450 nm with an increment of 1.2 nm every 300 ms. Regarding the fluorescence detector, 3 couples of excitation-emission wavelengths (Ex-Em) were retained for all the analyses. The fluorescence intensities were measured for the wavelengths corresponding to the humic-like components (315-443 nm), protein-type components (280-330 nm), and tryptophan-type components (278-354 nm) were measured every 400 ms.

To optimize the use of the data generated by HPSEC/UV-fluorescence, a Gaussian decomposition was performed. The aim was to decompose the size continuum obtained in each HPSEC chromatogram into distinct Gaussian components (Eq. 1). For this, the toolbox «peakfit.m» (O’Haver, 2020) was applied on MatLab *R2020a*.

$$\text{Eq. 1 : } A(x; \mu; \sigma; \varphi) = \sum_{i=0}^{nC} \varphi_i e^{\frac{-(x-\mu_i)^2}{2\sigma^2 i}} + \varepsilon$$

where nC is the number of Gaussian components, ε is the residuals which represent the variability not taken into account by the model, x is the peak area, μ is the peak position parameter (elution volume), σ is the peak width and φ is the proportion of each fitted Gaussian distribution.

The number of detected components was validated using the Bayesian Information Criterion (Asmala et al., 2021). Areas of the identified Gaussian components were summed in order to calculate the total sec absorbance (TSA) from chromatograms obtained with detection at UV 254 nm (Asmala et al., 2021). The total sec fluorescence TSF was calculated from chromatograms obtained with fluorescence detection at three excitation-emission (Ex-Em) wavelengths (nm): Ex: 315-Em: 443, Ex: 280-Em: 330 and Ex: 278-Em: 354. These wavelengths

are classically used to characterize DOM properties (Table 1). The calibration curve allowed classifying each Gaussian component in a molecular weight fraction, based on the peak position (elution volume). In this study, three molecular weight fractions were defined as: low molecular weight fraction (LMW) with an average molecular weight < 1 kDa, medium molecular weight fraction (MMW) with an average molecular weight ranging from 1 to 10 kDa, and high molecular weight (HMW) with an average molecular weight >10 kDa. The LMW, MMW and HMW fractions represent the relative percentages of respectively small, medium and large molecules detected at the measured wavelengths.

2.7. Mathematical and statistical analyses

Anova analysis coupled with a Tukey test (packages stats and rstatix; R 4.0.2 software) were used to compare both forest and vineyard soil extracts, the temporal kinetics of the incubation study and also for the comparison between soil extracts and adjacent natural waters for all UV-Vis, EEM fluorescence and HPSEC/UV-fluorescence indicators. Principal component analysis (PCA; package factoextra, R 4.0.2 software) was also used for the comparison of soil extracts vs. natural adjacent waters, and a linear discriminant analysis (LDA; packages MASS and klaR, R 4.0.2 software) for the selection of the most relevant DOM markers of forest and vineyard soil occupations.

3. Results

3.1. Dissolved organic carbon

Before incubation, dissolved organic carbon (DOC) concentrations (Figure 5-3.A) in the 5 forest soil extracts ranged from 22.7 to 48.2 mg.L⁻¹ with a mean value of 32.7 ± 9.4 mg.L⁻¹. For the 5 vineyard soil extracts, DOC concentrations varied between 16.14 and 34.14 mg.L⁻¹ with a mean value of 23.8 ± 6.6 mg.L⁻¹. These values are consistent with those obtained in previous studies (Rinot et al., 2021). No significant difference (Tukey test, $p > 0.05$, $n=50$) in DOC concentrations was found between forest and vineyard soil extracts before incubation.

The percentage of DOC (Figure 5-3.B) was calculated for the duration of the experiment (28 days). For forest soil extracts, the percentage of DOC decreased by 67 ± 8 %, with a significant decrease (Tukey test, $p < 0.05$, $n=50$) of 42 ± 9 % within the first three days of incubation. For vineyard soil extracts, the percentage of DOC decreased by 61 ± 3 %, with a significant decrease of 46 ± 5%, again during the first three days of incubation. Previous studies showed DOC losses range from 5 to 93% depending on soil types (forest, pasture, arable) and extraction protocols (pressurized hot-water, water extraction, leaching extraction) (Guigue et al., 2014; Kalbitz et al., 2003).

DOC concentrations of samples collected in streams adjacent to forest soils ranged from 2.2 to 3.0 mg.L⁻¹ during high water periods (Figure 5-3.A). For these same sites, DOC concentrations were significantly higher (Tukey test, $p < 0.05$, $n=20$) ranging between 3.3 mg.L⁻¹ and 4.6 mg.L⁻¹ during low water periods. In the streams adjacent to vineyard soils, DOC ranged from 3.1 to 4.4 mg.L⁻¹ during high water. Non-significant difference was observed during low water periods with DOC concentrations ranging between 3.5 and 4.6 mg.L⁻¹. In addition, DOC concentrations were significantly higher in the vineyard stream samples than in the forest stream samples only during high water.

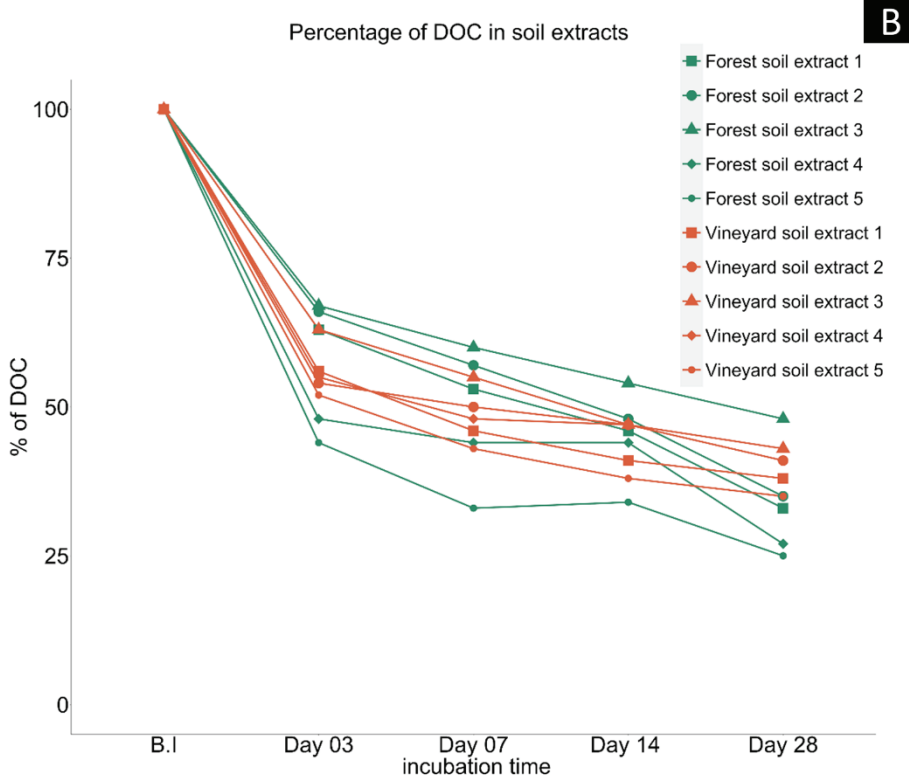
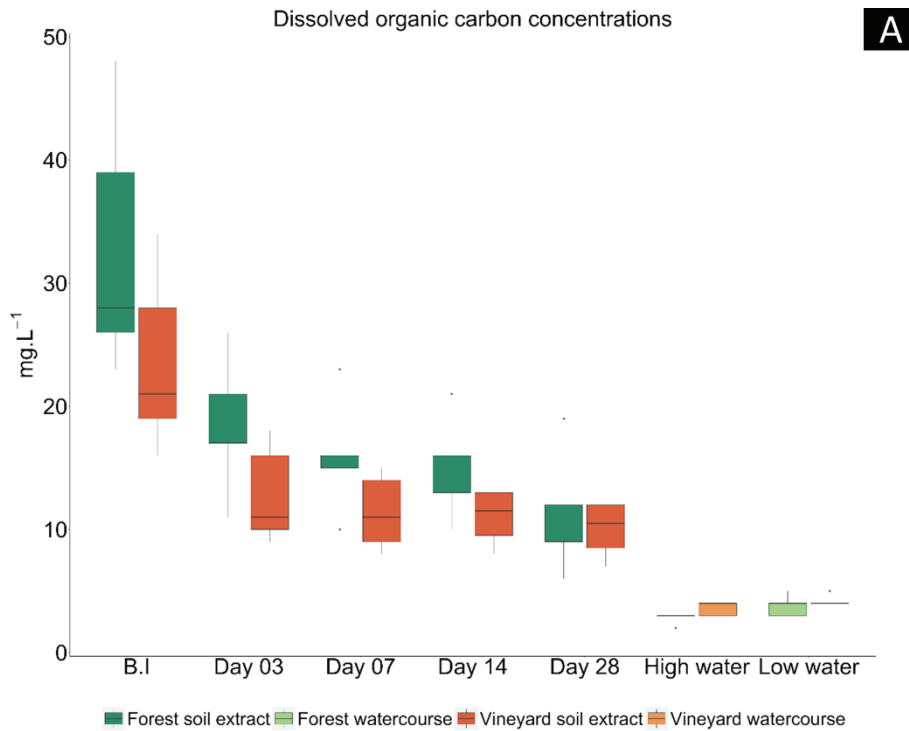


Figure 5-3. A: Dissolved organic carbon (DOC) concentrations measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods. B: Percentage of DOC for soil extracts during the 28 days incubation experiment relative to the DOC concentration measured before incubation.

3.2. UV-Visible

UV-Visible spectra (Supplementary Material 3) show a classical pattern, i.e. an exponential decrease between 200 and 250nm, followed by a slight decrease between 250 and 400 nm. Then the signal is close to the background for wavelength superior to 400 nm. A decrease in absorption is observed between the soil extracts before and after 28 days of incubation, at wavelengths from 250 to 400 nm, i.e. the range classically used to characterize the chromophoric DOM. $SUVA_{254}$ (Table 1) increased significantly (Tukey test, $p < 0.05$, $n=50$) over the 28-day incubation period of the experiment, from 0.98 ± 0.11 to 2.02 ± 0.12 $L \cdot mg^{-1} \cdot cm^{-1}$ for forest soil extracts, and from 1.44 ± 0.45 to 3.02 ± 0.51 $L \cdot mg^{-1} \cdot cm^{-1}$ for vineyard soil extracts (Figure 5-4.A). A significant difference (Tukey, test, $p < 0.05$, $n=50$) was observed between forest and vineyard soils only after the 3rd day of incubation. During the first 3 days, spectral slope $S_{275-295}$ and E2.E3 index (Table 1) values decreased significantly (Tukey test, $p < 0.05$, $n=50$) and then stabilized during the rest of the incubation period (Figure 5-4.B, C). Overall, values of $S_{275-295}$ for forest and vineyard extracts respectively decreased from 0.023 ± 0.002 to 0.017 ± 0.001 and from 0.022 ± 0.003 to 0.017 ± 0.003 ; and the E2.E3 index values decreased from 7.60 ± 0.47 to 6.33 ± 0.59 and from 6.86 ± 0.92 to 5.91 ± 0.90 . The values of the SR, E2.E4 and E3.E4 (Table 1) ratios did not vary significantly between the soil extraction and the last day of the experiment (Figure 5-4.D, E, F).

Regarding samples collected in adjacent streams, no significant difference was observed according to land use. However, the indicator SR showed significant differences between high water and low water sampling periods (Tukey test, $p < 0.05$, $n=20$). Indeed, values of SR varied between 0.751 ± 0.0497 to 0.808 ± 0.012 and 0.775 ± 0.027 to 0.833 ± 0.011 for both forest and vineyard adjacent water samples. Previous studies on soil extracts and stream water samples (Sanderman et al., 2009; Liu et al., 2021; Chen et al., 2021) have reported UV-Visible index values in the same ranges as those presented in this work.

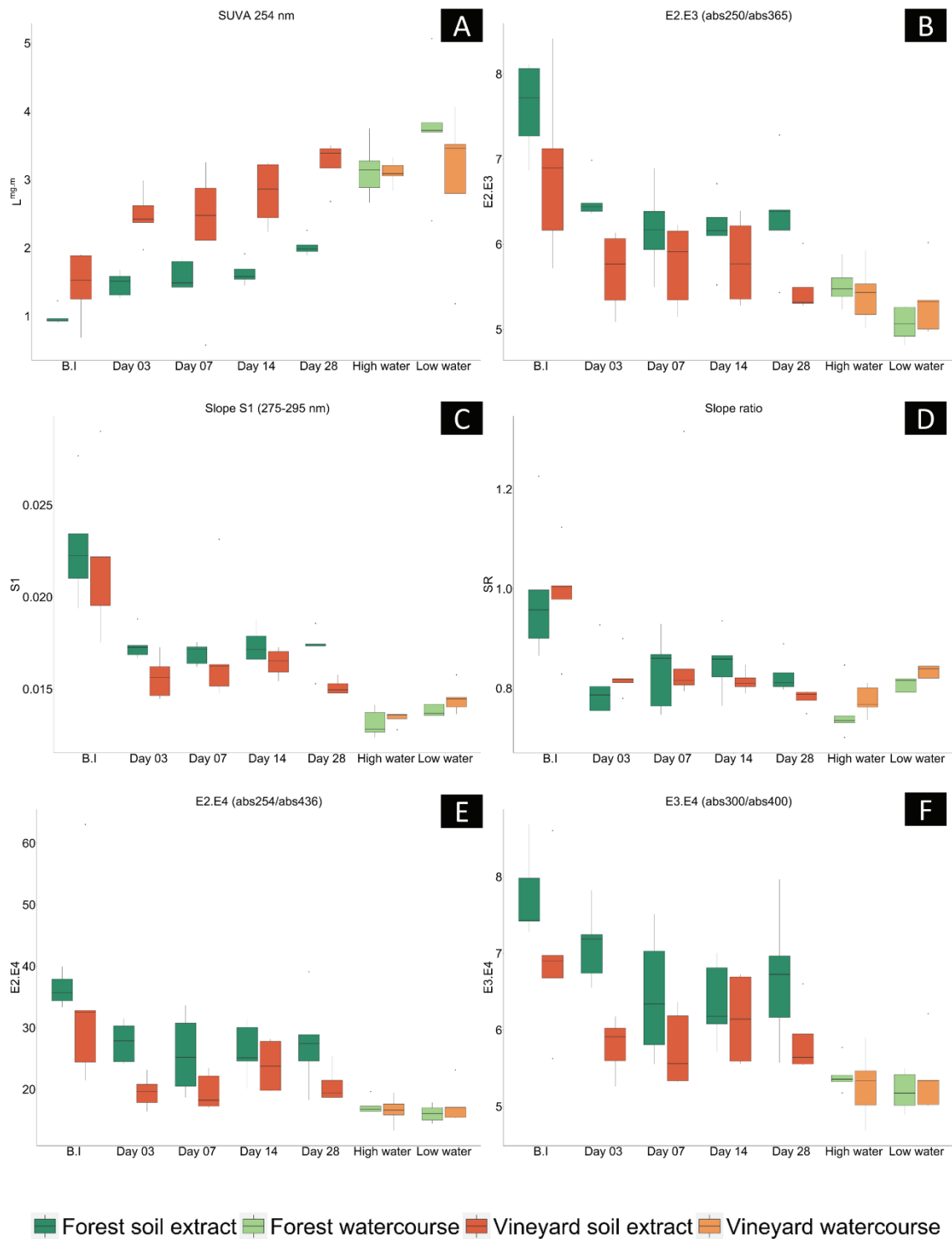


Figure 5-4. UV-Visible indicators measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods. A: SUVA₂₅₄, B: E2.E3 (abs250/abs365), C: Slope S1 (275_295 nm), D: Slope ratio SR (S275_295nm/S350_400nm), E: E2.E4 (abs254/abs436) and F: E3.E4 (abs300/abs400).

3.3. EEM fluorescence spectroscopy

The humification index (HIX; Table 1) showed a significant increasing trend during the first days of incubation (Tukey test, $p < 0.05$, $n=50$). These values stabilized for vineyard soil extracts while they decreased for forest soil extracts at the end of incubation (Figure 5-5.A). Indeed, the values increased from 3.1 ± 1.0 to 3.7 ± 1.5 for the forest soil extracts and from 2.9 ± 0.4 to 6.6 ± 0.7 for the vineyard soil extracts. Therefore, a significant difference (Tukey test, $p < 0.05$, $n=50$) is also observed between the two types of soils at the end of the experiment. No particular trend was observed for the indicator of biological activity (BIX; Table 1) and the Freshness index (FI; Table 1) during the incubation period (Figure 5-5.B, C). BIX values ranged from 0.71 ± 0.08 to 0.79 ± 0.08 for forest soil extracts; and from 0.63 ± 0.05 to 0.67 ± 0.02 for vineyard soil extracts; FI values from 1.40 ± 0.06 to 1.42 ± 0.04 for forest soil extracts; and from 1.31 ± 0.08 to 1.36 ± 0.04 for vineyard soil extracts. The YFI indicator (Table 1), showed a decrease according to incubation time (Figure 5-5.D). This decrease, however, is very limited for both soil extracts during the incubation experiment (YFI values from 0.89 ± 0.27 to 0.71 ± 0.06 and from 0.74 ± 0.04 to 0.61 ± 0.04 respectively for forest and vineyard soil extracts).

For water samples, only HIX showed a significant difference between water collected during high and low water (Tukey test, $p < 0.05$, $n=20$). HIX values varied between 8.44 ± 0.88 and 7.28 ± 1.11 for the streams adjacent to the forest, sampled during high and low water periods. For vineyard streams, it varied between 9.05 ± 0.34 and 7.39 ± 0.53 during the two sampling periods. The other indices BIX, FI and YFI showed no significant differences between forest and vineyard waters, nor between sampling periods. Values of fluorescence indices in this work are consistent with previous studies obtained on forest soils, peat soils, and subsoil extracts (Hansen et al., 2016; Olshansky et al., 2018; Zhang et al., 2022).

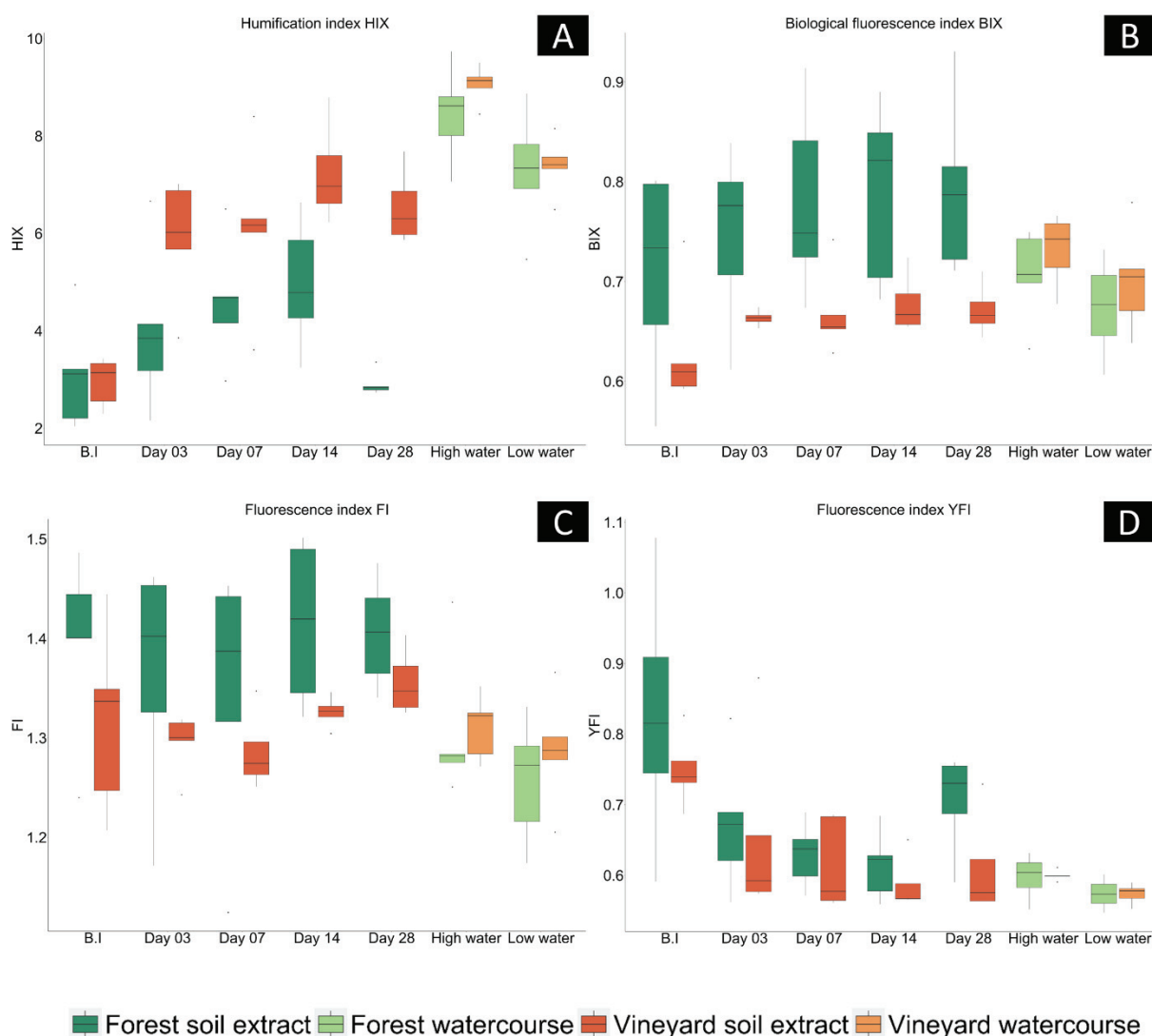


Figure 5-5. EEM fluorescence indicators measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods. A: humification index HIX, B: biological fluorescence index BIX, C: fluorescence index FI and D: fluorescence index YFI.

3.4. PARAFAC modeling

The contour plots of the 4 components determined by PARAFAC analysis are shown in supplementary material 4. Components C1 to C4 were identified from the literature and successfully matched in the OpenFluor database with similarity scores above 0.97. Component C1 showed two excitation maxima, a first one at ≤ 240 nm and a second one with lower intensity at 320 nm with an emission maxima at 420 nm. The component C1 has been described as terrestrial DOM degraded from lignin (Peleato et al., 2018) typical of soils and terrestrial sources in aquatic environments (Garcia et al., 2018; Yamashita et al., 2021). This

component has further been attributed to bioavailable organic matter, microbially derived and/or sensitive to photodegradation (Cawley et al., 2012; Eder et al., 2022).

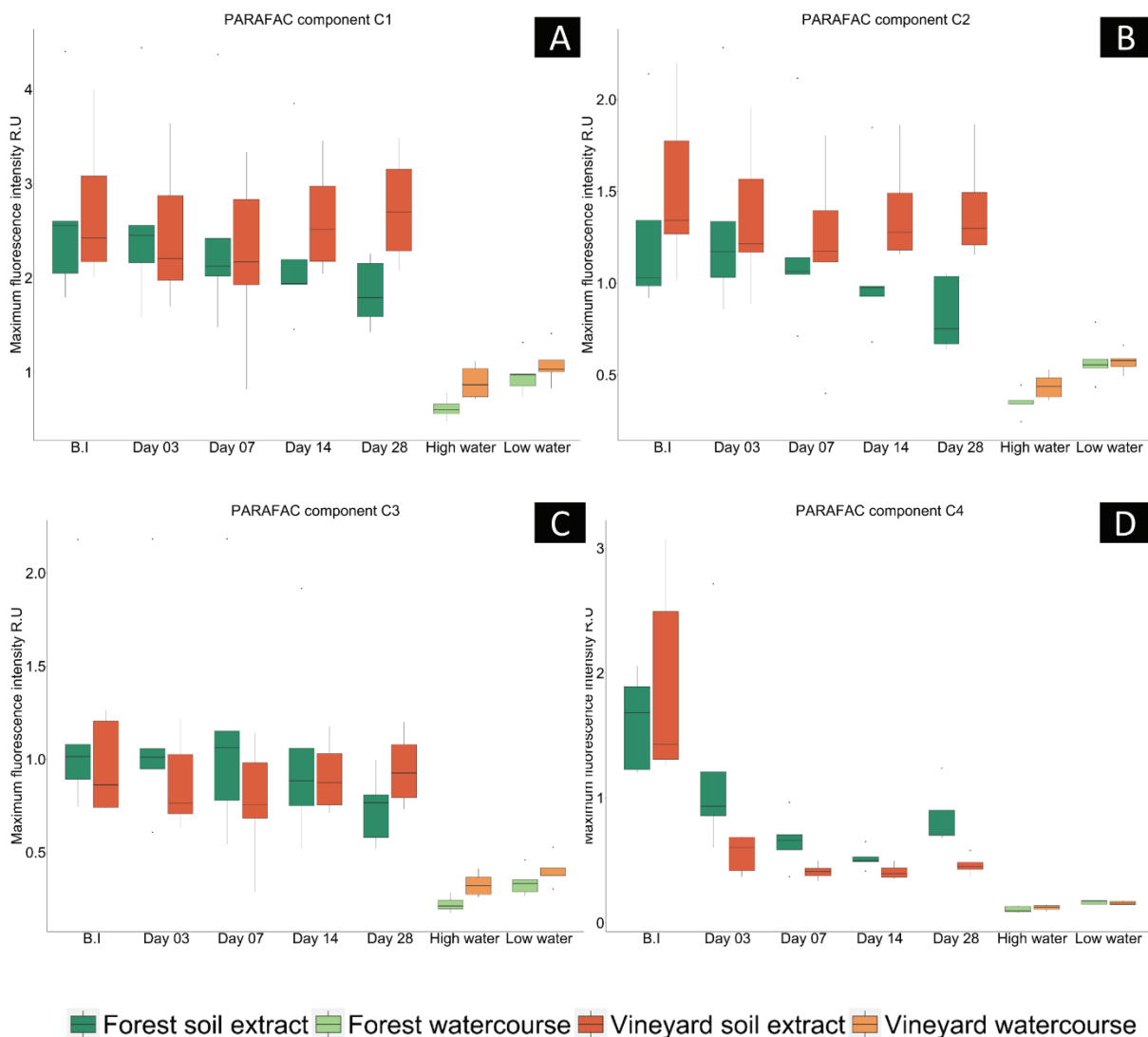


Figure 5-6. PARAFAC components C1 (<240(320)-420 nm), C2 (<240(390)-510 nm), C3 (365-443 nm) and C4 (280-330 nm) measured i) in soil extracts (forest in dark green and vineyard in dark orange) before incubation (B.I) and after at 3, 7, 14 and 28 days of incubation, and ii) in water samples collected from streams adjacent to the forest (in light green) and vineyard (in light orange) soils during high and low water periods.

Component C2 showed an excitation maxima at ≤ 240 and a lower intensity one at 390 nm with a maximum emission intensity at 510 nm. Previous studies have reported the component C2 as an ubiquitous terrestrial material with high molecular weight (MW) and aromatic content (Cawley et al., 2012; Wünsch et al., 2018; Yamashita et al., 2021). It has also been reported as fluorophores derived from higher plants and associated with photochemical

(Coulson et al., 2022; Eder et al., 2022) or microbial (Cory and McKnight, 2005) degradation. Component C3 showed an excitation maximum at 365 nm and an emission at 443 nm. It has been reported to be high MW, aromatic and hydrophobic molecules, corresponding to lignin-like compounds (Dainard et al., 2015; Derrien et al., 2018; Rinot et al., 2021). Other studies reported component C3 as organic matter of plant origin that has undergone little transformation in the soil after production (Catalá et al., 2021; Retelletti Brogi et al., 2019). Finally, the component C4 showed an excitation maximum at 280 nm and an emission at 330 nm. Numerous studies have reported this component to be attributed to amino acids, either free or bound to proteins, or to phenolic-type compounds (Coble, 1996; Kida et al., 2019; Parlanti et al., 2000; Wauthy et al., 2018).

Of the four components identified, only component C4 showed a decreasing trend during the incubation experiment (Figure 5-6). Indeed, for both soil extracts, the maximum fluorescence intensity of the C4 component decreased significantly (Tukey test, $p < 0.05$, $n=50$), from 2.6 ± 0.3 to 0.8 ± 0.2 for the forest soil extracts and from 1.9 ± 0.7 to 0.5 ± 0.1 for the vineyard soil extracts. On the other hand, all components showed significant differences between streams adjacent to the forest sampled during high and low water periods. In these samples, component C1 varied between 0.62 ± 0.1 and 0.97 ± 0.19 R.U, C2 between 0.34 ± 0.06 and 0.58 ± 0.11 R.U, component C3 between 0.22 ± 0.03 and 0.34 ± 0.06 R.U and component C4 varied between 0.10 ± 0.02 and 0.16 ± 0.01 R.U.

3.5. Size exclusion chromatography and Gaussian decomposition

The HPSEC/UV-fluorescence chromatograms provided additional information on the size distribution of DOM molecules contained in the samples. The Gaussian decomposition of the chromatograms revealed 6 to 8 fractions, depending on wavelength detection, distributed throughout the chromatogram. The percentage contribution of the LMW fraction (Section 2.5.4), for UV detection at 254 nm (Figure 5-7.A) and fluorescence at 280-330 and 278-354 nm (Figure 5-7.C, D), was significantly higher, before incubation, in forest soil extracts compared to vineyard soil extracts (Tukey test, $p < 0.05$, $n=50$). At the end of the experiment, the percentage of LMW was different only for UV detection at 254 nm (Tukey test, $p < 0.05$, $n=50$). For the adjacent water streams, significant differences were observed on the 280-330 and 278-354 nm fluorescence ranges according to the low vs. high water hydrological regime and this for both soil occupations (Figure 5-7.C,D). The MMW fraction, which corresponds to all

Gaussian components classified between 1 kDa and 10 kDa, showed the least variation. Nevertheless, the percentage of MMW at UV 254 nm were significantly different between the two types of soil extracts at the end of the incubation period (Tukey test, $p < 0.05$, $n=50$). The HMW fraction (> 10 kDa) showed very low percentages ($< 10\%$) compared to the LMW and MMW fractions for UV detection at 254 nm and fluorescence at 315-443 nm (Figure 5-7B). However, the HMW fraction was higher for 280-330 and 278-354 nm fluorescence detection with a significant difference (Tukey test, $p < 0.05$, $n=50$) between forest and vineyard soil extracts before incubation. For adjacent streams, the HMW fraction was also present at very low percentages and no difference was observed between soil occupations nor between low and high water regimes.

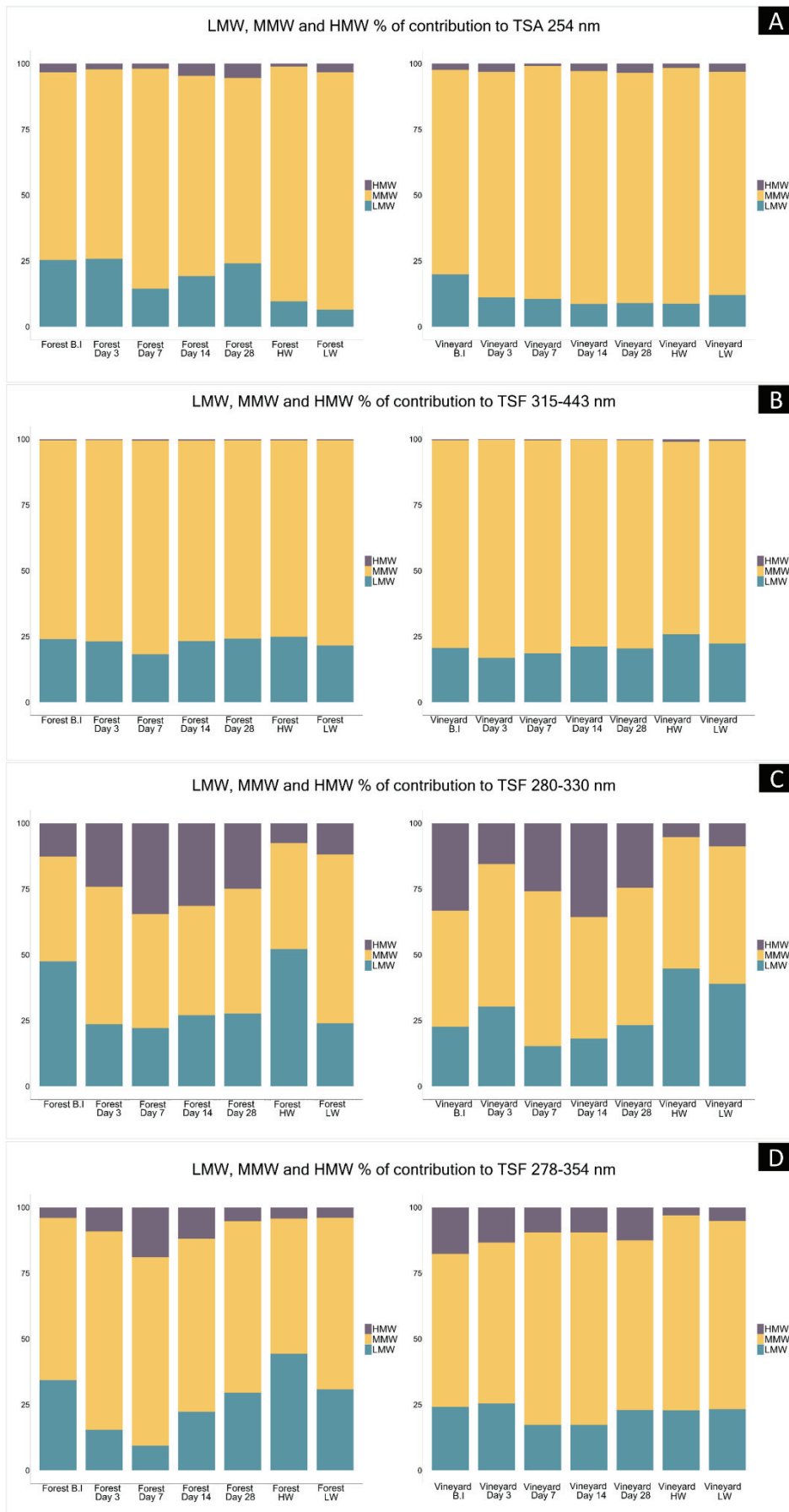


Figure 5-7. Low Molecular Weight (LMW; blue bars), Medium Molecular Weight (MMW; orange bars) and High Molecular Weight (HMW; purple bars) percentage of contribution to: A: Total Sec Absorbance TSA at PDA 254 nm, B: Total Sec Fluorescence TSF at Excitation Emission wavelengths 345-443 nm, C: TSF at fluorescence Excitation Emission wavelengths 280-330 nm, D: TSF at fluorescence Excitation Emission wavelengths 278-354 nm.. Samples plotted are soil extracts before incubation (B.I) and then at 3, 7, 14 and 28 days of incubation as well as for water samples collected from streams adjacent to the forest and vineyard soils during high and low water periods.

4. Discussion

4.1. Incubation experiment: from labile to refractory DOM in soil extracts

In the natural environment, the labile fraction of DOM is mainly degraded in the soil before being transported to the adjacent stream. Thus, the incubation experiment was designed to remove a possible labile fraction of DOM in the soil extracts before comparison with DOM from adjacent streams. During the 28-day incubation, most of the DOC mass was lost in the first 3 days for both soils (Figure 5-3A). The percentage of DOC loss stabilized between 40 and 70 % after 3 days incubation. These results are consistent with those obtained in previous studies investigating the biodegradation of water-extracted DOM from different soil types such as forest, pasture and agricultural soils (Guigue et al., 2014; Kalbitz et al., 2003). DOC loss is mostly explained by the degradation of labile DOM that is easily assimilated by microorganisms. It is well known that certain organic molecules, such as amino acids or carboxylic acids, are considered as labile while other molecules, such as lignins, are considered more recalcitrant (Eilers et al., 2010). The variability of the percentage of DOC loss observed in this study (between 40 and 70 %) could be explained by differences in the microorganisms communities contained in the soil samples. Indeed, it has been shown that a greater diversity of microorganisms in soils favors the DOC degradation (Raczka et al., 2021). Different proportions of labile and recalcitrant DOM pools in the soils could also explain the variability of DOC loss. Thus, soils containing higher levels of labile DOM will be degraded to a greater proportion than soils containing high levels of refractory DOM.

The degradation of labile molecules induced an increase in the average aromaticity content, average size and molecular weight of molecules, as highlighted by the indicators $SUVA_{254}$, E2.E3 and SR. Previous studies have also shown that degraded organic matter has relatively higher aromaticity content and molecular weight (Guarch-Ribot and Butturini, 2016; Hansen et al., 2016). The E3.E4 indicator being inversely correlated to the humification of DOM (Claret

et al., 2003) can be interpreted with the HIX indicator. The HIX and E3:E4 indicators show that the degree of humification of soil extracts was low at the beginning of incubation, and then increased to stabilize after 3 days of incubation for vineyard soils. For forest soils, a gradual increase was observed throughout the incubation period until 28 days where a significant decrease occurred. The unexpected decrease observed for forest soil extracts at the end of incubation could be explained by the very high values of γ band (280-330 nm); linked with protein-like compounds. The HIX index corresponds to the ratio of areas 435-480 nm and 300-345 nm of the emission spectrum for an excitation at 255 nm, thus a high γ band intensity can make the HIX values vary drastically. The other explanation could come from the humic fraction of forest soils. Indeed, it is a complex fraction consisting of functional groups such as phenols, quinones, methoxyls and carbonyls (Maccarthy, 2001), that are recalcitrant with a slow kinetic of degradation due to its complex chemical nature (Marschner et al., 2008). Humification is a process of DOM condensation resulting from several interactions such as plant alteration with partially degraded lignins or the transformation of polyphenols, amino acids and some sugars into quinones under the enzymatic action of microorganisms (Stevenson, 1995). This slow process can sometimes take years.

The BIX index showed a significant variability in forest soil extracts contrary to vineyard soils. This could be explained by a more contrasted biological activity in forest soil extracts than in vineyard soils. Based on the hypothesis that biological activity is directly related to the diversity of microbial communities, vineyard soil extracts may have less microbial diversity due to intensive soil exploitation allowing an homogenization of the communities present in the soil (Jangid et al., 2008). Other studies mention the concept of "priming effect", which reflects a greater mineralization of organic matter in soils enriched with fresh litter (Derrien et al., 2019). This could be the case for forest soils which are subject to a contribution of fresh organic matter through vegetation cover. In addition, the significant decrease of the YFI indicator shows a lower contribution of protein-like compounds at the end of the experiment. EEM fluorescence and PARAFAC decomposition bring details on the nature of the degraded molecules along the incubation. The component C4 (275-350) allows to affirm that some degraded labile molecules are tryptophan-like with recent biological production (Fellman et al., 2008). No significant difference was observed in the evolution of the other components C1, C2 and C3 which are characteristic of the following typical compounds: terrestrial humic-

like (plant derived), humic-like matter, commonly found in agricultural catchments and terrestrial high aromatic and hydrophobic macromolecules (Derrien et al., 2018; Jørgensen et al., 2011; Rinot et al., 2021). It has been reported in previous studies that the degraded molecules are generally low molecular weight compounds (Fujii et al., 2010; Ondrasek et al., 2019). However, the size exclusion chromatography showed a decrease in the contribution of low molecular weight molecules LMW < 1k Da, and also high molecular weight macromolecules HMW > 10 kDa at the same time. Based on these results, protein-like content of DOM is likely to be composed of both LMW and HMW fractions. The fraction of medium molecular weight (MMW) compounds, linked to C1, C2 and C3 PRAFAC components (lignin compounds and terrestrial derived DOM), did not vary significantly and seemed to be recalcitrant and not easily degradable.

4.2. Evolution of DOM characteristics from soils to adjacent streams

The evolution of soil extracted DOM during the incubation experiment in comparison with the adjacent watercourse DOM is shown in the principal component analysis (PCA), plotted with the 26 optical indicators calculated from UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence (Figure 5-8). The PCA shows that the optical fingerprints of forest and vineyard soil extracts are more distinct at the end of incubation experiment (82 days) than before. This could be explained by the nature and the proportion of labile vs refractory fractions of DOM before and after incubation. This suggests that the labile fraction of DOM is not so different between the two soil types. For all indicators, the evolution of DOM optical fingerprints during the incubation tends towards that of adjacent watercourse (Figure 5-8). The adjacent stream samples show relatively high contribution of medium weighted molecules MMW and a lower contribution of both low and high molecules of the protein and humic types (LMW and HMW). This could be explained by the fact that watercourse DOM has already undergone mineralization processes, mainly through microorganisms in top soil layers, before it reaches the adjacent watercourses. Other abiotic ways such as photodegradation and adsorption can influence DOM composition, which can explain the observed differences between samples at the end of incubation and the adjacent watercourse samples.

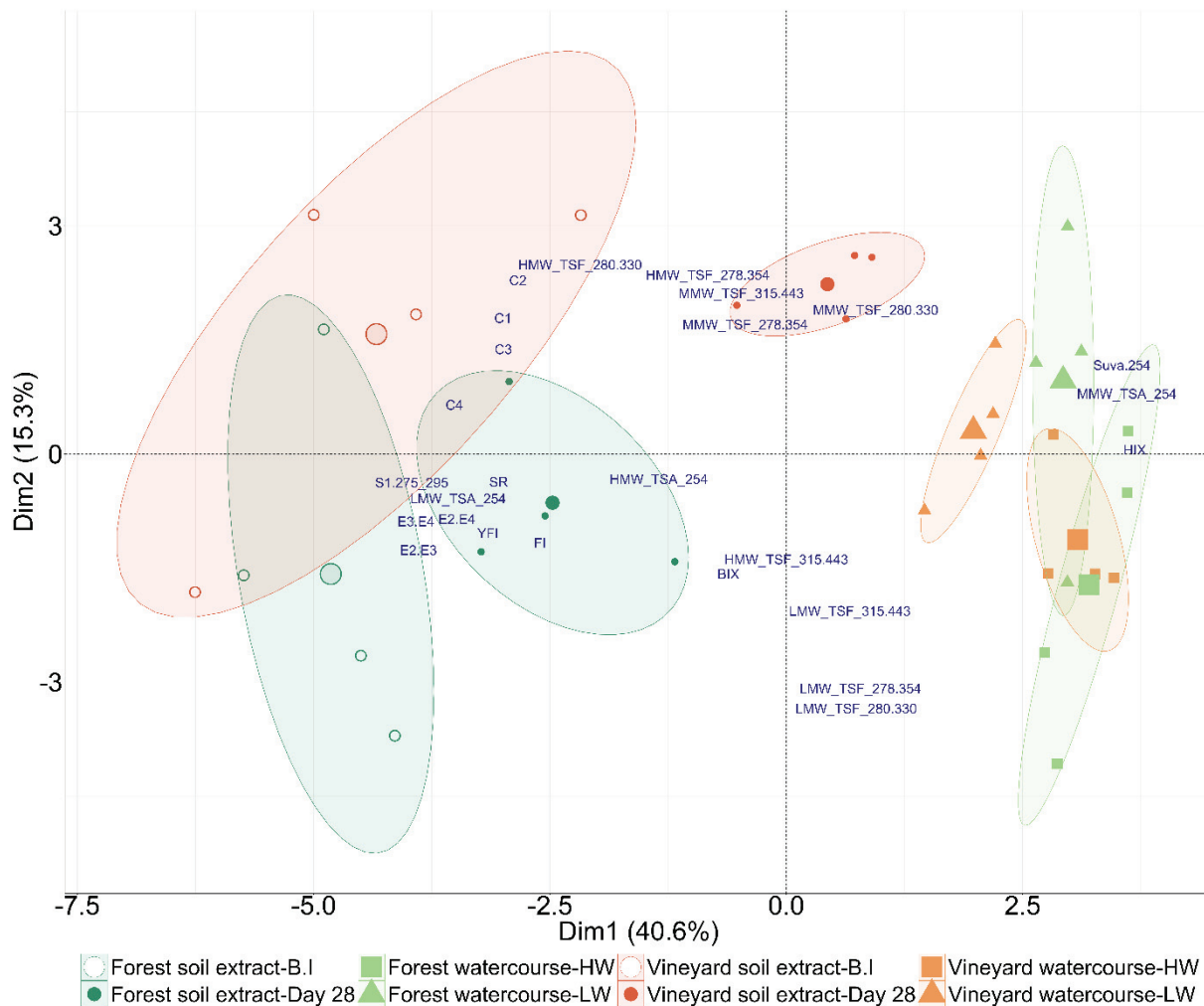


Figure 5-8. First factorial plane of the principal component analysis (PCA) performed on data from forest and vineyard soil extracts before incubation (B.I), after 28 days of incubation and their adjacent water stream samples collected during high and low water periods. The variables considered in the PCA are the UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence indicators.

The PCA also showed that water samples taken downstream of the forest soils seem more heterogeneous than downstream of the vineyard soils, whatever the hydrological regime. This heterogeneity could be explained by in-situ complex abiotic processes as well as by the difference in vegetation cover in the Ardières-Morcille watershed. Indeed, the forest sources considered in this study are mostly influenced by coniferous forests, but broad-leaved forests and natural grasslands can represent up to 35% of the total sub-catchment areas (calculation based on land use maps, CLC 2018). A previous study highlighted different DOM quality between deciduous and coniferous forests (Thieme et al., 2019). On the contrary, the viticultural sub-catchments are very homogeneous in terms of vegetation cover as they are

subject to strict local regulations. The hydrologic regime also seems to have an influence on DOM quality especially for vineyard adjacent watercourses. This could be explained by the contribution of an autochthonous source being more heightened during low water periods. Autochthonous DOM is essentially produced locally by the microbial activities (Münster, 1993) and it was demonstrated previously that there is an impact of agriculture on increasing carbon and nutrient inputs to downstream watercourses (Williams et al., 2010). These nutrient inputs provide energy support for microbial activity which can affect DOM quantity and quality.

4.3. How do the fingerprints match between DOM in soil extracts after incubation and that in downstream watercourses during high and low water periods?

As discussed in the previous section, hydrological regime influences the quantity and quality of the DOM in the watercourses. It is therefore more appropriate to look for markers of forest and vineyard soils downstream of the waterways draining these soils considering both high and low water sampling periods. To do this, a linear discriminant analysis (LDA) was applied in two steps. In a first step, the construction of the LDA 1 was conducted with 3 groups: i) a first group, called Forest DOM HW (high water), containing samples of degraded forest soils and adjacent forest streams collected at high water. ii) A second group, called Vineyard DOM HW, which contained samples of degraded vineyard soils with water from adjacent vineyard streams collected at high water, and iii) a third group called low water samples, which contained all the samples from adjacent watercourses taken during low water periods. The purpose of the first LDA (Figure 5-9.A) was to identify if there was similarities between DOM markers from soil extracts and DOM from the adjacent watercourse during high water period. In a second step, a LDA 2 was carried out by reversing the groups of water samples, thus considering i) a first group Forest DOM LW (low water), which contained the samples of forest soil extracts and forest adjacent watercourses taken during low water. ii) a second group of Vineyard DOM LW with samples from degraded vineyard soils and low water vineyard watercourses; and iii) a third group of high water samples containing all the samples of the adjacent watercourses taken during high water. The second LDA (Figure 5-9.B) was conducted to determine any similarities between soil extracts and adjacent watercourses during low water periods.

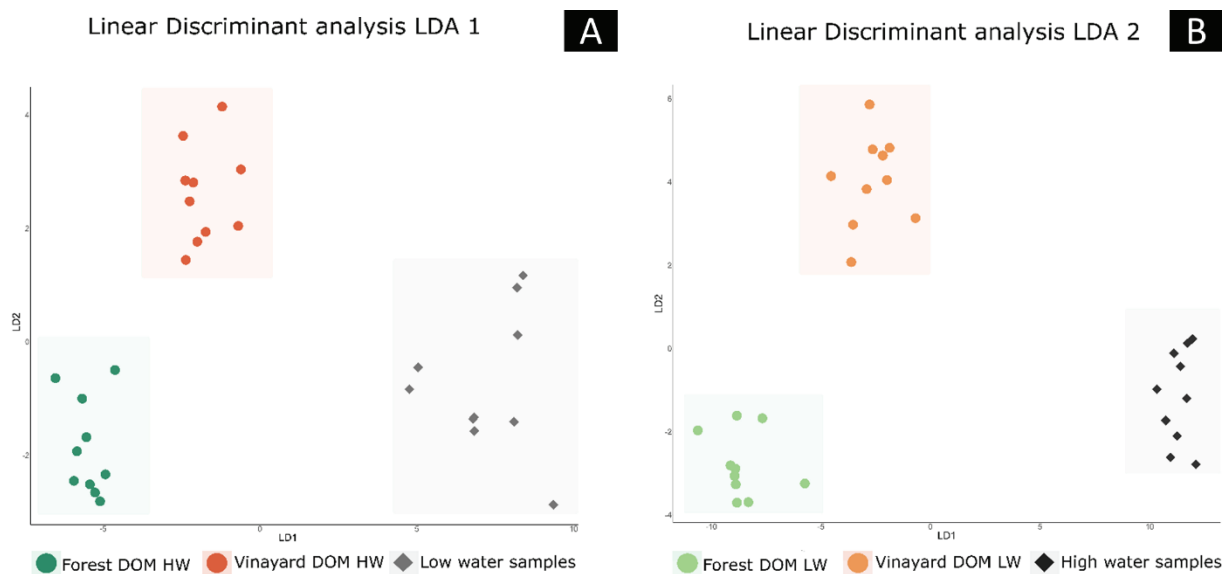


Figure 5-9. A. Linear discriminant analysis (LDA1) calculated with 26 indicators from UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence to discriminate forest and vineyard DOM taken during high water from low water DOM. B. Linear discriminant analysis (LDA2) calculated with the same indicators to discriminate forest and vineyard DOM taken during low water periods from High water DOM.

Among the 26 optical indicators (UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence) considered in this study, 6 indicators were able to correctly classify the samples into Forest DOM HW, Vineyard DOM HW and Low water samples groups (Figure 5-9). They correspond to one UV-Visible indicator (slope ratio SR), two EEM fluorescence indicators (HIX and PARAFAC component C1) and three HPSEC/UV-fluorescence indicators (LMW-TSA 254, MMW TSF 278-354 and MMW TSF 280-330). The results from LDA 2 also allowed the selection of 6 indicators that classified the samples into Forest DOM LW, Vineyard DOM LW and High water samples groups (Figure 5-9.B). There were one UV-Visible indicator ($SUVA_{254}$), three EEM fluorescence indicators (HIX, FI and YFI) and two HPSEC/UV-fluorescence indicators (MMW TSF 278-354 and MMW TSF 280-330). Through this supervised cross-sectional analysis, we have highlighted a series of indicators of terrigenous contribution in both high water and low water periods. These results reveal that: (i) there is an influence of land use on the quality of DOM in adjacent streams using selected optical indicators, and (ii) this influence is present as well in periods of high water as in periods of low water. During high water, DOM sampled downstream of forest soils comprise of relatively lower molecular weight compounds with low degrees of humification and higher contribution of protein-type molecules smaller than 1000 Da. On the contrary, the vineyard DOM carries a higher molecular weight, a higher degree of humification

and a lower contribution of small protein-type molecules. During the low water period, DOM sampled downstream of the forest soils has higher aromaticity and molecular weight than the vineyard DOM which contains relatively low molecules. Indeed, vineyard DOM could be impacted by higher microbial activity (Chavarria et al., 2021; Williams et al., 2010). The differences observed between DOM sampled downstream forest and vineyard regardless the hydrological conditions can be explained by several mechanisms. First, the presence of a biodegraded form of DOM in vineyard soils probably due to the intensive exploitation of the soils and the low rooting networks. Indeed, previous studies reported an higher organic matter oxidation in agricultural soils which are more prone to erosion than forest soils and to an accelerated phenomenon of open-air oxidation in the upper soil layers (Graeber et al., 2012; Shang et al., 2018). Secondly, given the very steep slopes that favor surface runoff in the Ardières-Morcille watershed, the differences observed between high and low water could be explained by a contribution of surface soil organic matter during high water and a contribution of deeper and more degraded soil organic matter during low water period.

It is important to note that DOM fingerprints presented in this study are not generalizable to other watersheds. Considerable differences can be observed from one watershed to another depending on topography, flow direction and soil characteristics. However, the fingerprints of DOM sampled in a stream downstream of a homogeneous watershed while accounting for hydrologic conditions provides a good representation of land use type. This sampling method can be applied to other land use types with different agricultural and livestock types and pressures. Moreover, given the complexity of DOM composition and reactivity, the combination of optical indicators was essential. The result of the LDA, when calculated with UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence indicators separately, could not provide a good differentiation between forest and vineyard DOM (Supplementary Material 5). At last, indicators from HPSEC/UV-fluorescence proposed in this study, have shown potential as good markers of diffuse sources of DOM in streams as discussed more in depth in the next part 4.4 .

4.4. Relevance of HPSEC/UV-fluorescence indicators to study DOM

The HPSEC/UV-fluorescence indicators presented in this study provided additional information regarding classical optical proxies (Table 1). For example, the SUVA indicator highlight the aromatic content of DOM and the LMW TSA 254 indicator reflects the

contribution of molecules lower than 1 k Da (Supplementary Material 6). A negative correlation (Pearson test, $p < 0.05$, $n=70$) between the two indicators demonstrates that the low molecular weight compounds detected at 254 nm are poorly aromatic and that the aromatic content of DOM is mostly found in compounds larger than 1 k Da. The correlation matrix (Supplementary Material 6) also shows that the indicator HMW TSF 280-330 is positively correlated with the component PARAFAC C4 associated with protein-like compounds. This suggests that the C4 component is composed of relatively large molecules (> 10 k Da). In the context of this study, this showed that DOM labile pool was also composed of high molecular weight compounds in contrast to what has been reported in previous studies (Curtis-Jackson et al., 2009; Derrien et al., 2019). Other indicators of HPSEC/UV-fluorescence (e.g. HMW TSF 278-354) point out the limitations of using classic indicators like slope ratio SR indicator (Table 1). Indeed, in some soil extract, the SR indicator showed the presence of DOM mostly consisting of low molecular weight compounds, while the HMW TSF 278-354 indicator showed a relatively high contribution of amino-acid type compounds larger than 10 k Da. This result suggests that the SR indicator could be limited to the description of certain types of molecules that exclude acid-amine like compounds. Finally, the indicators from HPSEC/UV-fluorescence could be used on a larger scale like in monitoring studies of agricultural, industrial and urban discharges. The information provided by these indicators would allow better identification of anthropogenic compounds (e.g. pesticides, pharmaceuticals) that are very complicated to detect using classical optical proxies as most of these compounds have almost similar optical properties (Hassoun et al., 2017). Also, the establishment of databases of HPSEC/UV-fluorescence indicators specific to anthropogenic compounds of interest, could facilitate their detection at lower cost than with current techniques using chromatography coupled with mass spectrometry.

5. Conclusion

The objective of this study was to provide insight into the sampling of terrigenous diffuse sources of DOM directly into streams using relevant land use markers. These markers were identified by comparing DOM from soil extracts with adjacent watercourse DOM, sampled at the outlet of homogenous sub-catchments during low and high water.

The UV-Visible, EEM fluorescence and HPSEC/UV-fluorescence techniques showed an evolution of DOM extracted from forest and vineyard soils during a 28-day incubation experiment. Indicators calculated from these techniques showed that the evolution of DOM mainly occurred in the three first days of incubation. This resulted in the degradation of high (>10 k Da) and low (<1 k Da) MW protein-like molecules (labile fraction of DOM), leading to an increase of aromaticity and humification mostly observed in medium molecular weight compounds (from 1 to 10 k Da). The labile fraction of DOM did not differ according to the land use.

Adjacent stream samples had distinct optical fingerprints influenced mostly by the hydrological regime, as DOM sampled during high water periods showed a large contribution of LMW and HMW than DOM sampled during low water periods. At the end of the incubation study, forest and vineyard soil extract tend to fingerprints of the adjacent watercourse samples. Linear discriminant analysis (LDA) revealed the relevance of 9 indicators to differentiate forest DOM and vineyard DOM fingerprints during high and low water sampling. These results demonstrate that a representative sampling of a diffuse source of DOM in rivers should take into account both the homogeneity of the upstream sub-catchments in terms of land use but also the hydrological conditions. The cross exploitation of the data from these three techniques seems appropriate to establish DOM fingerprints. However, the fingerprints of forest and vineyard diffuse sources studied in this work are not exhaustive contrary to the methodological approach, which can be applicable to other types of diffuse sources. Indeed, application of this approach on a larger scale would provide specific fingerprints for terrigenous diffuse sources in streams.

Acknowledgements

This study was part of the “CHYPSTER” project supported by the French National Research Agency (ANR-21-CE34-0013-01) and “IDESOC” project granted by Zone Atelier Bassin du Rhône (ZABR) and Rhône Méditerranée Corse Water Agency (AERMC). The authors wish to acknowledge Loïc Richard and Alexandra Gruat for their assistance in field sampling, samples preparation and chemical analysis.

Credit Author Statement

Amine Boukra: Conceptualization, Methodology, Software, Investigation, Formal analysis, Data curation, Writing original draft, Visualization. **Mathieu Masson:** Conceptualization, Writing-review and editing, Supervision, Project administration, Funding acquisition. **Corinne Brosse:** Resources. **Mahaut Sourzac :** Resources. **Edith Parlanti:** Software, Writing-review and editing, Resources. **Cécile Miege:** Conceptualization, Writing-review and editing, Supervision, Project administration, Funding acquisition.

6. References

- Aiken, G.R., 2014. 1.11 - Dissolved Organic Matter in Aquatic Systems, in: Ahuja, S. (Ed.), *Comprehensive Water Quality and Purification*. Elsevier, Waltham, pp. 205–220. <https://doi.org/10.1016/B978-0-12-382182-9.00014-1>
- Asmala, E., Massicotte, P., Carstensen, J., 2021. Identification of dissolved organic matter size components in freshwater and marine environments. *Limnology and Oceanography* 66, 1381–1393. <https://doi.org/10.1002/lno.11692>
- Berg, S.M., Whiting, Q.T., Herrli, J.A., Winkels, R., Wammer, K.H., Remucal, C.K., 2019. The Role of Dissolved Organic Matter Composition in Determining Photochemical Reactivity at the Molecular Level. *Environ. Sci. Technol.* 53, 11725–11734. <https://doi.org/10.1021/acs.est.9b03007>
- Bolan, N.S., Adriano, D.C., Kunhikrishnan, A., James, T., McDowell, R., Senesi, N., 2011. Chapter One - Dissolved Organic Matter: Biogeochemistry, Dynamics, and Environmental Significance in Soils, in: Sparks, D.L. (Ed.), *Advances in Agronomy*. Academic Press, pp. 1–75. <https://doi.org/10.1016/B978-0-12-385531-2.00001-3>
- Boukra, A., Masson, M., Brosse, C., Sourzac, M., Parlanti, E., Miège, C., 2023. Sampling terrigenous diffuse sources in watercourse: Influence of land use and hydrological conditions on dissolved organic matter characteristics. *Science of The Total Environment* 162104. <https://doi.org/10.1016/j.scitotenv.2023.162104>
- Brezinski, K., Gorczyca, B., 2019. An overview of the uses of high performance size exclusion chromatography (HPSEC) in the characterization of natural organic matter (NOM) in potable water, and ion-exchange applications. *Chemosphere* 217, 122–139. <https://doi.org/10.1016/j.chemosphere.2018.10.028>
- Carson, J.K., Rooney, D., Gleeson, D.B., Clipson, N., 2007. Altering the mineral composition of soil causes a shift in microbial community structure. *FEMS Microbiology Ecology* 61, 414–423. <https://doi.org/10.1111/j.1574-6941.2007.00361.x>
- Carstea, E.M., Popa, C.L., Baker, A., Bridgeman, J., 2020. In situ fluorescence measurements of dissolved organic matter: A review. *Science of The Total Environment* 699, 134361. <https://doi.org/10.1016/j.scitotenv.2019.134361>
- Catalá, T.S., Shorte, S., Dittmar, T., 2021. Marine dissolved organic matter: a vast and unexplored molecular space. *Appl Microbiol Biotechnol* 105, 7225–7239. <https://doi.org/10.1007/s00253-021-11489-3>
- Cawley, K.M., Ding, Y., Fourqurean, J., Jaffé, R., 2012. Characterising the sources and fate of dissolved organic matter in Shark Bay, Australia: a preliminary study using optical properties and stable carbon isotopes. *Mar. Freshwater Res.* 63, 1098. <https://doi.org/10.1071/MF12028>
- Chavarria, K.A., Saltonstall, K., Vinda, J., Batista, J., Lindmark, M., Stallard, R.F., Hall, J.S., 2021. Land use influences stream bacterial communities in lowland tropical watersheds. *Sci Rep* 11, 21752. <https://doi.org/10.1038/s41598-021-01193-7>
- Chen, X., Yu, M., He, X., Zheng, M., Xi, B., Sun, Y., Fu, X., Su, J., 2021. Fate of dissolved organic matter substructure in a full-scale wastewater treatment plant by using size exclusion chromatography multi-excitation-emission matrix analysis. *Journal of Cleaner Production* 328, 129677. <https://doi.org/10.1016/j.jclepro.2021.129677>
- Claret, F., Schäfer, T., Bauer, A., Buckau, G., 2003. Generation of humic and fulvic acid from Callovo-Oxfordian clay under high alkaline conditions. *Science of The Total Environment* 317, 189–200. [https://doi.org/10.1016/S0048-9697\(03\)00337-1](https://doi.org/10.1016/S0048-9697(03)00337-1)
- Coble, A.A., Wymore, A.S., Potter, J.D., McDowell, W.H., 2022. Land Use Overrides Stream Order and Season in Driving Dissolved Organic Matter Dynamics Throughout the Year in a River Network. *Environ. Sci. Technol.* 56, 2009–2020. <https://doi.org/10.1021/acs.est.1c06305>

- Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry* 51, 325–346. [https://doi.org/10.1016/0304-4203\(95\)00062-3](https://doi.org/10.1016/0304-4203(95)00062-3)
- Cory, R.M., McKnight, D.M., 2005. Fluorescence Spectroscopy Reveals Ubiquitous Presence of Oxidized and Reduced Quinones in Dissolved Organic Matter. *Environ. Sci. Technol.* 39, 8142–8149. <https://doi.org/10.1021/es0506962>
- Coulson, L.E., Weigelhofer, G., Gill, S., Hein, T., Griebler, C., Schelker, J., 2022. Small rain events during drought alter sediment dissolved organic carbon leaching and respiration in intermittent stream sediments. *Biogeochemistry* 159, 159–178. <https://doi.org/10.1007/s10533-022-00919-7>
- Curtis-Jackson, P.K., Massé, G., Gledhill, M., Fitzsimons, M.F., 2009. Characterization of low molecular weight dissolved organic nitrogen by liquid chromatography-electrospray ionization-mass spectrometry. *Limnology and Oceanography: Methods* 7, 52–63. <https://doi.org/10.4319/lom.2009.7.52>
- Dainard, P.G., Guéguen, C., McDonald, N., Williams, W.J., 2015. Photobleaching of fluorescent dissolved organic matter in Beaufort Sea and North Atlantic Subtropical Gyre. *Marine Chemistry* 177, 630–637. <https://doi.org/10.1016/j.marchem.2015.10.004>
- D’Andrilli, J., Foreman, C.M., Sigl, M., Priscu, J.C., McConnell, J.R., 2017. A 21 000-year record of fluorescent organic matter markers in the WAIS Divide ice core. *Clim. Past* 13, 533–544. <https://doi.org/10.5194/cp-13-533-2017>
- Derrien, M., Kim, M.-S., Ock, G., Hong, S., Cho, J., Shin, K.-H., Hur, J., 2018. Estimation of different source contributions to sediment organic matter in an agricultural-forested watershed using end member mixing analyses based on stable isotope ratios and fluorescence spectroscopy. *Science of The Total Environment* 618, 569–578. <https://doi.org/10.1016/j.scitotenv.2017.11.067>
- Derrien, M., Shin, K.-H., Hur, J., 2019. Biodegradation-induced signatures in sediment pore water dissolved organic matter: Implications from artificial sediments composed of two contrasting sources. *Science of The Total Environment* 694, 133714. <https://doi.org/10.1016/j.scitotenv.2019.133714>
- Du, Y., Chen, F., Xiao, K., Song, C., He, H., Zhang, Q., Zhou, Y., Jang, K.-S., Zhang, YiBo, Xing, P., Liu, Z., Zhang, YunLin, Lu, Y., 2021a. Water Residence Time and Temperature Drive the Dynamics of Dissolved Organic Matter in Alpine Lakes in the Tibetan Plateau. *Global Biogeochemical Cycles* 35, e2020GB006908. <https://doi.org/10.1029/2020GB006908>
- Du, Y., Lu, Y., Roebuck, J.A., Liu, D., Chen, F., Zeng, Q., Xiao, K., He, H., Liu, Z., Zhang, Y., Jaffé, R., 2021b. Direct versus indirect effects of human activities on dissolved organic matter in highly impacted lakes. *Science of The Total Environment* 752, 141839. <https://doi.org/10.1016/j.scitotenv.2020.141839>
- Eder, A., Weigelhofer, G., Pucher, M., Tiefenbacher, A., Strauss, P., Brandl, M., Blöschl, G., 2022. Pathways and composition of dissolved organic carbon in a small agricultural catchment during base flow conditions. *Ecohydrology & Hydrobiology* 22, 96–112. <https://doi.org/10.1016/j.ecohyd.2021.07.012>
- Eilers, K.G., Lauber, C.L., Knight, R., Fierer, N., 2010. Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. *Soil biology & biochemistry*.
- Escribano-Viana, R., López-Alfaro, I., López, R., Santamaría, P., Gutiérrez, A.R., González-Arenzana, L., 2018. Impact of Chemical and Biological Fungicides Applied to Grapevine on Grape Biofilm, Must, and Wine Microbial Diversity. *Frontiers in Microbiology* 9.
- Fellman, J.B., D’Amore, D.V., Hood, E., Boone, R.D., 2008. Fluorescence characteristics and biodegradability of dissolved organic matter in forest and wetland soils from coastal temperate watersheds in southeast Alaska. *Biogeochemistry* 88, 169–184. <https://doi.org/10.1007/s10533-008-9203-x>
- Fujii, K., Hayakawa, C., Van Hees, P.A.W., Funakawa, S., Kosaki, T., 2010. Biodegradation of low molecular weight organic compounds and their contribution to heterotrophic soil respiration

- in three Japanese forest soils. *Plant Soil* 334, 475–489. <https://doi.org/10.1007/s11104-010-0398-y>
- Garcia, R.D., Diéguez, M. del C., Gereá, M., Garcia, P.E., Reissig, M., 2018. Characterisation and reactivity continuum of dissolved organic matter in forested headwater catchments of Andean Patagonia. *Freshwater Biology* 63, 1049–1062. <https://doi.org/10.1111/fwb.13114>
- Gouy, V., Liger, L., Ahrouch, S., Bonnineau, C., Carluer, N., Chaumot, A., Coquery, M., Dabrin, A., Margoum, C., Pesce, S., 2021. Ardières-Morcille in the Beaujolais, France: A research catchment dedicated to study of the transport and impacts of diffuse agricultural pollution in rivers. *Hydrological Processes* 35, e14384. <https://doi.org/10.1002/hyp.14384>
- Graeber, D., Gelbrecht, J., Pusch, M.T., Anlanger, C., von Schiller, D., 2012. Agriculture has changed the amount and composition of dissolved organic matter in Central European headwater streams. *Science of The Total Environment* 438, 435–446. <https://doi.org/10.1016/j.scitotenv.2012.08.087>
- Guarch-Ribot, A., Butturini, A., 2016. Hydrological conditions regulate dissolved organic matter quality in an intermittent headwater stream. From drought to storm analysis. *Science of The Total Environment* 571, 1358–1369. <https://doi.org/10.1016/j.scitotenv.2016.07.060>
- Guigue, J., Mathieu, O., Lévêque, J., Mounier, S., Laffont, R., Maron, P.-A., Navarro, N., Chateau-Smith, C., Suchet, P., Lucas, Y., 2014. A comparison of extraction procedures for water-extractable organic matter in soils. *European Journal of Soil Science* 65, 520–530. <https://doi.org/10.1111/ejss.12156>
- Hansen, A.M., Kraus, T.E.C., Pellerin, B.A., Fleck, J.A., Downing, B.D., Bergamaschi, B.A., 2016. Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. *Limnology and Oceanography* 61, 1015–1032. <https://doi.org/10.1002/lno.10270>
- Hassoun, H., Lamhasni, T., Foudeil, S., El Bakkali, A., Ait Lyazidi, S., Haddad, M., Choukrad, M., Hnach, M., 2017. Total Fluorescence Fingerprinting of Pesticides: A Reliable Approach for Continuous Monitoring of Soils and Waters. *J Fluoresc* 27, 1633–1642. <https://doi.org/10.1007/s10895-017-2100-8>
- Hawkes, J.A., Sjöberg, P.J.R., Bergquist, J., Tranvik, L.J., 2019. Complexity of dissolved organic matter in the molecular size dimension: insights from coupled size exclusion chromatography electrospray ionisation mass spectrometry. *Faraday Discuss.* 218, 52–71. <https://doi.org/10.1039/C8FD00222C>
- Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography* 53, 955–969. <https://doi.org/10.4319/lo.2008.53.3.0955>
- Heo, J., Yoon, Y., Kim, D.-H., Lee, H., Lee, D., Her, N., 2016. A new fluorescence index with a fluorescence excitation-emission matrix for dissolved organic matter (DOM) characterization. *Desalination and Water Treatment* 57, 20270–20282. <https://doi.org/10.1080/19443994.2015.1110719>
- Hu, A., Choi, M., Tanentzap, A.J., Liu, J., Jang, K.-S., Lennon, J.T., Liu, Y., Soininen, J., Lu, X., Zhang, Y., Shen, J., Wang, J., 2022. Ecological networks of dissolved organic matter and microorganisms under global change. *Nat Commun* 13, 3600. <https://doi.org/10.1038/s41467-022-31251-1>
- Huang, Y., Du, Y., Ma, T., Deng, Y., Tao, Y., Xu, Y., Leng, Z., 2021. Dissolved organic matter characterization in high and low ammonium groundwater of Dongting Plain, central China. *Ecotoxicology and Environmental Safety* 208, 111779. <https://doi.org/10.1016/j.ecoenv.2020.111779>
- Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J.-M., Parlanti, E., 2009. Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Organic Geochemistry* 40, 706–719. <https://doi.org/10.1016/j.orggeochem.2009.03.002>
- Jaffé, R., Boyer, J.N., Lu, X., Maie, N., Yang, C., Scully, N.M., Mock, S., 2004. Source characterization of dissolved organic matter in a subtropical mangrove-dominated estuary by fluorescence analysis. *Marine Chemistry* 84, 195–210. <https://doi.org/10.1016/j.marchem.2003.08.001>

- Jangid, K., Williams, M.A., Franzluebbers, A.J., Sanderlin, J.S., Reeves, J.H., Jenkins, M.B., Endale, D.M., Coleman, D.C., Whitman, W.B., 2008. Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biology and Biochemistry* 40, 2843–2853. <https://doi.org/10.1016/j.soilbio.2008.07.030>
- Jørgensen, L., Stedmon, C.A., Kragh, T., Markager, S., Middelboe, M., Søndergaard, M., 2011. Global trends in the fluorescence characteristics and distribution of marine dissolved organic matter. *Marine Chemistry* 126, 139–148. <https://doi.org/10.1016/j.marchem.2011.05.002>
- Kalbitz, K., Schmerwitz, J., Schwesig, D., Matzner, E., 2003. Biodegradation of soil-derived dissolved organic matter as related to its properties. *Geoderma, Ecological aspects of dissolved organic matter in soils* 113, 273–291. [https://doi.org/10.1016/S0016-7061\(02\)00365-8](https://doi.org/10.1016/S0016-7061(02)00365-8)
- Kida, M., Kojima, T., Tanabe, Y., Hayashi, K., Kudoh, S., Maie, N., Fujitake, N., 2019. Origin, distributions, and environmental significance of ubiquitous humic-like fluorophores in Antarctic lakes and streams. *Water Research* 163, 114901. <https://doi.org/10.1016/j.watres.2019.114901>
- Lambert, T., Bouillon, S., Darchambeau, F., Morana, C., Roland, F.A.E., Descy, J.-P., Borges, A.V., 2017. Effects of human land use on the terrestrial and aquatic sources of fluvial organic matter in a temperate river basin (The Meuse River, Belgium). *Biogeochemistry* 136, 191–211. <https://doi.org/10.1007/s10533-017-0387-9>
- Lee, S.-A., Kim, T.-H., Kim, G., 2020. Tracing terrestrial versus marine sources of dissolved organic carbon in a coastal bay using stable carbon isotopes. *Biogeosciences* 17, 135–144. <https://doi.org/10.5194/bg-17-135-2020>
- Li, P., Hur, J., 2017. Utilization of UV-Vis spectroscopy and related data analyses for dissolved organic matter (DOM) studies: A review. *Critical Reviews in Environmental Science and Technology* 47. <https://doi.org/10.1080/10643389.2017.1309186>
- Liu, Z., Sheng, L., Zhang, X., Duan, L., Jiang, Y., Xiao, Q., 2023. Dissolved Organic Carbon Dynamics Variability from Ponds Draining Different Landscapes in a Typical Agricultural Watershed. *Atmosphere* 14, 363. <https://doi.org/10.3390/atmos14020363>
- Maccarthy, P., 2001. The Principles of Humic Substances. *Soil Science* 166, 738–751. <https://doi.org/10.1097/00010694-200111000-00003>
- Marschner, B., Brodowski, S., Dreves, A., Gleixner, G., Gude, A., Grootes, P.M., Hamer, U., Heim, A., Jandl, G., Ji, R., Kaiser, K., Kalbitz, K., Kramer, C., Leinweber, P., Rethemeyer, J., Schäffer, A., Schmidt, M.W.I., Schwark, L., Wiesenberg, G.L.B., 2008. How relevant is recalcitrance for the stabilization of organic matter in soils? *Journal of Plant Nutrition and Soil Science* 171, 91–110. <https://doi.org/10.1002/jpln.200700049>
- Massicotte, P., Markager, S., 2016. Using a Gaussian decomposition approach to model absorption spectra of chromophoric dissolved organic matter. *Marine Chemistry* 180, 24–32. <https://doi.org/10.1016/j.marchem.2016.01.008>
- McDowell, W.H., Zsolnay, A., Aitkenhead-Peterson, J.A., Gregorich, E.G., Jones, D.L., Jödemann, D., Kalbitz, K., Marschner, B., Schwesig, D., 2006. A comparison of methods to determine the biodegradable dissolved organic carbon from different terrestrial sources. *Soil Biology and Biochemistry* 38, 1933–1942. <https://doi.org/10.1016/j.soilbio.2005.12.018>
- McEnroe, N.A., Williams, C.J., Xenopoulos, M.A., Porcal, P., Frost, P.C., 2013. Distinct Optical Chemistry of Dissolved Organic Matter in Urban Pond Ecosystems. *PLOS ONE* 8, e80334. <https://doi.org/10.1371/journal.pone.0080334>
- Mcknight, D., Boyer, E., Westerhoff, P., Doran, P., Kulbe, T., Andersen, D., 2001. Spectrofluorometric Characterization of Dissolved Organic Matter for Indication of Precursor Organic Material and Aromaticity. *Limnology and Oceanography* 46, 38–48. <https://doi.org/10.4319/lo.2001.46.1.0038>
- Michael-Kordatou, I., Michael, C., Duan, X., He, X., Dionysiou, D.D., Mills, M.A., Fatta-Kassinos, D., 2015. Dissolved effluent organic matter: Characteristics and potential implications in wastewater treatment and reuse applications. *Water Research* 77, 213–248. <https://doi.org/10.1016/j.watres.2015.03.011>

- Mladenov, N., Parsons, D., Kinoshita, A.M., Pinongcos, F., Mueller, M., Garcia, D., Lipson, D.A., Grijalva, L.M., Zink, T.A., 2022. Groundwater-surface water interactions and flux of organic matter and nutrients in an urban, Mediterranean stream. *Science of The Total Environment* 811, 152379. <https://doi.org/10.1016/j.scitotenv.2021.152379>
- Mosher, J.J., Klein, G.C., Marshall, A.G., Findlay, R.H., 2010. Influence of bedrock geology on dissolved organic matter quality in stream water. *Organic Geochemistry* 41, 1177–1188. <https://doi.org/10.1016/j.orggeochem.2010.08.004>
- Münster, U., 1993. Concentrations and fluxes of organic carbon substrates in the aquatic environment. *Antonie Van Leeuwenhoek* 63, 243–274. <https://doi.org/10.1007/BF00871222>
- Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Anal. Methods* 5, 6557–6566. <https://doi.org/10.1039/C3AY41160E>
- Murphy, K.R., Stedmon, C.A., Waite, T.D., Ruiz, G.M., 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Marine Chemistry* 108, 40–58. <https://doi.org/10.1016/j.marchem.2007.10.003>
- Nebbioso, A., Piccolo, A., 2013. Molecular characterization of dissolved organic matter (DOM): a critical review. *Anal Bioanal Chem* 405, 109–124. <https://doi.org/10.1007/s00216-012-6363-2>
- Olshansky, Y., Root, R.A., Chorover, J., 2018. Wet–dry cycles impact DOM retention in subsurface soils. *Biogeosciences* 15, 821–832. <https://doi.org/10.5194/bg-15-821-2018>
- Ondrasek, G., Bakić Begić, H., Zovko, M., Filipović, L., Meriño-Gergichevich, C., Savić, R., Rengel, Z., 2019. Biogeochemistry of soil organic matter in agroecosystems & environmental implications. *Science of The Total Environment* 658, 1559–1573. <https://doi.org/10.1016/j.scitotenv.2018.12.243>
- Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M., 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Organic Geochemistry* 31, 1765–1781. [https://doi.org/10.1016/S0146-6380\(00\)00124-8](https://doi.org/10.1016/S0146-6380(00)00124-8)
- Peleato, N.M., Legge, R.L., Andrews, R.C., 2018. Neural networks for dimensionality reduction of fluorescence spectra and prediction of drinking water disinfection by-products. *Water Research* 136, 84–94. <https://doi.org/10.1016/j.watres.2018.02.052>
- Peyrard, X., Liger, L., Guillemain, C., Gouy, V., 2016. A trench study to assess transfer of pesticides in subsurface lateral flow for a soil with contrasting texture on a sloping vineyard in Beaujolais. *Environ Sci Pollut Res* 23, 14–22. <https://doi.org/10.1007/s11356-015-4917-5>
- Qi, G., Zhang, B., Tian, B., Yang, R., Baker, A., Wu, P., He, S., 2023. Characterization of Dissolved Organic Matter from Agricultural and Livestock Effluents: Implications for Water Quality Monitoring. *International Journal of Environmental Research and Public Health* 20, 5121. <https://doi.org/10.3390/ijerph20065121>
- Rabiet, M., Coquery, M., Carluer, N., Gahou, J., Gouy, V., 2015. Transfer of metal(loid)s in a small vineyard catchment: contribution of dissolved and particulate fractions in river for contrasted hydrological conditions. *Environ Sci Pollut Res* 22, 19224–19239. <https://doi.org/10.1007/s11356-015-5079-1>
- Raczka, N.C., Piñeiro, J., Tfaily, M.M., Chu, R.K., Lipton, M.S., Pasa-Tolic, L., Morrissey, E., Brzostek, E., 2021. Interactions between microbial diversity and substrate chemistry determine the fate of carbon in soil. *Sci Rep* 11, 19320. <https://doi.org/10.1038/s41598-021-97942-9>
- Retelletti Brogi, S., Derrien, M., Hur, J., 2019. In-Depth Assessment of the Effect of Sodium Azide on the Optical Properties of Dissolved Organic Matter. *J Fluoresc* 29, 877–885. <https://doi.org/10.1007/s10895-019-02398-w>
- Reuter, J.H., Perdue, E.M., 1977. Importance of heavy metal-organic matter interactions in natural waters. *Geochimica et Cosmochimica Acta* 41, 325–334. [https://doi.org/10.1016/0016-7037\(77\)90240-X](https://doi.org/10.1016/0016-7037(77)90240-X)
- Rinot, O., Borisover, M., Levy, G.J., Eshel, G., 2021. Fluorescence spectroscopy: A sensitive tool for identifying land-use and climatic region effects on the characteristics of water-extractable soil

- organic matter. *Ecological Indicators* 121, 107103. <https://doi.org/10.1016/j.ecolind.2020.107103>
- Sanderman, J., Lohse, K.A., Baldock, J.A., Amundson, R., 2009. Linking soils and streams: Sources and chemistry of dissolved organic matter in a small coastal watershed. *Water Resources Research* 45. <https://doi.org/10.1029/2008WR006977>
- Schwarz, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461–464.
- Shang, P., Lu, Y., Du, Y., Jaffé, R., Findlay, R.H., Wynn, A., 2018. Climatic and watershed controls of dissolved organic matter variation in streams across a gradient of agricultural land use. *Science of The Total Environment* 612, 1442–1453. <https://doi.org/10.1016/j.scitotenv.2017.08.322>
- Smith, M.A., Kominoski, J.S., Gaiser, E.E., Price, R.M., Troxler, T.G., 2021. Stormwater Runoff and Tidal Flooding Transform Dissolved Organic Matter Composition and Increase Bioavailability in Urban Coastal Ecosystems. *Journal of Geophysical Research: Biogeosciences* 126, e2020JG006146. <https://doi.org/10.1029/2020JG006146>
- Stevenson, F.J., 1995. *Humus Chemistry: Genesis, Composition, Reactions*, Second Edition. *J. Chem. Educ.* 72, A93. <https://doi.org/10.1021/ed072pA93.6>
- Thieme, L., Graeber, D., Hofmann, D., Bischoff, S., Schwarz, M.T., Steffen, B., Meyer, U.-N., Kaupenjohann, M., Wilcke, W., Michalzik, B., Siemens, J., 2019. Dissolved organic matter characteristics of deciduous and coniferous forests with variable management: different at the source, aligned in the soil. *Biogeosciences* 16, 1411–1432. <https://doi.org/10.5194/bg-16-1411-2019>
- Toming, K., Tuvikene, L., Vilbaste, S., Agasild, H., Viik, M., Kisand, A., Feldmann, T., Martma, T., Jones, R.I., Nöges, T., 2013. Contributions of autochthonous and allochthonous sources to dissolved organic matter in a large, shallow, eutrophic lake with a highly calcareous catchment. *Limnology and Oceanography* 58, 1259–1270. <https://doi.org/10.4319/lo.2013.58.4.1259>
- Verkh, Y., Rozman, M., Petrovic, M., 2018. Extraction and cleansing of data for a non-targeted analysis of high-resolution mass spectrometry data of wastewater. *MethodsX* 5, 395–402. <https://doi.org/10.1016/j.mex.2018.04.008>
- Wauthy, M., Rautio, M., Christoffersen, K.S., Forsström, L., Laurion, I., Mariash, H.L., Peura, S., Vincent, W.F., 2018. Increasing dominance of terrigenous organic matter in circumpolar freshwaters due to permafrost thaw. *Limnol Oceanogr Lett* 3, 186–198. <https://doi.org/10.1002/lo2.10063>
- Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Fram, M.S., Fujii, R., Mopper, K., 2003. Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. *Environ. Sci. Technol.* 37, 4702–4708. <https://doi.org/10.1021/es030360x>
- Williams, C.J., Frost, P.C., Morales-Williams, A.M., Larson, J.H., Richardson, W.B., Chiandet, A.S., Xenopoulos, M.A., 2016. Human activities cause distinct dissolved organic matter composition across freshwater ecosystems. *Global Change Biology* 22, 613–626. <https://doi.org/10.1111/gcb.13094>
- Williams, C.J., Yamashita, Y., Wilson, H.F., Jaffé, R., Xenopoulos, M.A., 2010. Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. *Limnology and Oceanography* 55, 1159–1171. <https://doi.org/10.4319/lo.2010.55.3.1159>
- Wilson, H.F., Saiers, J.E., Raymond, P.A., Sobczak, W.V., 2013. Hydrologic Drivers and Seasonality of Dissolved Organic Carbon Concentration, Nitrogen Content, Bioavailability, and Export in a Forested New England Stream. *Ecosystems* 16, 604–616. <https://doi.org/10.1007/s10021-013-9635-6>
- Wünsch, U.J., Geuer, J.K., Lechtenfeld, O.J., Koch, B.P., Murphy, K.R., Stedmon, C.A., 2018. Quantifying the impact of solid-phase extraction on chromophoric dissolved organic matter composition. *Marine Chemistry* 207, 33–41. <https://doi.org/10.1016/j.marchem.2018.08.010>

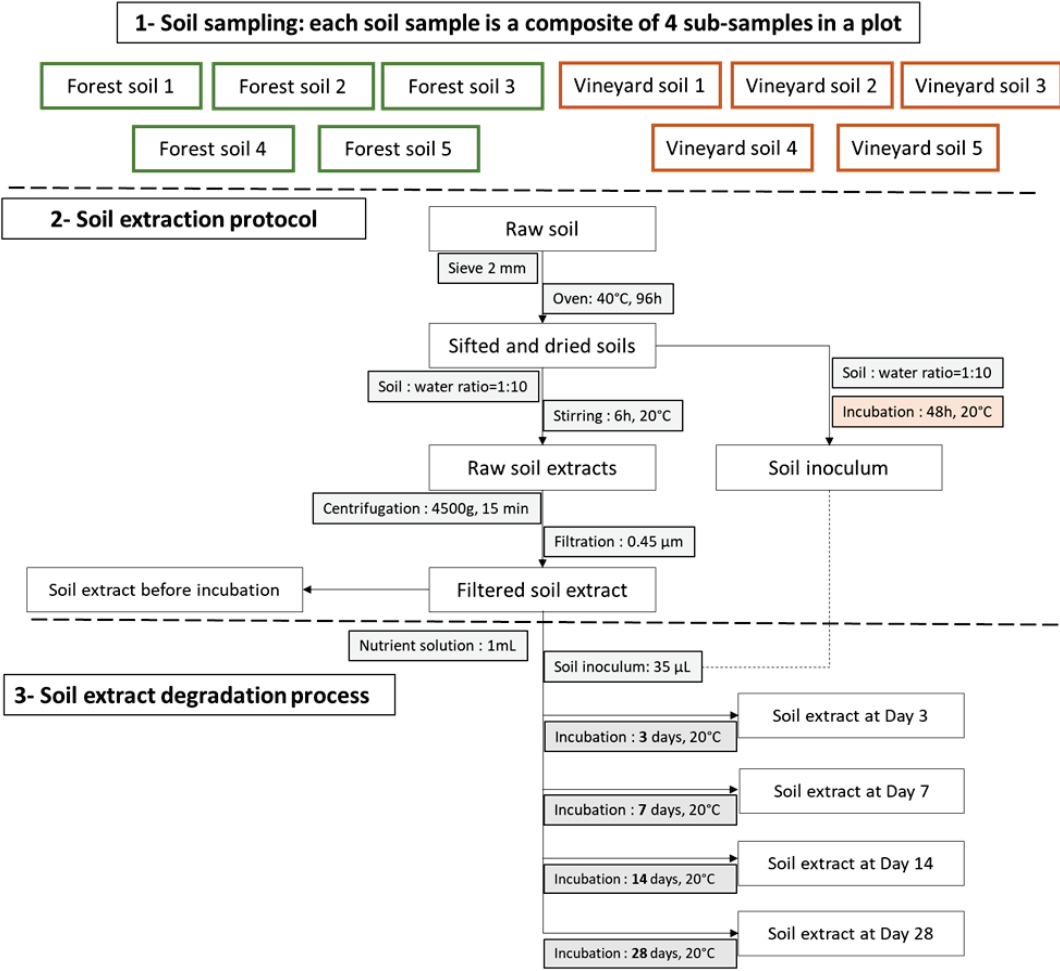
- Wünsch, U.J., Murphy, K., 2021. A simple method to isolate fluorescence spectra from small dissolved organic matter datasets. *Water Research* 190, 116730. <https://doi.org/10.1016/j.watres.2020.116730>
- Yamashita, Y., Cory, R.M., Nishioka, J., Kuma, K., Tanoue, E., Jaffé, R., 2010. Fluorescence characteristics of dissolved organic matter in the deep waters of the Okhotsk Sea and the northwestern North Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography, Ecological and Biogeochemical Interactions in the Dark Ocean* 57, 1478–1485. <https://doi.org/10.1016/j.dsr2.2010.02.016>
- Yamashita, Y., Kojima, D., Yoshida, N., Shibata, H., 2021. Relationships between dissolved black carbon and dissolved organic matter in streams. *Chemosphere* 271, 129824. <https://doi.org/10.1016/j.chemosphere.2021.129824>
- Zark, M., Dittmar, T., 2018. Universal molecular structures in natural dissolved organic matter. *Nat Commun* 9, 1–8. <https://doi.org/10.1038/s41467-018-05665-9>
- Zhang, X., Li, Y., Ye, J., Chen, Z., Ren, D., Zhang, S., 2022. The spectral characteristics and cadmium complexation of soil dissolved organic matter in a wide range of forest lands. *Environmental Pollution* 299, 118834. <https://doi.org/10.1016/j.envpol.2022.118834>
- Zhao, L., Li, N., Huang, T., Zhang, H., Si, F., Li, K., Qi, Y., Hua, F., Huang, C., 2022. Effects of artificially induced complete mixing on dissolved organic matter in a stratified source water reservoir. *Journal of Environmental Sciences* 111, 130–140. <https://doi.org/10.1016/j.jes.2021.03.024>
- Zsolnay, Á., 2003. Dissolved organic matter: artefacts, definitions, and functions. *Geoderma, Ecological aspects of dissolved organic matter in soils* 113, 187–209. [https://doi.org/10.1016/S0016-7061\(02\)00361-0](https://doi.org/10.1016/S0016-7061(02)00361-0)
- Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere* 38, 45–50. [https://doi.org/10.1016/S0045-6535\(98\)00166-0](https://doi.org/10.1016/S0045-6535(98)00166-0)

7. Supplementary material

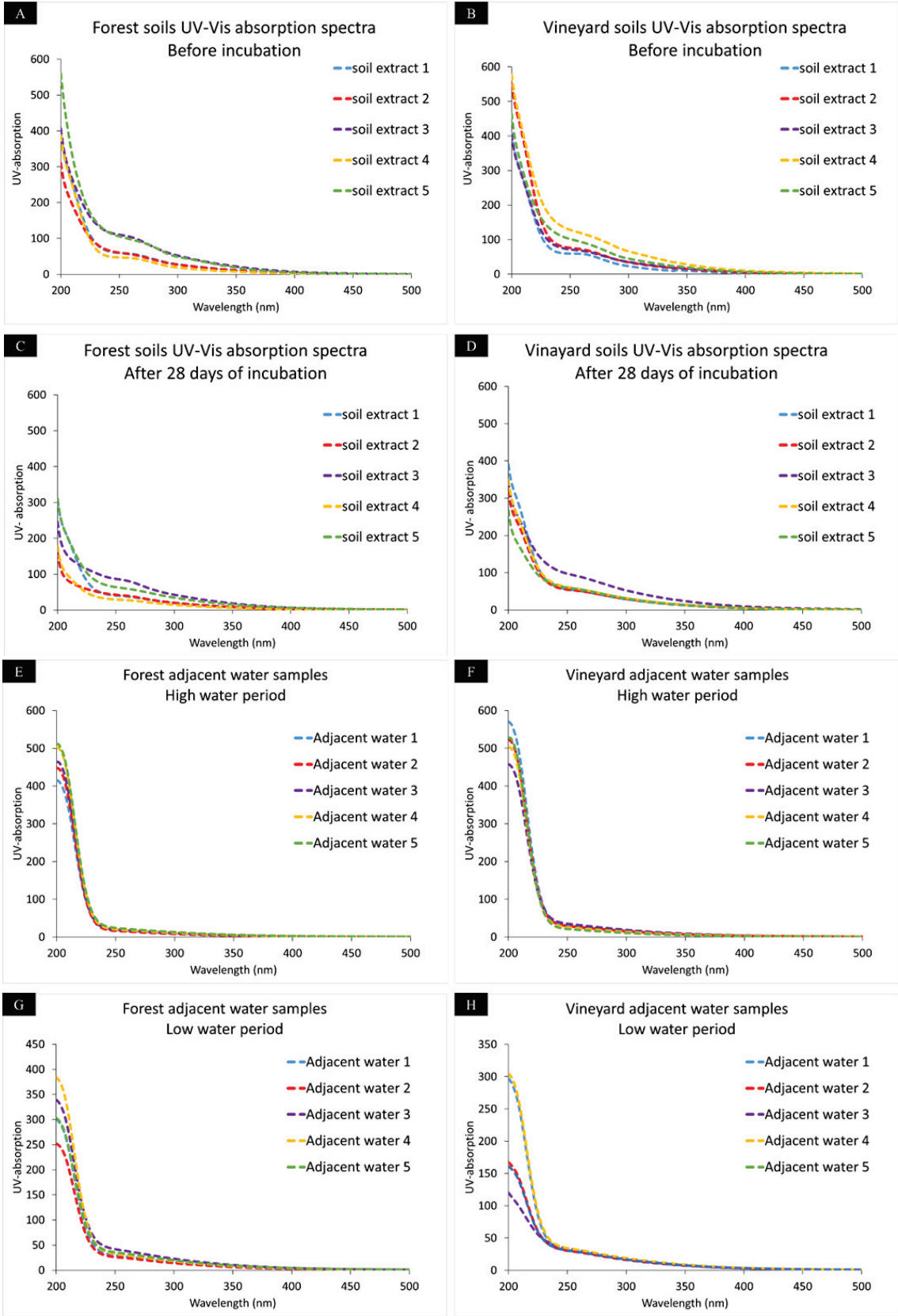
Supplementary material 1: detailed geographic coordinate of soil sampling areas with associated particular organic carbon POC concentrations (Flash Dynamic Combustion) in each soil.

Soil	Geographic coordinates	POC = g/kg MS
Forest soil 1 (FP1)	46.182, 4.63098	28
Forest soil 2 (FP2)	46.19197, 4.54902	30
Forest soil 3 (FP3)	46.18642, 4.52236	38
Forest soil 4 (FP4)	46.16876, 4.52858	22
Forest soil 5 (FP5)	46.15804, 4.58192	56
Vineyard soil 1 (VP1)	46.13965, 4.67481	18
Vineyard soil 2 (VP2)	46.14673, 4.65093	36
Vineyard soil 3 (VP3)	46.14719, 4.64498	51
Vineyard soil 4 (VP4)	46.13271, 4.6248	36
Vineyard soil 5 (VP5)	46.17415, 4.64304	35

Supplementary material 2: Summary diagram of forest and vineyard soil sampling, organic matter extraction (soil extracts) and degradation study over 28 days. Soil extract subsamples were made after 3, 7, 14 and 28 days of incubation.

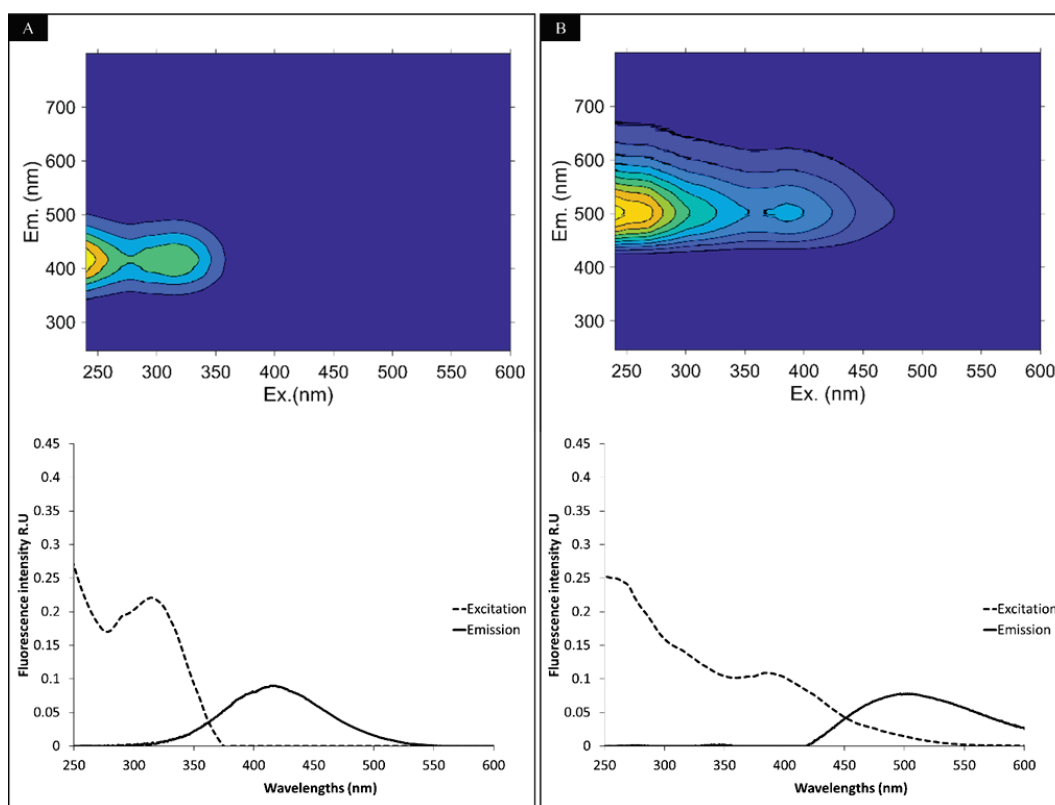


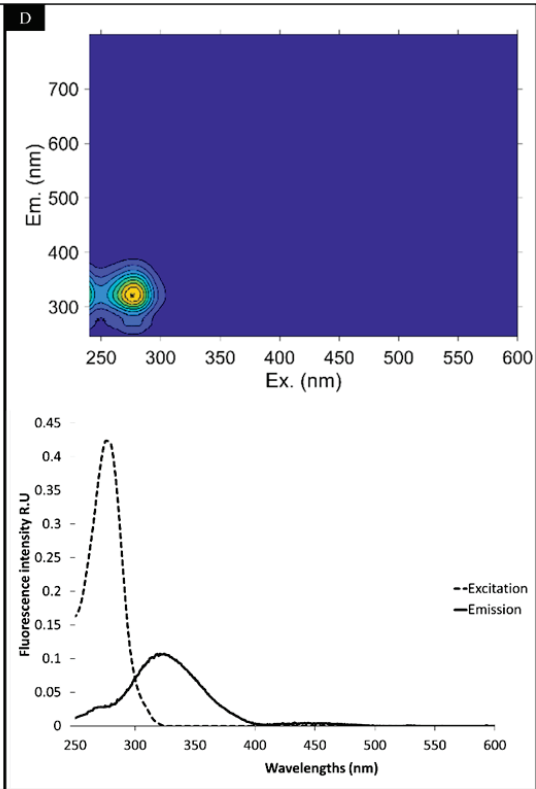
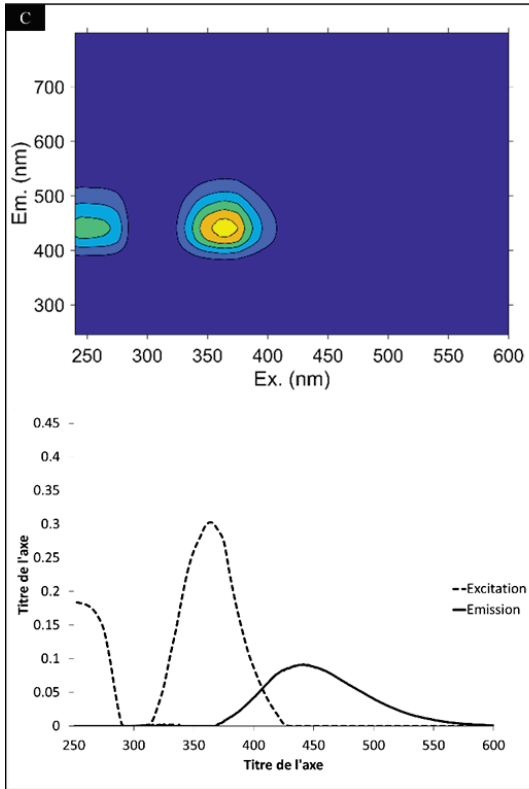
Supplementary material 3: UV-visible spectra for forest and vineyard soil extracts. A, B: forest and vineyard soil extracts before incubation, C, D: forest and vineyard soil extracts after 28



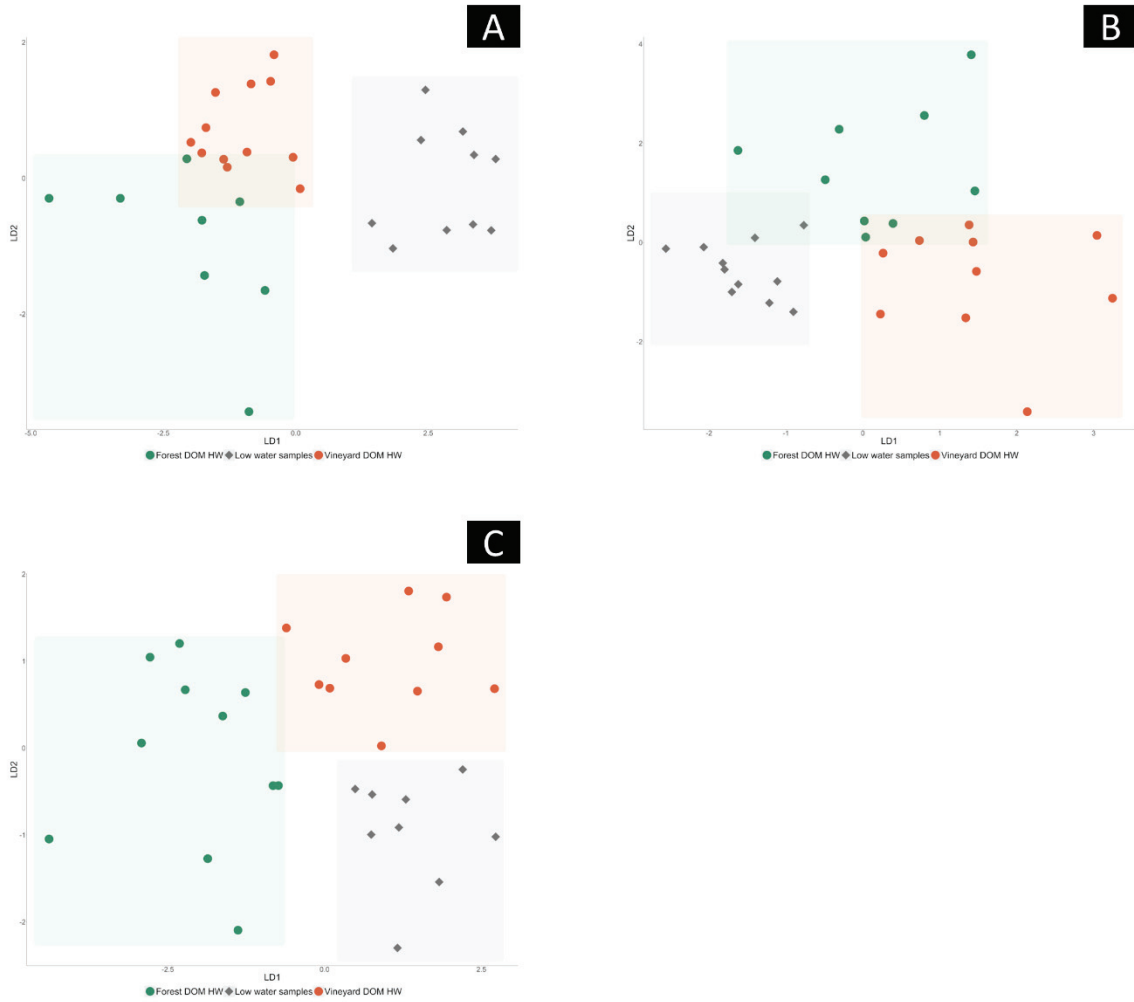
days of incubation, E, F: forest and vineyard adjacent water sampled during high water, G, H: forest and vineyard adjacent water sampled during low water.

Supplementary material 4: PARAFAC decomposition component matrices and excitation-emission spectra. A: component C1(240(320)-420 nm), B : component C2(240(390)-510 nm), C: component C3(365-443 nm), D : component C4 (280-330 nm)

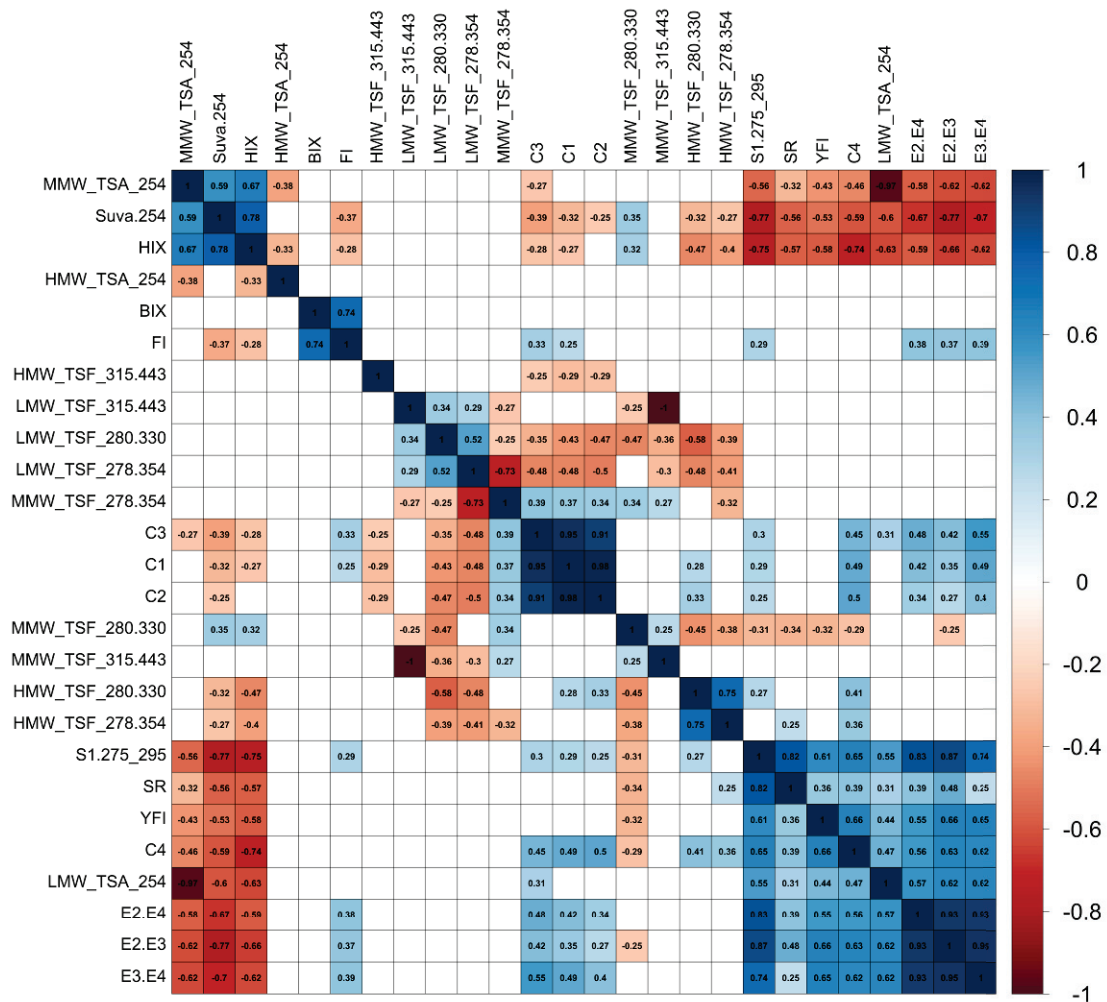




Supplementary material 5: linear discriminant analysis LDA 1 with A : UV-visible indicators only, B: EEM fluorescence and PRAFAC indicators only and C : HPSEC/UV-fluorescence indicators only.



Supplementary material 6: Correlation matrix of UV-Visible, fluorescence and HPSEC/UV-fluorescence indicators. White cells indicate a non-significant correlation based on a Pearson correlation test ($p > 0.05$, $n=70$)



5.2 Etude de l’empreinte globale des sources diffuses terrigènes à l’échelle de plusieurs bassins versants

La méthodologie d’échantillonnage des sources diffuses proposée dans la première partie a permis de mettre en évidence des empreintes de MOD spécifiques aux usages des sols, avec un apport terrigène en période de hautes eaux et un apport autochtone (non lié aux usages des sols) plus visible en période de basses eaux. Par exemple, l’étude réalisée sur le BV de l’Ardières-Morcille a montré l’impact d’un sol agricole par rapport à un sol forestier sur les caractéristiques de la MOD dans les cours d’eau. Néanmoins, cela pose la question de savoir si les empreintes obtenues pourraient être généralisables à une plus large échelle. En effet, en raison de la forte réactivité et complexité de la MOD, associées aux conditions environnementales et aux caractéristiques des bassins versants (topographie, géologie, hydrologie), l’établissement d’empreintes globales de l’usage des sols reste à ce jour un défi. L’obtention d’empreintes spécifiques de MOD pour les sources diffuses terrigènes demeure néanmoins essentielle dans les approches de suivi de la qualité des eaux naturelles et de l’impact des activités anthropiques sur l’environnement.

Dans ce contexte, nous avons élargi l’étude de l’empreinte des sources diffuses terrigènes aux bassins versants de l’Yzeron et du Gier (Rhône et Rhône/Loire). Le bassin versant de l’Yzeron présente un gradient amont-aval d’occupation très marqué, de sols forestiers vers des zones urbaines, en passant par l’agriculture et l’élevage. Il en est de même pour le bassin versant du Gier, avec une forte densité urbaine en aval (Chapitre 3, Parties 3.1.1.2 et 3.1.1.3). Les échantillons étudiés dans cette partie sont ceux prélevés en aval des occupations de sols forestières/prairies, et des usages de sols agricoles et retenues d’eau dans les trois bassins versants de l’Ardières-Morcille, de l’Yzeron et du Gier. Nous avons ajouté les retenues d’eau, car leur présence est très importante au niveau des trois bassins versants, et des études précédentes ont montré que ces retenues pouvaient avoir un impact considérable sur la qualité de la MOD dans les cours d’eau (McEnroe et al., 2013; Zhao et al., 2022; Liu et al., 2023). Tous les points ont été prélevés durant une période de hautes eaux et une période de basses eaux pour déterminer l’apport de la production autochtone. Le descriptif des échantillons est rapporté dans le Tableau 3-1 du Chapitre 3. La comparaison des empreintes de MOD a été réalisée à l’aide d’analyses multivariées en utilisant les données issues des

techniques optiques et moléculaires et en prenant en compte de nouveaux indicateurs présentés dans le Chapitre 4. Avant l'interprétation des résultats, la Partie 5.2.1 rassemble et permet de présenter les résultats de DOC et la description globale des indicateurs classiques UV-Visible et EEM de fluorescence. Par la suite, l'ensemble des indicateurs a été intégré à des analyses multivariées de type analyses en composantes principales (PCA) et classification hiérarchique ascendante (CAH) (Parties 5.2.2, 5.2.3). Les valeurs médianes, maximales et minimales de tous les indicateurs (n = 47) sont rapportées en Annexes 5-1, 5-2 et 5-3.

5.2.1 Teneur en DOC et description globale des indicateurs optiques classiques dans différents types de sources diffuses de MOD

Pour les sources représentatives des occupations de sols forêts/prairies, les concentrations médianes de DOC étaient respectivement de 2.7, 6.7 et 2.2 mg. L⁻¹ pour le BV de l'Ardières-Morcille, de l'Yzeron et du Gier en période de hautes eaux et de 3.6, 5.2 et 3.3 mg. L⁻¹ en période de basses eaux. Les concentrations médianes de DOC pour les sources agricoles étaient respectivement 3.8, 9.6 et 6.4 mg. L⁻¹ en période de hautes eaux et de 4, 7.3 et 4 mg. L⁻¹ en période de basses eaux. Enfin, pour les sources représentatives des retenues d'eau, les concentrations de DOC étaient respectivement de 5.2, 10.8 et 8.3 mg. L⁻¹ pour les BV de l'Ardières-Morcille, de l'Yzeron et du Gier en période de hautes eaux et de 6.9, 6.6 et 3.8 mg. L⁻¹ en période de basses eaux. La gamme des concentrations de DOC observée dans ces échantillons est comparable à celle observée dans les études sur les cours d'eau provenant de sous-bassins versants présentant des usages des sols mixtes (Williams et al., 2010; Smith et al., 2021).

L'augmentation des concentrations de carbone observée entre les périodes de basses et de hautes eaux peut s'expliquer par une plus grande contribution des sources terrigènes de MOD pendant les périodes de hautes eaux. En effet, plusieurs études ont déjà rapporté une augmentation du carbone pendant les périodes humides en raison d'apports terrigènes plus importants dus au lessivage des sols adjacents (Wilson et al., 2013).

La variabilité de DOC observée dans les retenues d'eau des trois BV pourrait quant à elle être due au fait que ces retenues représentent des systèmes assez clos, alimentés par de faibles débits d'eau. En effet, cette configuration peut entraîner un apport très variable en DOC selon

les sources avoisinant les retenues. De plus, le temps de résidence au niveau de ces retenues peut induire différents processus de dégradation et transformation qui peuvent être différents d'un milieu à un autre.

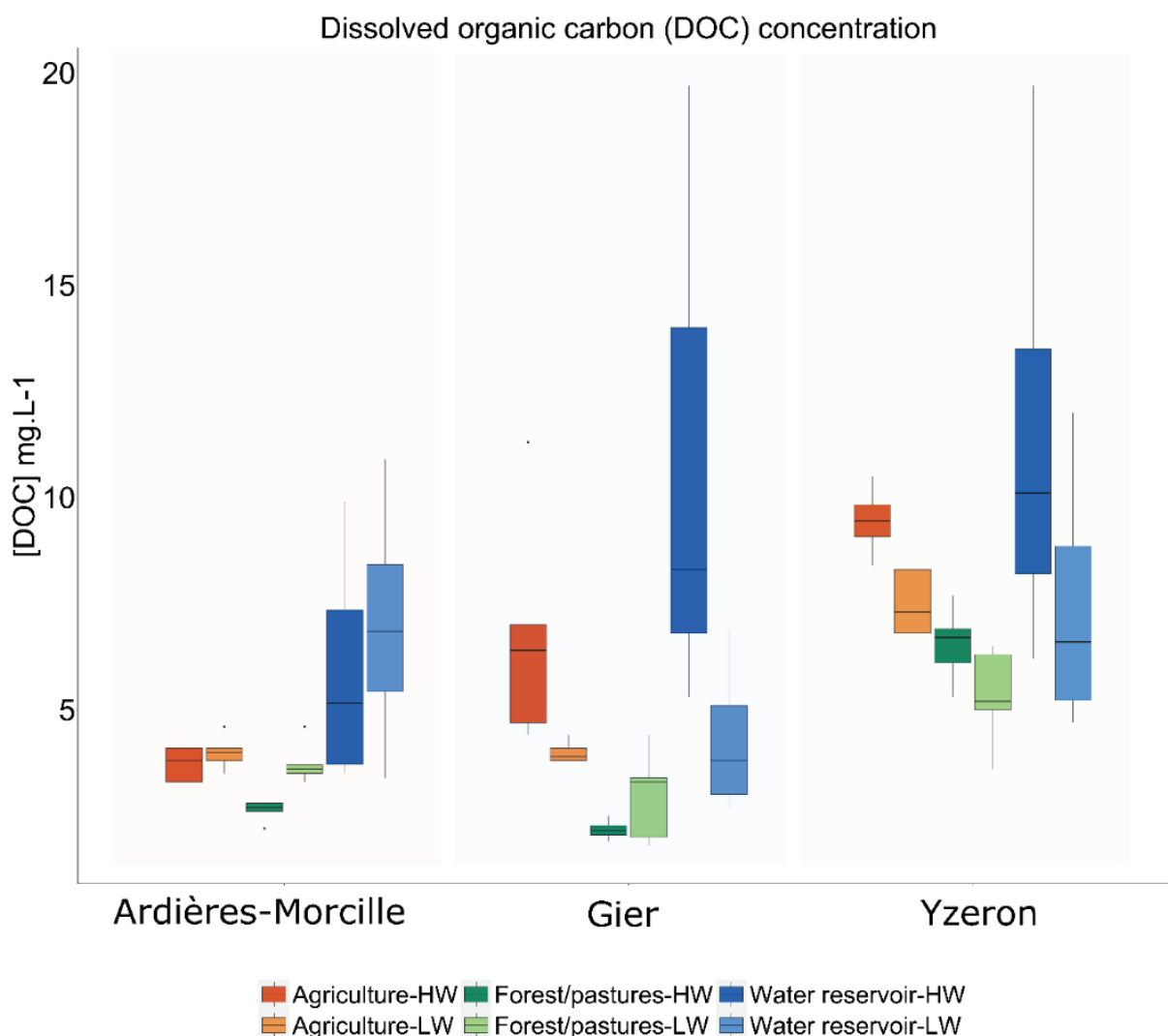


Figure 5-10. Boxplots des concentrations de carbone organique dissous (DOC) en mg.L⁻¹ dans les échantillons représentatifs des sources diffuses

Les indicateurs UV-Visible E2.E4 et E3.E4 (Figure 5-11) mettent en évidence une MOD de nature terrigène humifiée avec une hétérogénéité en termes de poids moléculaire (E2.E3) pour les échantillons prélevés en hautes eaux, en aval des sources forestières et agricoles. La MOD prélevée au niveau des retenues d'eau est quant à elle principalement caractérisée par la présence de petites molécules peu humifiées d'origine autochtone. Cette caractéristique s'applique également aux échantillons prélevés en période de basses eaux et dont les empreintes spectrales semblent relativement proches, indépendamment du type de source. Le degré d'aromaticité est relativement plus élevé dans les échantillons prélevés en période

de basses eaux en comparaison avec les hautes eaux. Concernant les différences entre BV, les échantillons représentatifs des sources forestières et agricoles prélevés dans le Gier semblent être relativement différents de ceux prélevés dans les BV de l'Ardières et de l'Yzeron. Cette différence, principalement visible avec les indicateurs E2.E4 et E3.E4, indique que la MOD forestière et agricole du Gier présente un degré d'humification plus faible que celle de l'Ardières et de l'Yzeron. La MOD du Gier, quant à elle, est principalement issue de sources terrigènes avec très peu de contribution autochtone, en comparaison avec l'Ardières et l'Yzeron. Les valeurs des indicateurs UV-Visible observées dans cette étude sont similaires à celles rapportées dans de précédentes études ayant travaillé sur des matrices similaires (Chen et al., 2021; Liu et al., 2023).

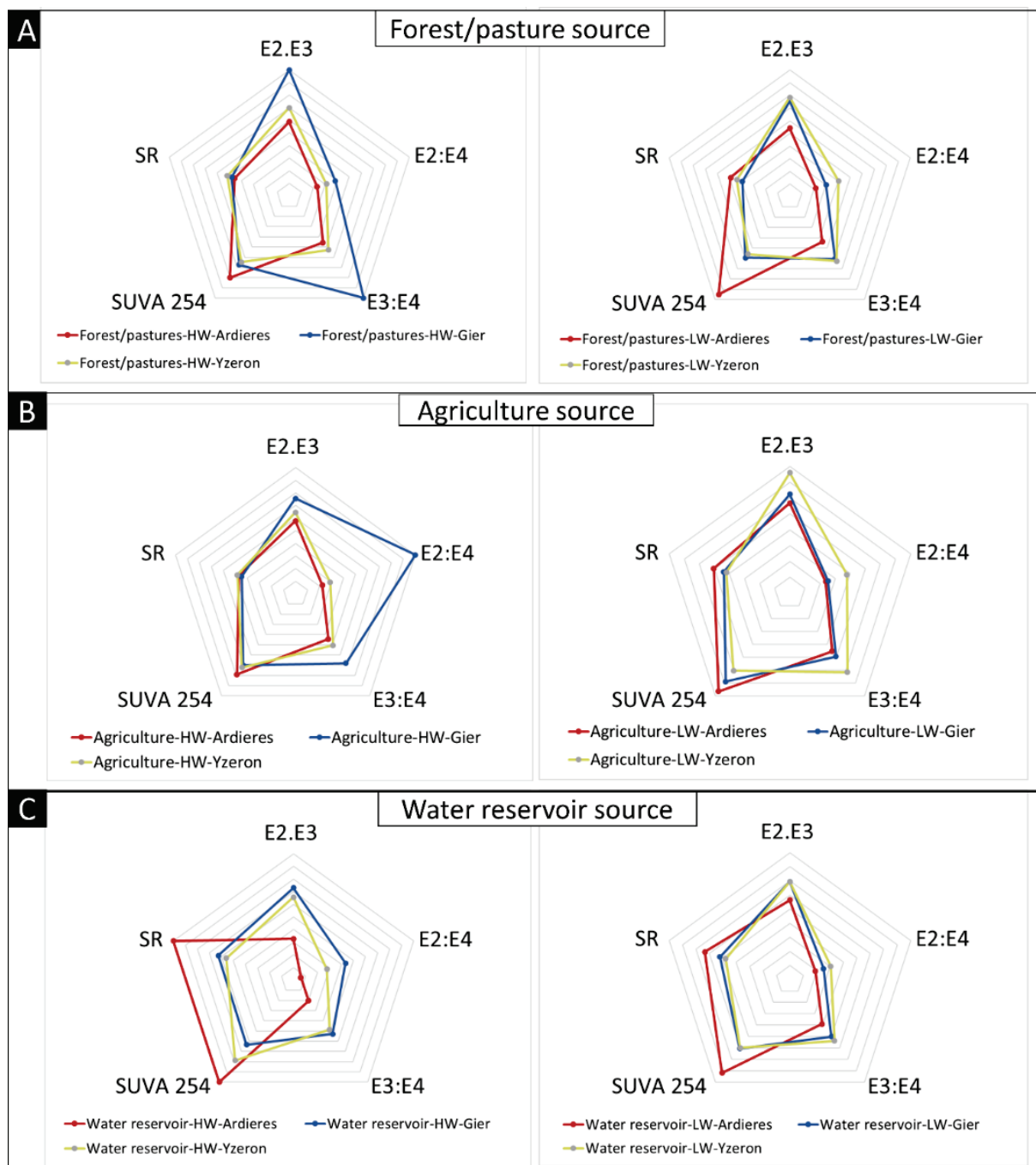


Figure 5-11. Valeurs médianes des indicateurs UV-Visible dans les échantillons de sources A. agricoles en hautes eaux (HW) et basses eaux (LW), B. forestier/prairies en HW et LW et C. retenues d'eau en HW et LW. **E2.E3** : inversement corrélé à la taille des molécules, **E2.E4** : inversement corrélé à la production autochtone, **E3.E4** : inversement corrélé à l'humification, **SUVA** : aromaticité de la DOM ($L \cdot mg^{-1} \cdot cm^{-1}$), **SR** : inversement corrélé au poids moléculaire.

En ce qui concerne la spectroscopie EEM de fluorescence, la décomposition PARAFAC a permis de valider cinq composantes (Figure 5-12) qui ont été comparées aux résultats issus de la littérature en utilisant la base de données *OpenFluor* avec des scores de similarité supérieurs à 97 %.

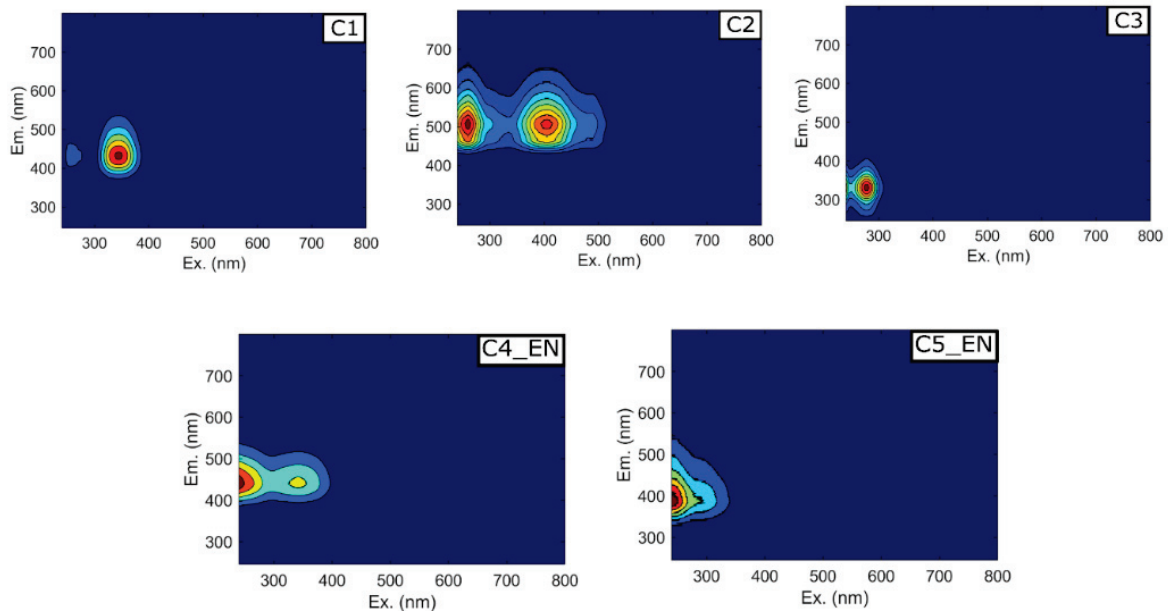


Figure 5-12. Composantes PARAFAC validées dans les échantillons de sources diffuses. C1 (Ex240-350 Em440), C2 (Ex260-Em505), C3 (Ex280-Em330), C4 (Ex325-Em395 nm), C5_EN (Ex<240-Em390)

La composante C1 (Ex240-350 Em440) a été décrite comme un composé d'origine terrestre dégradé à partir de la lignine (Peleato et al., 2018), typique des sols et des sources terrestres dans les environnements aquatiques (Garcia et al., 2018; Yamashita et al., 2021). Cette composante a également été attribuée à la matière organique biodisponible, d'origine microbienne et/ou sensible à la photodégradation (Cawley et al., 2012; Eder et al., 2022). La composante C2 (Ex260-Em505) a été rapportée dans des études antérieures comme étant caractéristique des composés de type humique dégradés par voie photochimique (Coulson et al., 2022). La composante C3 (Ex280-Em330) a quant à elle été associée à des composés de type protéique (Wünsch and Murphy, 2021). La composante C4_EN (Ex325-Em395 nm) a été précédemment rapportée comme étant typique des composés de type humique obtenus par activité biologique (Yamashita et al., 2010), tandis que d'autres études ont décrit la composante C4_EN comme étant associée aux composés transformés par les microorganismes (Du et al., 2021b). Enfin, la composante C5_EN (Ex<240-Em390) a été

rapportée comme étant caractéristique des composés d'origine terrestre et végétale ayant un poids moléculaire élevé et une forte teneur aromatique (D'Andrilli et al., 2017), tandis que d'autres études ont relié cette composante à un hydrocarbure aromatique polycyclique présent dans les stations d'épuration industrielles ou municipales (Carstea et al., 2020).

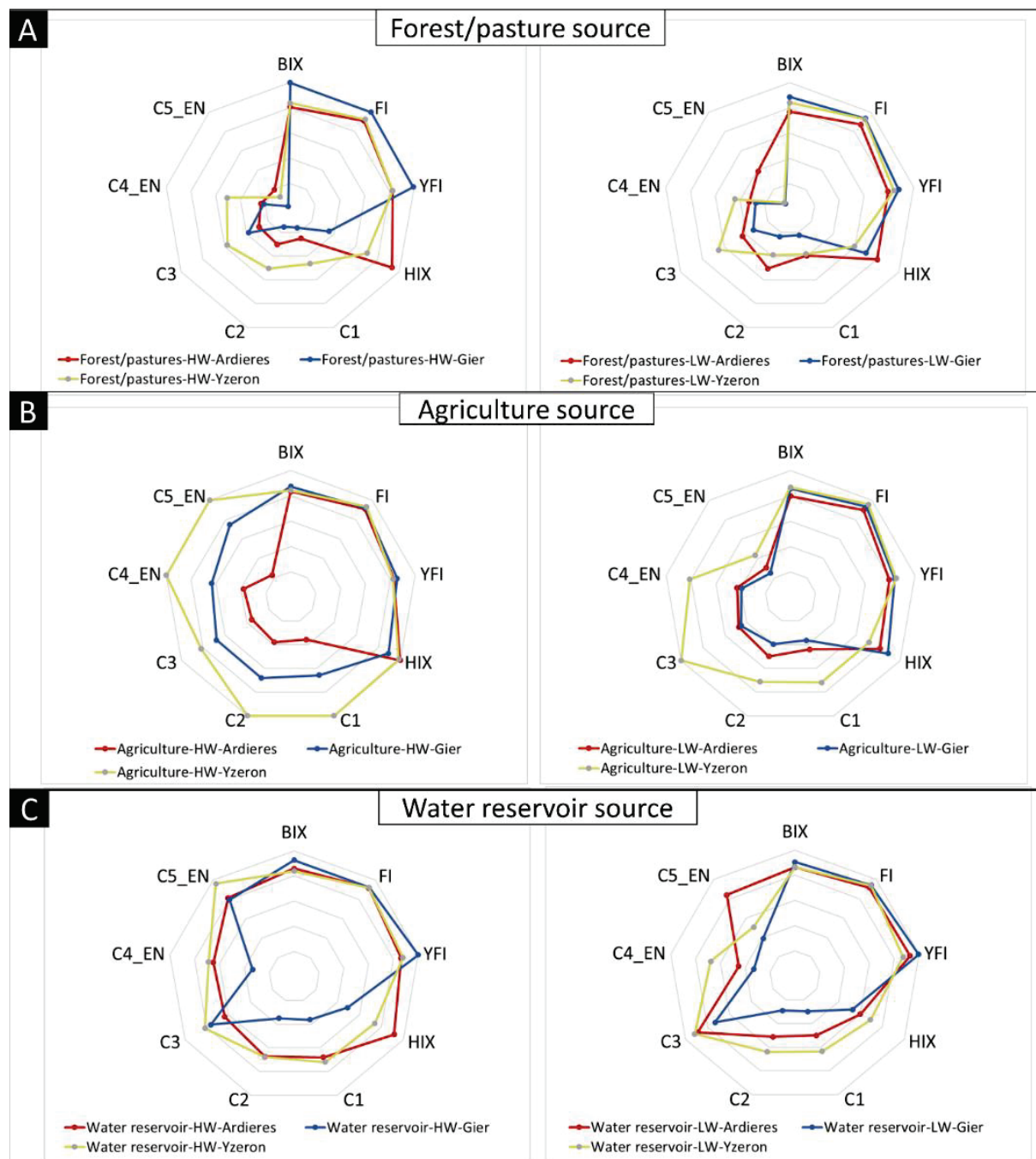


Figure 5-13. Valeurs médianes des indicateurs de fluorescence dans les échantillons de sources A. agricoles en hautes eaux (HW) et basses eaux (LW), B. forestiers/prairies, fractions HW et LW et C. retenues d'eau, fractions HW et LW. **BIX** : corrélé à l'activité microbienne, **FI** et **YFI** : indicateurs d'apports terrigènes, **HIX** : indicateur d'humification, **C1** : composé terrigène dégradé (R.U), **C2** : composé type humique, **C3** : composé type protéique, **C4_EN** : composé type humique issue de l'activité microbienne, **C5_EN** : composé terrigène aromatique.

Les composantes PARAFAC et les indicateurs d'EEM de fluorescence (Figure 5-13) montrent la présence de différences entre les empreintes de fluorescence des MOD en fonction des types de sources et des conditions hydrologiques.

Les échantillons prélevés en période de hautes eaux, en aval des sources forestières, présentent des contributions plus faibles pour l'ensemble des composantes PARAFAC (C1, C2, C3, C4_EN et C5_EN), en comparaison avec ceux en aval des sources agricoles et dans les retenues d'eau. Le degré d'humification est légèrement plus faible dans les retenues et s'accompagne d'une forte contribution en composés de type protéique (C3). L'empreinte des MOD en période de basses eaux tend vers une homogénéisation, avec une contribution relativement élevée des composés de type protéique (C3). La composante C5 (composé d'origine terrestre ou végétale) est fortement présente dans les eaux agricoles et les retenues d'eau en période de hautes eaux, mais diminue durant la période de basses eaux. Les valeurs de HIX, BIX et FI obtenues dans cette étude sont dans les mêmes gamme de valeurs que celles observées dans de précédentes études (Lambert et al., 2017; Coble et al., 2022).

5.2.2 Caractérisation intégrée de la MOD issue des sources diffuses : identification d'un effet bassin versant

A une échelle d'étude plus large, l'intégration de l'ensemble des indicateurs dans une analyse en composantes principales (PCA) et une classification ascendante hiérarchique (CAH) (Figure 5-14) a permis d'identifier deux groupes distincts d'échantillons de MOD : un premier groupe (CAH 1), avec une majorité d'échantillons prélevés dans les BV de l'Yzeron et du Gier, et un deuxième groupe (CAH 2) qui contient principalement les échantillons du BV de l'Ardières-Morcille.

Contrairement aux échantillons prélevés dans l'Yzeron et le Gier, la MOD de l'Ardières-Morcille est principalement caractérisée par une MOD de poids moléculaire plus élevé avec des degrés d'humification et d'aromaticité plus élevés. Les indicateurs LC-HRMS mettent également en évidence la présence de composés principalement apolaires (élution en fin d'acquisition) répartis sur une large gamme de masses (<500 Da et >1000 Da). La MOD prélevée dans les BV de l'Yzeron et du Gier est caractérisée par contribution plus élevée en petites molécules de type acides aminés (< 1 kDa) aux longueurs d'onde d'excitation-émission 275-354 nm. Les indicateurs FI, BIX et YFI montrent que la MOD prélevée dans ces deux BV est

produite récemment, avec une activité microbienne élevée en comparaison avec celle prélevée dans le BV de l'Ardières-Morcille. Les indicateurs UV-Visible E2.E3 et E3.E4, qui sont inversement corrélés à la taille des molécules et au degré d'humification, confirment ces résultats. Les indicateurs LC-HRMS RT0.6-MW0.500 et RT12.18-MW0.500 montrent la présence d'une MOD principalement constituée de composés dont la masse est inférieure à 500 Da, avec une gamme large de polarité. A l'échelle des bassins versants, les résultats révèlent bien la présence d'un effet « bassin versant » sur la qualité de la MOD. Celui-ci pourrait provenir des paramètres suivants :

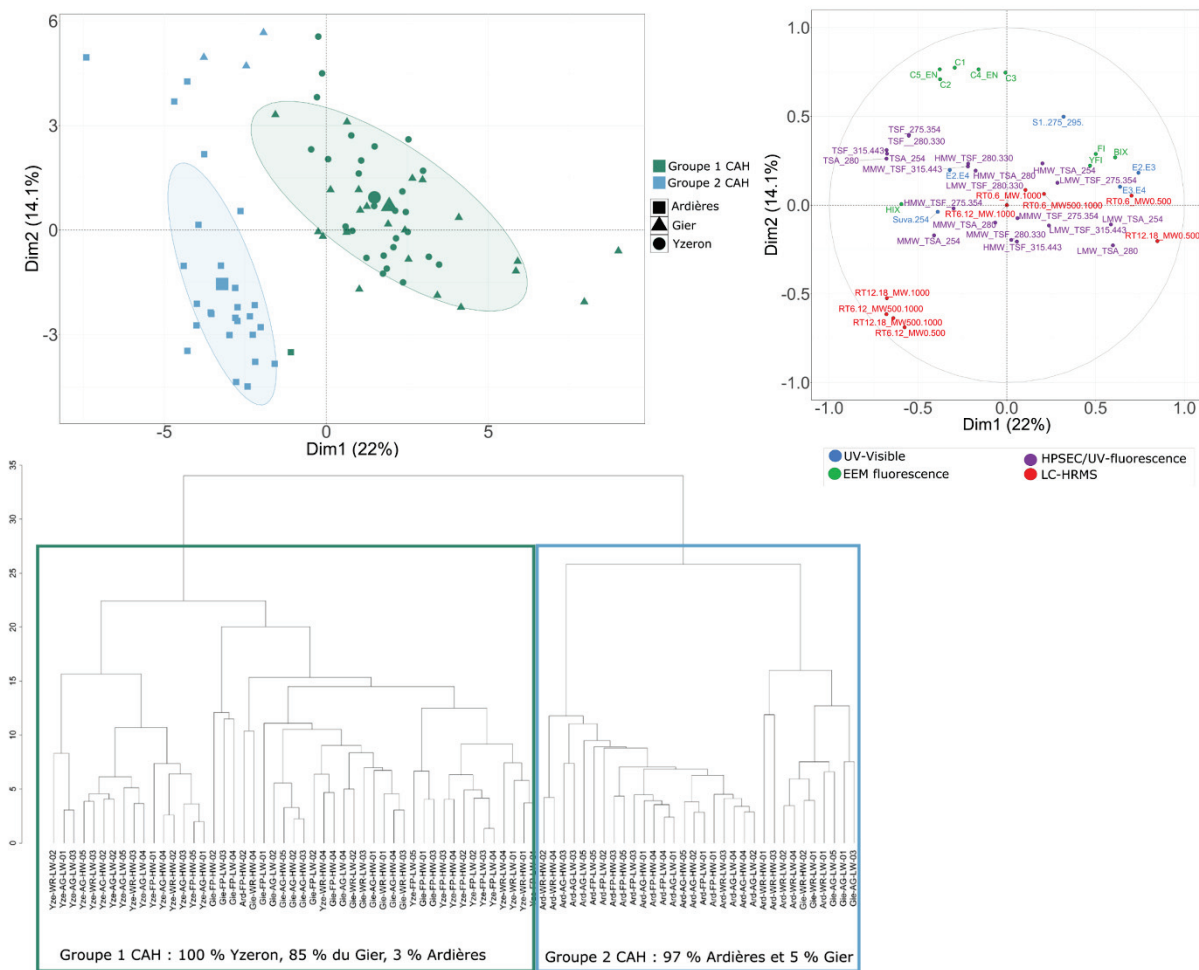


Figure 5-14. Analyse en composantes principales des échantillons de DOM représentatifs des sources diffuses prélevés dans les bassins versants de l'Ardières-Morcille, de l'Yzeron et du Gier-Grand Malval.

- Les usages des sols : bien que nous ayons considéré les mêmes types de sources, il existe des différences d'usage des sols entre les 3 BV, notamment dans les pratiques agricoles, à savoir la présence d'élevage et de maraichage pour l'Yzeron et le Gier et une forte pratique viticole dans le BV de l'Ardières-Morcille. Cela peut avoir un impact sur la qualité de la MOD dans les

cours d'eau adjacents. Par exemple, un ruissellement d'eau dans des zones d'élevages peut entraîner un apport de déchets issus d'animaux et d'excréments contenant des agents pathogènes et d'autres contaminants organiques (e.g. produits pharmaceutiques) qui altèrent la composition de la MOD (Qi et al., 2023). Les pratiques viticoles impliquent l'utilisation de pesticides et d'engrais qui peuvent également modifier la composition de la MOD, entraînant des changements de types de composés présents dans les échantillons (Peyrard et al., 2016). Certaines propriétés physiques du sol, qui sont susceptibles d'être modifiées par les pratiques agricoles (par exemple en raison du piétinement du bétail ou le tassement des tracteurs qui circulent dans les vignes), peuvent également impacter la structuration des communautés microbiennes, et donc les processus de biodégradation de la MOD (Escribano-Viana et al., 2018; Hu et al., 2022).

- Les conditions hydrologiques : les conditions hydrologiques des cours d'eau, telles que le débit, le temps de séjour de l'eau et les dynamiques de mélanges, peuvent également modifier la qualité de la MOD (Shang et al., 2018; Coble et al., 2022). Dans cette étude, nous avons tenté de limiter au mieux l'influence de ces paramètres en allant prélever au plus près des sources en tête de bassin (afin de s'assurer d'un temps de résidence dans le cours d'eau très faible).

- La géologie et la pédologie : la géologie des bassins versants peut également influencer la qualité de la MOD (Mosher et al., 2010). Par exemple, si un bassin versant présente des sols plus riches en minéraux que les autres, cela peut conduire à des apports plus importants de minéraux, ce qui peut affecter l'activité microbienne et les processus de dégradation qui influencent la qualité de la MOD (Carson et al., 2007). La pédologie des sols peut également affecter la vitesse de transfert des eaux dans le sol et, de ce fait, la dégradation plus ou moins poussée de la MOD transférée vers les cours d'eau (Sanderman et al., 2009).

5.2.3 Sur quel type de source diffuse l'effet bassin versant est-il visible ?

Afin de mieux évaluer la différence d'empreintes des sources diffuses entre plusieurs bassins versants, chaque type d'occupation de sol a été étudié séparément. Le but était de voir sur quel type d'occupation de sol l'effet bassin versant était le plus visible. L'analyse ACP (Figure 5-15) montre la répartition des échantillons de MOD prélevés en aval des sous-bassins versants caractéristiques des usages de sols agricoles dans les trois BV (en période de hautes

eaux et basses eaux), en fonction de l'ensemble des paramètres optiques et moléculaires. Le premier plan de l'ACP (composantes principales 1 et 2), qui explique 41 % de variance totale, met en évidence des empreintes distinctes entre les bassins versants à la fois en période de hautes eaux et en période de basses eaux. Les empreintes de sources agricoles sont relativement proches pour les BV de l'Yzeron et du Gier, au contraire de celles de l'Ardières-Morcille.

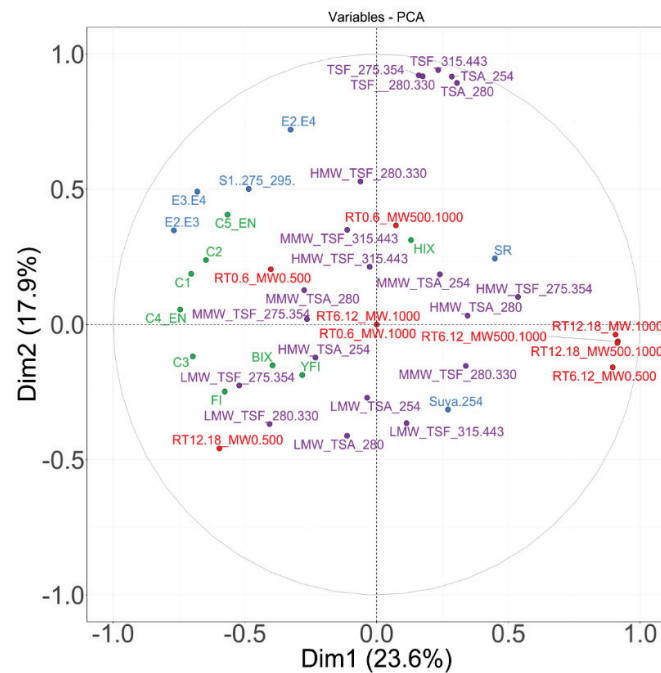
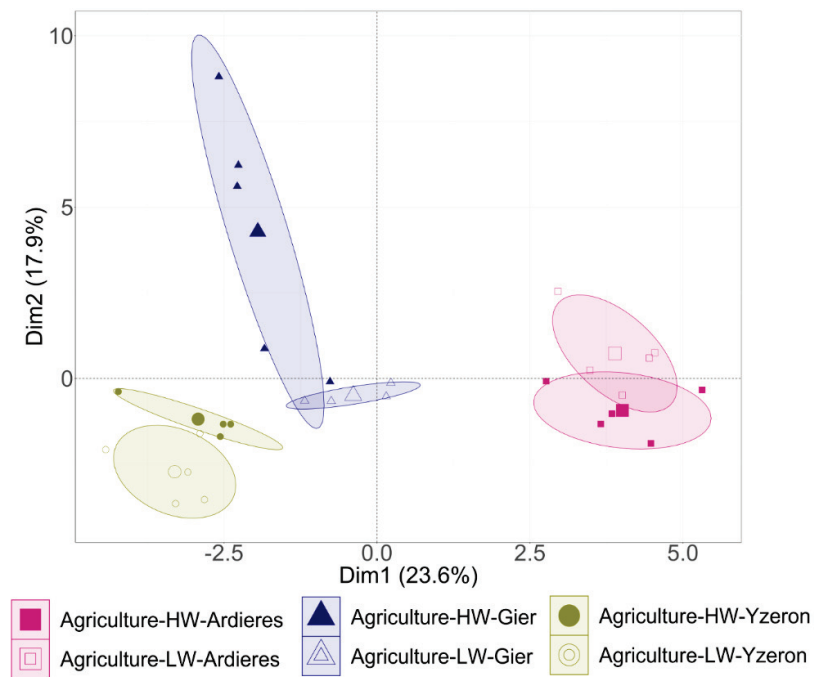


Figure 5-15. Analyse en composantes principales des échantillons de DOM représentatifs des sources diffuses agricoles, prélevés dans les bassins versants de l'Ardières-Morcille, de l'Yzeron et du Gier-Grand Malval en périodes de hautes eaux (HW) et de basses eaux (LW)

La MOD prélevée au niveau de l'Yzeron et du Gier est principalement caractérisée par la présence de molécules de petites tailles faiblement aromatiques, avec une contribution relativement élevée des composantes PARAFAC caractéristiques d'un apport terrigène et également de composés de type protéique. La MOD agricole prélevée dans l'Yzeron se distingue de celle prélevée dans le Gier par la présence de composés apolaires de faibles masses moléculaires (<500 Da) dans les échantillons de l'Yzeron. De plus, les indicateurs HPSEC/UV-fluorescence LMW275-354 et LMW280-330 indiquent une contribution plus élevée en composés de type protéine (< 1 k Da) dans les échantillons de l'Yzeron (Figure 5-16).

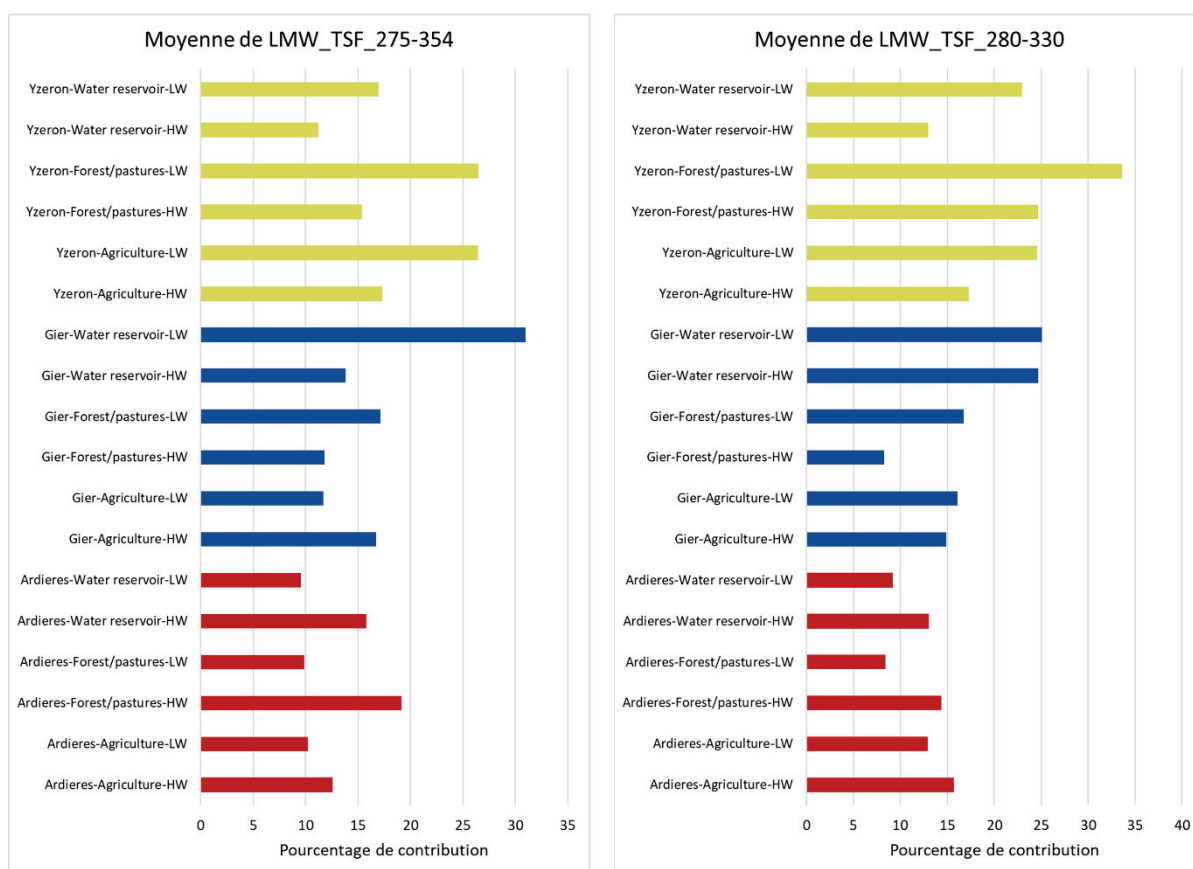


Figure 5-16. Pourcentage de contribution des petites molécules (< 1 k Da) à la fluorescence totale aux longueurs d'onde d'excitation-émission 275-354 nm (à gauche) et 280-330 nm (à droite), dans les sources des 3 bassins versants (jaune : Yzeron, bleu : Gier, rouge : Ardieres-Morcille).

En parallèle, les échantillons prélevés dans le Gier présentent une contribution élevée en composés de poids moléculaires élevés (> 10 kDa) de types humique et protéique. Au niveau de l'Ardières-Morcille, la MOD prélevée en aval des sols à usage viticole semble être principalement constituée de molécules apolaires de masses hétérogènes, ayant des degrés d'aromaticité et d'humification élevés. En outre, des différences entre les empreintes

agricoles des trois BV sont également observées en période de basses eaux. Cela montre que l'apport terrigène est également visible en période de basses eaux. Il est possible que la MOD produite pendant les périodes d'étiage soit influencée non seulement par l'activité microbienne locale et celle des producteurs primaires, mais aussi par un effet cumulatif, qui peut être tamponné et différé dans le temps, des sources et des affluents dans les sous-bassins en amont (signature autochtone et terrigène). A notre connaissance, aucune étude n'a porté sur la qualité de la MOD en période d'étiage sur différents bassins versants. Le plus souvent, les comparaisons sont en effet faites entre la MOD autochtone et la MOD allochtone (Murphy et al., 2008; Toming et al., 2013; Lee et al., 2020).

Le même constat peut être dressé pour les sources prélevées dans des sous-bassins versants forestiers (Figure 5-17). Le premier plan de l'ACP, expliqué par 50.1 % de variance totale, montre également la présence de trois empreintes de sols forestiers, bien distinctes entre les bassins versants. La dimension 1 de l'ACP (34%) permet de distinguer les empreintes de forêts du BV de l'Ardières-Morcille de celles des deux autres BV ; et la dimension 2 permet de différencier l'Yzeron des deux autres BV. Contrairement aux sources agricoles, le couvert végétal se ressemble en niveau des trois BV, à savoir, principalement des forêts de feuillus et de conifères avec la présence de prairies. La différence observée entre les empreintes de MOD de sources forestières dans les trois bassins versants, même avec une végétation similaire, peut être attribuée à une combinaison complexe de plusieurs facteurs interdépendants, comme le type de sol, le climat, l'hydrologie et l'activité et la diversité microbienne (Thieme et al., 2019).

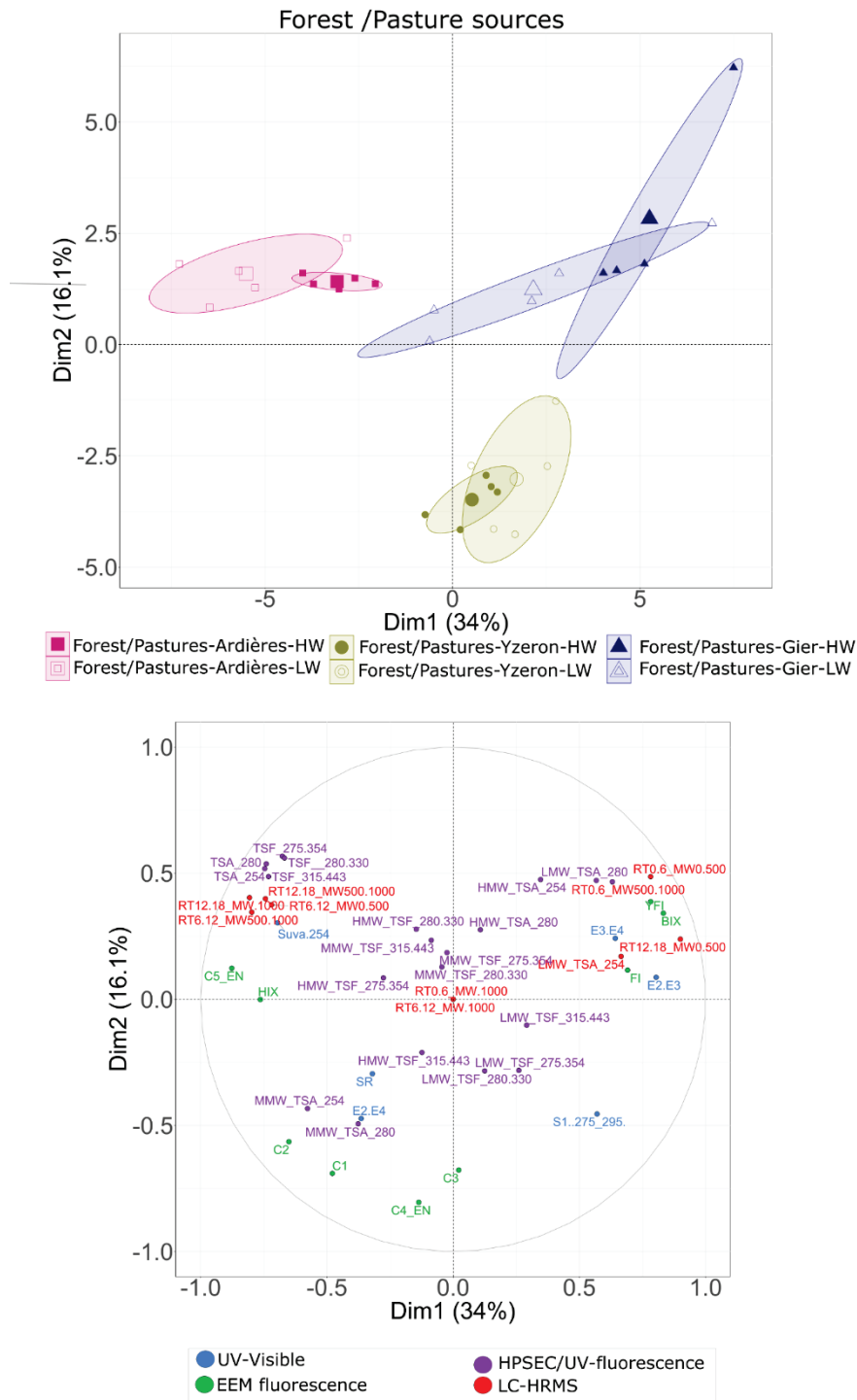


Figure 5-17. Analyse en composantes principales des échantillons de DOM représentatifs des sources diffuses forestières/pâturages, prélevés dans les bassins versants de l'Ardières-Morville, de l'Yzeron et du Gier-Grand Malval en périodes de hautes eaux (HW) et de basses eaux (LW)

Le dernier type de source considéré dans cette étude concerne les retenues d'eau. L'analyse ACP (Figure 5-18) montre la répartition des échantillons représentatifs de ces sources dans les trois BV en période de hautes et basses eaux. Le résultat permet de mettre en évidence une

empreinte de MOD relativement homogène pour les bassins de l'Yzeron et du Gier et ce, indépendamment des conditions hydrologiques lors du prélèvement.

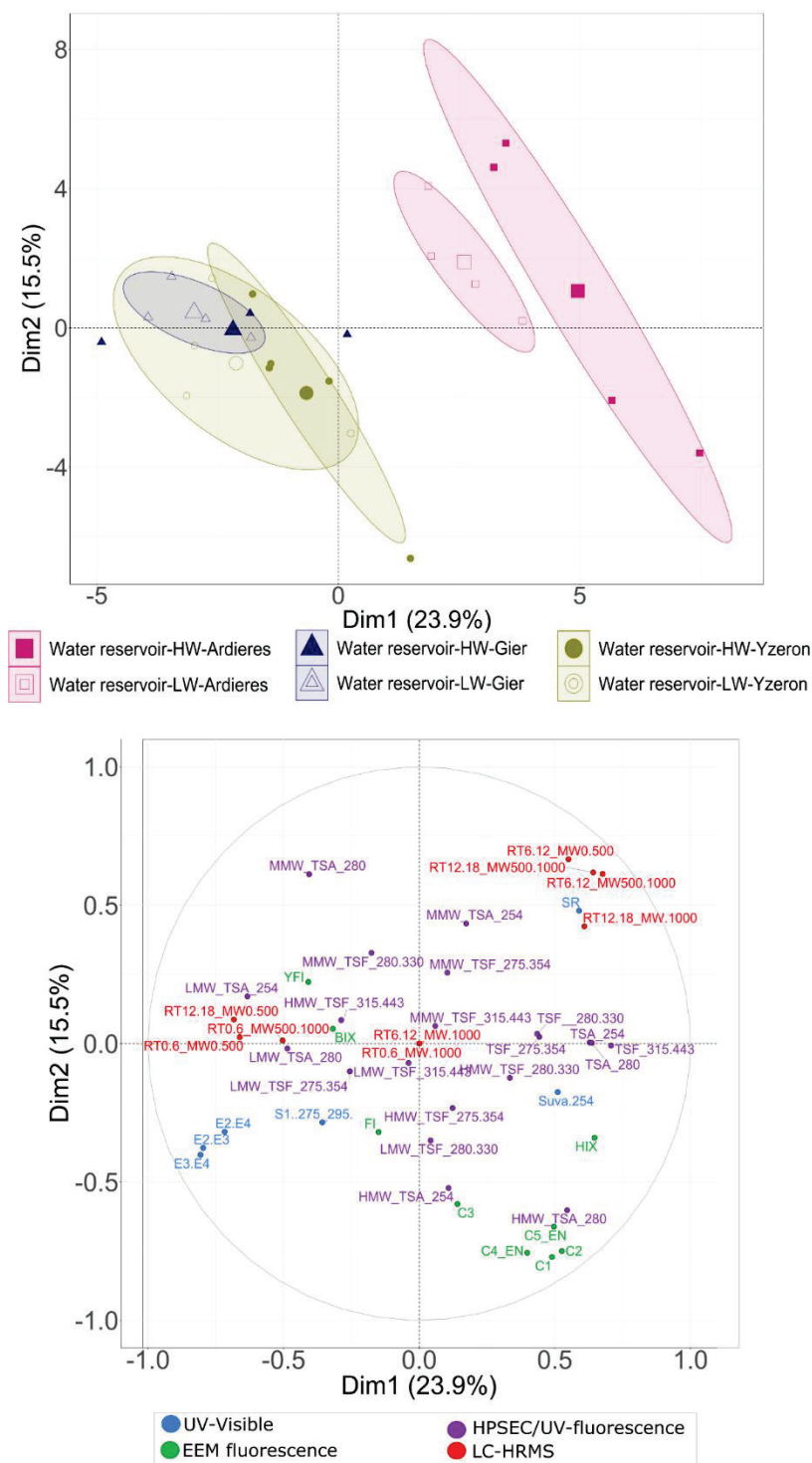


Figure 5-18. Analyse en composantes principales des échantillons de DOM représentatifs des retenues d'eau, prélevés dans les bassins versants de l'Ardières-Morcille, de l'Yzeron et du Gier-Grand Malval en périodes de hautes eaux (HW) et de basses eaux (LW)

Cette empreinte est principalement caractérisée par la présence d'un large panel de composés variables en masses et en polarités, avec une contribution relativement importante de la production autochtone mise en évidence par les indicateurs de fluorescence FI et BIX. En ce qui concerne le BV de l'Ardières-Morcille, les empreintes des retenues d'eau sont hétérogènes entre elles en période de hautes eaux, et plus homogène en période de basses eaux. Ces empreintes sont caractérisées par une forte contribution en composés apolaires et aromatiques.

Si l'empreinte de la MOD dans les retenues d'eau est pratiquement similaire dans deux des trois bassins versants de l'étude, cela pourrait être dû à un mélange de MOD provenant de différentes sources, y compris les forêts, l'agriculture et les zones urbaines. En effet, les retenues d'eau correspondent à des petites masses d'eau isolées qui ont généralement un débit d'entrée et de sortie limité. Ces masses d'eau sont souvent peu profondes et ont un temps de séjour relativement élevé, ce qui peut conduire à l'accumulation de MOD provenant de plusieurs sous-bassins versants environnants. Le temps de séjour de la MOD dans ces compartiments est alors susceptible d'avoir un impact sur sa qualité. L'allongement de ce temps de séjour renforce la production autochtone et la dégradation / transformation des mélanges complexes de composés constituant la MOD et conduit à une homogénéisation de son empreinte, homogénéisation observée même à l'échelle de plusieurs bassins versants (Du et al., 2021a).

Pour résumer, l'effet de bassin versant est observé à la fois au niveau des sources forêts/prairies et des sources agricoles, ce qui peut être attribué à divers facteurs tels que le type de sol, la géologie, ou encore l'activité et la diversité des microorganismes entre les bassins. Les différences de pratiques agricoles telles que les cultures, l'élevage ou la viticulture peuvent également expliquer les différences observées dans ces sources. Toutefois, l'impact de l'effet de bassin versant semble être moins visible pour les sources de retenues d'eau, sans doute en raison des différents mécanismes de dégradation et de production autochtone au niveau de ces milieux relativement stagnants.

5.2.4 Est-il possible de construire des empreintes d'occupations de sols indépendamment du bassin versant ?

L'effet bassin versant est prédominant sur l'empreinte des sources diffuses terrigènes. Cependant, il est encore possible de mettre en avant les similitudes (si elles existent) dans les compositions et propriétés physico-chimiques de la MOD des trois BV, et ce pour chaque type d'occupation des sols. Pour cela, on utilise une analyse linéaire discriminante (LDA) pour sélectionner une nouvelle série d'indicateurs pour discriminer les sources en s'affranchissant de cet effet BV. L'analyse LDA est une technique statistique utilisée pour analyser les différences entre deux ou plusieurs groupes d'échantillons sur la base d'un ensemble de variables quantitatives. Elle permet d'identifier les variables qui discriminent le mieux les groupes définis *a priori* par l'opérateur. Ainsi, la première étape de la LDA consiste à définir les groupes qui doivent être comparés. Les groupes sélectionnés dans le cadre de notre étude sont les échantillons impactés par les usages de sols tout BV confondus, à savoir six groupes d'échantillons issus des sources forêts/prairies, agriculture et retenues d'eau prélevés en hautes eaux et basses eaux. Le récapitulatif des groupes est présenté en Annexe 4. La LDA permet ensuite de déterminer une combinaison de variables qui distingue le mieux ces groupes en produisant un ensemble de fonctions discriminantes, qui sont des combinaisons linéaires de variables. Le résultat de l'analyse (Figure 5-19) a permis de classer les échantillons de MOD dans les groupes sélectionnés *a priori* (Annexe 4) avec un pourcentage d'erreur de 27 %. La combinaison d'indicateurs qui a permis de classer les sources en s'affranchissant de l'effet BV sont : les indicateurs UV-Visible SR et pente spectrale $S_{275-295}$, les indicateurs EEM de fluorescence HIX, YFI, C5_EN, les indicateurs HPSEC/UV-fluorescence TSF_315-443, HMW_TSF_275-354, MMW_TSF_315-443 et l'indicateur LC-HRMS RT12.18_MW500.1000. Ce résultat montre qu'il est ainsi possible de construire une typologie de MOD issue des sources diffuses, même si la variabilité inter-bassins versant est très présente. Les échantillons prélevés en période de hautes eaux (HW) sont différenciés selon les usages de sols *a priori*. Un quatrième groupe d'échantillons, prélevés en période de basses eaux, est également visible sur la Figure 5-19. L'empreinte de ces échantillons se confond avec les échantillons forestiers prélevés en période de hautes eaux. En période de basses eaux, l'impact des usages de sols est moins visible en raison de la contribution de la source autochtone. La combinaison linéaire de des neuf indicateurs issus de l'ensemble des techniques montre également la

nécessité de disposer d'une information complémentaire entre les approches optique et moléculaires.

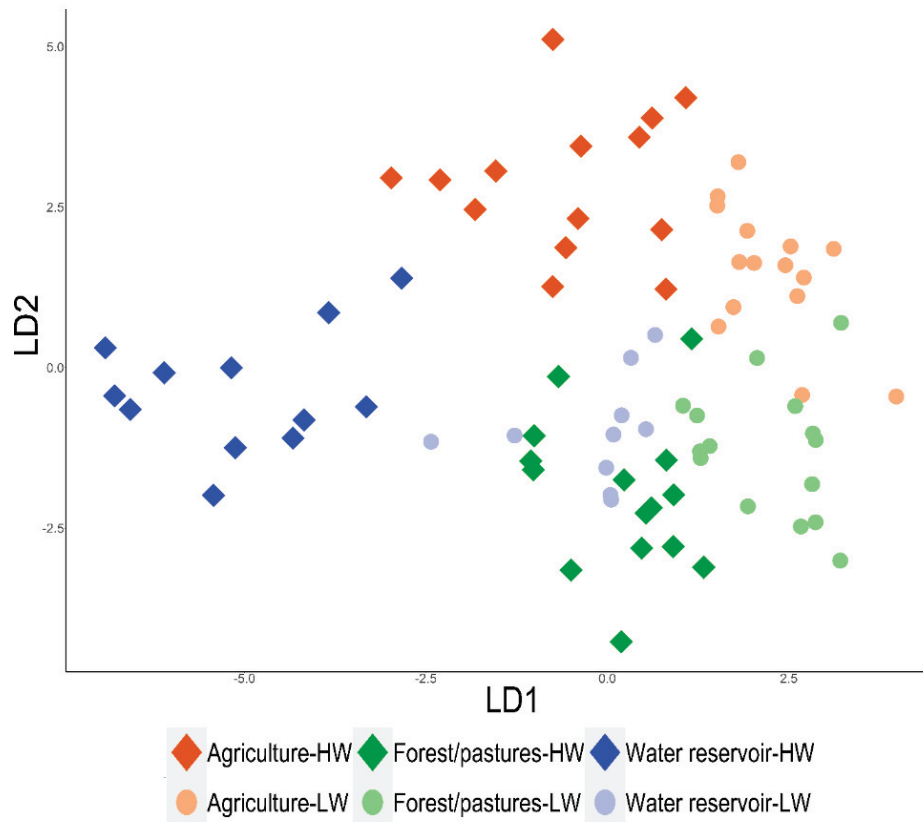


Figure 5-19. Analyse linéaire discriminante des échantillons de DOM représentatifs des sources diffuses terrigènes prélevés dans les BV de l'Ardières-Morcille, de l'Yzeron et du Gier.

5.2.5 Conclusion

L'objectif de cette deuxième partie de chapitre était d'étudier les empreintes de sources diffuses terrigènes présentées dans la Partie 5.1 à une échelle plus globale. Pour ce faire, la méthode d'échantillonnage basée sur des sous-bassins homogènes en termes d'usage des sols, et tenant compte des conditions hydrologiques, a été appliquée dans trois bassins versants : l'Ardières-Morcille, l'Yzeron et le Gier. Les sources échantillonnées sont caractérisées par un couvert végétal naturel (forêts/prairies), des usages agricoles (élevage, cultures, vignobles) et des retenues d'eau (étangs) très présents à l'échelle du BV. La comparaison des empreintes de sources a été réalisée à l'aide d'analyses multivariées en prenant en compte un large panel d'indicateurs issus des techniques optiques et moléculaires.

Le résultat a montré la présence d'un effet de bassin versant sur les sources forêts/prairies et les sources agricoles. Ces différences peuvent s'expliquer par divers facteurs, tels que le type de sol, la géologie ou l'activité et la diversité des micro-organismes, mais également par la différence des pratiques agricoles entre les BV. L'effet de bassin versant était en revanche moins visible pour les échantillons prélevés dans les retenues d'eau. Cela peut être dû à un temps de séjour de MOD plus élevé, qui peut induire des mécanismes de dégradation ainsi qu'une production autochtone plus importante et plus homogène dans ces écosystèmes.

Malgré l'effet BV très présent, il a été possible de construire des empreintes des sources diffuses terrigènes en utilisant une analyse LDA en minimisant la variabilité existante entre les BV pour une même source *a priori*. Le résultat de la LDA a montré qu'avec neuf indicateurs issus de l'ensemble des techniques, il était possible d'obtenir une typologie de MOD issue de sources forestières/prairies, agricoles et retenues d'eau. Une quatrième empreinte autochtone prélevée en période de basses eaux a également été identifiée. La distinction entre MOD terrigène et MOD autochtone nous permet de mieux comprendre les processus biogéochimiques qui se déroulent à l'échelle des écosystèmes aquatiques. Les résultats présentés ici ouvrent ainsi des perspectives pour identifier et tracer les sources de pollution et évaluer l'impact des usages des sols sur la qualité de l'eau.

5.2.6 Annexes

Annexe 5-1. Valeurs médianes, maximales et minimales du DOC et des indicateurs optiques et moléculaires des échantillons prélevés dans le BV de l'Ardières-Morcille

	Indicators	Forest/pastures High water			Forest/pastures Low water			Agriculture High water			Agriculture Low water			Water reservoir High water			Water reservoir Low water		
		Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max
Dissolved organic carbon	[DOC]	2.70	2.20	3.00	3.60	3.30	4.60	3.80	3.10	4.40	4.00	3.50	4.60	5.15	3.50	9.90	6.85	3.40	10.90
	E2:E3	5.48	5.24	5.88	5.06	4.82	5.27	5.43	5.02	5.93	5.32	4.97	6.02	2.98	2.59	3.72	5.30	4.89	5.63
	E2:E4	16.74	16.35	19.61	16.06	14.44	17.85	16.62	13.32	19.40	17.07	15.24	23.14	4.04	3.22	6.84	14.40	12.50	14.99
	E3:E4	5.36	5.18	5.77	5.18	4.90	5.50	5.34	4.69	5.89	5.34	5.01	6.21	2.23	1.95	3.20	4.72	4.26	4.83
	Suva 254	3.14	2.66	3.75	3.59	2.28	4.92	3.09	2.84	3.33	3.32	1.13	3.92	4.06	3.18	4.41	3.57	1.51	4.21
UV-Visible	S1 (275_295)	0.013	0.012	0.014	0.013	0.013	0.013	0.014	0.013	0.014	0.014	0.013	0.015	0.011	0.010	0.012	0.016	0.014	0.017
	SR	0.73	0.70	0.85	0.73	0.71	0.75	0.77	0.74	0.81	0.76	0.75	0.76	1.75	1.06	2.03	1.08	0.89	1.13
	BIX	0.71	0.63	0.75	0.68	0.61	0.73	0.74	0.68	0.77	0.70	0.64	0.78	0.76	0.69	0.80	0.76	0.72	0.78
	FI	1.28	1.25	1.44	1.27	1.17	1.33	1.32	1.27	1.35	1.29	1.21	1.37	1.32	1.28	1.36	1.30	1.28	1.38
	YFI	0.60	0.55	0.63	0.57	0.55	0.60	0.60	0.59	0.61	0.58	0.55	0.59	0.63	0.58	0.65	0.69	0.60	0.71
EEM fluorescence	HIX	8.62	7.07	9.74	7.34	5.47	8.87	9.14	8.45	9.50	7.41	6.49	8.15	8.06	7.33	9.77	5.38	4.95	6.03
	C1	0.54	0.43	0.71	0.87	0.67	1.19	0.78	0.64	0.99	0.94	0.77	1.26	1.35	0.60	2.74	1.06	0.71	1.61
	C2	0.23	0.17	0.30	0.38	0.30	0.53	0.30	0.25	0.36	0.39	0.34	0.44	0.49	0.23	0.88	0.39	0.29	0.52
	C3	0.09	0.08	0.13	0.16	0.14	0.17	0.14	0.10	0.16	0.17	0.16	0.18	0.19	0.13	0.41	0.33	0.17	0.46
	C4_EN	0.27	0.22	0.36	0.37	0.29	0.52	0.46	0.34	0.58	0.49	0.38	0.75	0.65	0.35	1.52	0.51	0.38	0.80
HPSEC/UV-fluorescence	C5_EN	0.14	0.06	0.18	0.30	0.14	0.38	0.15	0.12	0.21	0.20	0.17	0.25	0.52	0.10	1.10	0.57	0.13	1.04
	TSA_254	643.85	293.57	811.25	1110.71	865.51	1466.69	916.77	700.88	1105.78	1160.47	1043.00	1210.46	1040.42	731.21	2713.29	1494.84	839.08	1874.56
	TSA_280	465.78	245.02	586.91	807.96	620.70	1094.69	698.30	504.43	743.77	776.76	752.51	950.40	750.84	533.56	2007.22	1054.95	590.15	1296.65
	TSF_315-443	2609.18	1159.15	3491.17	4564.08	3140.72	6297.72	3908.73	3232.01	5195.96	4960.44	4120.88	6958.77	5190.58	2890.94	10238.37	5532.90	3685.49	8796.17
	TSF_275-354	966.03	618.97	1769.67	1684.50	1175.60	2173.45	1468.16	1230.60	2156.28	1860.67	1533.51	2562.41	2247.04	1092.08	4552.61	3008.09	1243.22	5305.51
	TSF_280-330	998.14	553.93	1829.69	1681.17	1202.25	2195.52	1350.95	1264.81	2168.41	1866.22	1616.38	2513.93	2190.47	1168.77	4445.08	3052.33	1415.03	5175.66
	HMW_TSA_254	1.26	0.16	2.19	0.77	0.47	4.21	0.20	0.00	1.17	0.48	0.22	4.18	1.13	0.80	3.49	2.19	0.47	3.31
	MMW_TSA_254	90.01	88.03	95.40	91.41	90.91	94.34	90.83	86.19	96.17	90.64	86.58	94.34	89.46	87.27	94.46	91.09	87.60	93.87
	LMW_TSA_254	8.16	4.44	11.03	5.71	4.63	8.12	8.26	3.23	13.81	8.89	5.18	12.12	9.11	4.37	10.23	6.72	5.65	9.09

HMW_TSA_280	2.25	0.60	2.94	3.05	0.59	61.87	1.80	0.29	49.43	1.13	0.71	3.81	24.68	1.00	67.47	1.59	0.03	3.51
MMW_TSA_280	84.81	80.40	86.22	90.93	28.59	91.57	87.53	41.06	88.91	92.54	82.14	98.13	64.32	22.91	91.87	92.00	89.37	93.03
LMW_TSA_280	13.17	11.69	18.49	7.83	6.03	9.55	10.54	9.51	11.20	6.75	0.73	14.06	10.19	7.12	11.24	6.10	4.53	10.16
HMW_TSF_315-443	0.73	0.42	1.15	0.50	0.21	17.95	0.78	0.26	0.98	0.52	0.00	0.77	0.39	0.09	0.70	0.41	0.13	0.58
MMW_TSF_315-443	82.30	81.34	88.99	84.44	68.20	89.95	83.34	78.45	84.54	86.45	84.74	91.79	87.12	71.56	90.52	85.43	80.51	89.92
LMW_TSF_315-443	16.93	9.87	18.03	13.85	9.55	16.00	16.39	14.66	20.58	13.04	7.81	14.59	12.44	8.90	28.35	14.04	9.95	19.17
HMW_TSF_275-354	6.62	0.67	14.37	6.08	0.14	24.57	24.86	0.83	57.07	4.48	0.00	46.70	3.70	2.91	45.30	7.67	0.08	14.11
MMW_TSF_275-354	80.13	64.49	81.84	84.85	63.87	90.50	62.32	31.65	87.66	83.73	45.33	89.07	74.68	37.02	94.83	82.23	74.52	93.21
LMW_TSF_275-354	19.09	10.93	33.15	9.36	6.24	14.88	11.28	9.13	19.10	10.90	7.93	13.44	18.50	1.72	24.45	10.10	6.71	11.37
HMW_TSF_280-330	5.97	0.00	27.45	2.98	1.77	7.93	0.00	0.00	3.82	5.27	0.33	8.03	4.25	1.64	30.78	11.99	0.10	16.74
MMW_TSF_280-330	78.47	18.81	91.35	87.21	81.87	93.40	82.27	75.53	89.05	81.82	75.93	89.02	79.65	9.92	91.20	80.90	75.35	85.22
LMW_TSF_280-330	11.19	8.65	80.20	8.89	4.39	11.70	14.70	10.51	24.47	13.82	7.66	16.04	17.76	3.84	59.30	8.75	4.63	14.68
RT0-6_MW0-500	4.666	4.528	4.736	4.631	4.504	4.820	4.636	4.619	4.658	4.650	4.618	4.743	4.554	4.253	4.737	4.544	4.420	4.714
RT6-12_MW0-500	6.325	6.294	6.410	6.250	6.131	6.269	6.208	6.168	6.307	6.249	6.184	6.284	6.196	5.755	6.328	6.043	5.892	6.233
RT12-18_MW0-500	6.821	6.763	6.901	6.684	6.585	6.736	6.672	6.607	6.761	6.645	6.596	6.696	6.596	6.258	6.724	6.447	6.246	6.723
RT0-6_MW500-1000	0	0	0	0	0	0	0	0	0	0.301	0.000	3.127	0.000	0.000	0.301	0.000	0.000	0.000
RT6-12_MW500-1000	4.215	4.117	4.225	4.115	3.992	4.203	4.045	3.990	4.206	4.085	3.998	4.234	3.987	3.595	4.124	3.958	3.854	4.106
RT12-18_MW500-1000	5.905	5.822	5.955	5.783	5.660	5.820	5.666	5.596	5.902	5.765	5.693	5.793	5.675	5.339	5.785	5.535	5.345	5.881
RT0-6_MW+1000	0	0	0	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0	0.000	0.000	0	0.000
RT6-12_MW+1000	0	0	0	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RT12-18_MW+1000	3.229	3.189	3.276	3.229	3.092	3.342	3.071	2.947	3.200	3.130	3.021	3.307	2.853	2.607	3.065	3.068	2.842	3.421

Annexe 5-2. Valeurs médianes, maximales et minimales du DOC et des indicateurs optiques et moléculaires des échantillons prélevés dans le BV de l'Yzeron

	Indicators	Forest/pastures High water			Forest/pastures Low water			Agriculture High water			Agriculture Low water			Water reservoir High water			Water reservoir Low water		
		Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max
Dissolved organic carbon	[DOC]	6.70	5.30	7.70	5.20	3.60	6.50	9.60	8.40	10.50	7.30	6.70	9.20	10.85	6.20	19.70	6.60	4.70	12.00
	E2:E3	6.78	5.96	7.00	7.16	6.91	8.24	5.94	5.75	6.57	6.95	6.83	7.68	6.31	5.86	6.58	6.40	5.51	7.73
UV-Visible	E2:E4	23.43	18.70	28.45	28.72	24.99	38.78	21.07	17.92	23.55	27.47	24.60	32.67	21.73	17.96	23.71	22.34	14.99	30.48
	E3:E4	6.35	5.58	7.04	7.11	6.87	8.63	5.92	5.36	6.39	7.31	6.78	8.04	6.04	5.15	6.34	6.33	5.08	7.83
	Suva 254	2.17	2.03	3.91	2.24	1.98	2.42	2.89	2.51	4.66	2.40	2.15	2.49	2.21	1.06	5.32	2.25	2.08	2.84
	S1 (275_295)	0.017	0.016	0.017	0.015	0.015	0.016	0.016	0.016	0.014	0.016	0.015	0.014	0.015	0.017	0.016	0.018	0.015	0.015
	SR	0.86	0.81	0.89	0.73	0.67	0.79	0.83	0.74	0.93	0.69	0.67	0.74	0.88	0.83	1.10	0.77	0.71	0.91
EEM fluorescence	BIX	0.73	0.72	0.76	0.75	0.66	0.79	0.74	0.71	0.76	0.75	0.72	0.82	0.74	0.68	0.78	0.75	0.70	0.83
	FI	1.32	1.30	1.37	1.34	1.27	1.38	1.33	1.29	1.39	1.35	1.34	1.43	1.34	1.25	1.38	1.32	1.29	1.45
	YFI	0.59	0.58	0.61	0.61	0.57	0.64	0.61	0.56	0.64	0.60	0.58	0.67	0.62	0.58	0.71	0.63	0.58	0.68
	HIX	6.33	6.13	6.73	4.40	3.39	7.84	8.71	6.98	10.29	6.70	3.38	9.43	7.34	3.75	8.06	6.43	5.55	6.61
	C1	1.01	0.78	1.26	0.83	0.54	1.22	2.05	1.63	2.77	1.55	1.44	1.84	1.44	0.69	2.71	1.21	0.68	2.52
	C2	0.39	0.32	0.48	0.31	0.19	0.43	0.73	0.57	0.91	0.55	0.49	0.66	0.48	0.24	0.87	0.42	0.25	0.90
	C3	0.21	0.20	0.23	0.19	0.11	0.44	0.30	0.25	0.37	0.38	0.25	0.47	0.28	0.19	0.38	0.29	0.19	0.56
	C4_EN	0.64	0.45	0.72	0.49	0.36	0.77	1.10	0.89	1.54	0.94	0.89	1.11	0.82	0.35	1.31	0.78	0.37	1.37
	C5_EN	0.10	0.00	0.15	0.03	0.00	0.13	0.57	0.44	1.22	0.22	0.17	0.67	0.44	0.15	1.56	0.22	0.08	0.88
	HPSEC/UV-fluorescence	TSA_254	191.67	161.84	220.67	183.74	149.25	198.63	249.61	237.98	280.31	198.46	178.98	219.39	201.08	47.63	289.79	187.32	165.36
TSA_280		134.56	110.42	153.67	122.28	101.75	139.79	178.68	158.08	202.21	146.21	125.37	151.47	131.45	31.78	201.24	127.70	113.05	170.96
TSF_315-443		902.70	739.18	943.62	782.26	620.35	953.91	1241.08	955.67	1492.02	1059.52	835.04	1091.09	1011.86	175.78	1193.97	890.75	676.39	1097.67
TSF_275-354		319.98	266.53	334.65	265.06	233.06	403.17	506.43	461.11	671.41	409.76	381.89	458.29	436.77	95.18	606.41	385.17	289.84	471.40
TSF_280-330		301.96	268.34	336.68	281.05	243.44	389.90	507.86	443.33	666.81	423.92	393.92	490.32	449.97	104.58	601.64	379.50	279.61	474.28
HMW_TSA_254		0.70	0.12	1.65	1.33	0.28	1.80	3.94	0.08	50.26	0.35	0.17	0.74	1.48	0.78	40.40	1.35	0.13	3.62
MMW_TSA_254		87.22	87.07	89.59	89.29	88.66	90.79	89.21	40.90	92.97	89.57	86.61	90.77	86.96	52.82	91.94	88.52	86.28	92.29
LMW_TSA_254		11.47	9.10	12.81	9.63	7.42	10.52	6.85	5.79	10.08	9.89	8.73	13.22	8.25	6.78	13.91	9.29	5.76	13.59
HMW_TSA_280		0.62	0.16	3.23	0.80	0.30	3.25	1.39	0.69	7.09	0.37	0.00	0.45	20.23	0.61	46.97	1.65	0.31	3.06
MMW_TSA_280		87.92	85.48	91.93	87.71	82.62	91.80	88.42	81.21	93.53	88.74	85.25	89.91	74.52	43.56	90.89	85.55	82.31	89.12

LMW_TSA_280	10.44	7.65	11.92	11.89	7.41	14.13	9.20	5.65	11.69	11.26	9.72	14.30	8.50	2.00	9.46	12.79	7.82	17.39
HMW_TSF_315-443	0.94	0.56	22.47	0.54	0.00	1.95	0.36	0.27	0.76	0.41	0.30	0.81	0.45	0.08	0.79	0.84	0.62	1.32
MMW_TSF_315-443	80.91	58.70	85.78	81.73	75.31	92.53	86.89	85.67	90.21	82.97	77.00	88.83	88.29	84.73	92.61	79.95	76.73	83.13
LMW_TSF_315-443	18.14	12.32	18.83	17.06	7.47	24.17	12.80	9.48	13.87	16.64	10.36	22.59	11.55	6.73	14.48	19.19	16.00	22.24
HMW_TSF_275-354	0.48	0.15	30.46	1.18	0.28	1.83	1.30	0.00	4.09	0.30	0.03	0.68	12.21	3.30	41.58	6.47	0.19	40.35
MMW_TSF_275-354	83.10	54.67	94.09	74.23	62.98	79.23	82.64	71.47	86.32	75.92	58.33	84.53	78.58	49.01	84.99	78.55	41.34	80.29
LMW_TSF_275-354	15.60	5.76	24.27	25.49	19.59	36.40	16.07	13.12	26.19	23.78	14.79	41.01	10.56	5.41	13.01	16.67	13.77	20.70
HMW_TSF_280-330	0.28	0.00	1.75	0.46	0.00	13.24	0.97	0.63	5.81	0.13	0.02	1.14	4.39	0.00	17.87	1.62	0.17	4.42
MMW_TSF_280-330	86.04	80.98	87.55	62.99	35.36	83.67	81.13	80.23	84.70	74.61	65.12	85.49	80.49	69.10	91.97	77.52	62.32	83.02
LMW_TSF_280-330	13.68	10.70	19.02	23.77	15.87	64.49	17.46	13.64	18.80	25.27	14.49	33.74	13.66	8.03	15.94	20.19	14.54	36.88
RT0-6_MW0-500	5.720	5.646	5.768	5.852	5.782	6.080	5.787	5.713	5.837	5.922	5.774	5.940	5.740	5.643	5.958	5.911	5.765	6.254
RT6-12_MW0-500	5.661	5.605	5.747	5.755	5.671	5.939	5.573	5.545	5.607	5.684	5.563	5.738	5.573	5.325	5.684	5.703	5.457	5.882
RT12-18_MW0-500	6.919	6.771	6.977	7.064	6.875	7.244	6.817	6.761	6.889	6.952	6.774	7.016	6.818	6.565	6.934	6.924	6.677	7.202
RT0-6_MW500-1000	0	0	0	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RT6-12_MW500-1000	3.317	3.214	3.574	2.787	2.530	2.877	3.109	2.784	3.296	2.561	2.431	2.657	3.127	2.650	3.479	2.814	2.556	2.997
RT12-18_MW500-1000	4.593	4.582	4.772	4.663	4.577	4.861	4.487	4.390	4.504	4.492	4.455	4.627	4.571	4.138	4.763	4.652	4.367	4.742
RT0-6_MW+1000	0	0	0	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0	0.000	0.000	0	0.000
RT6-12_MW+1000	0	0	0	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RT12-18_MW+1000	0	0	0	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

LC-HRMS
(LOG10)

Annexe 5-3. Valeurs médianes, maximales et minimales du DOC et des indicateurs optiques et moléculaires des échantillons prélevés dans le BV du Gier

Technique	Indicators	Forest/pastures High water			Forest/pastures Low water			Agriculture High water			Agriculture Low water			Water reservoir High water			Water reservoir Low water		
		Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max
DOC	[DOC]	2.20	1.90	4.10	3.30	1.80	4.40	6.40	4.40	11.30	4.00	3.80	4.40	8.30	5.30	19.70	3.80	2.70	6.90
	E2:E3	8.37	0.42	12.70	6.47	5.94	9.35	6.79	6.07	8.30	6.14	5.41	6.38	6.15	5.70	8.71	6.47	5.69	7.39
	E2:E4	27.76	0.64	29.06	19.17	17.15	31.44	36.86	18.52	63.87	17.60	16.72	20.69	22.77	21.60	52.24	18.74	14.16	22.86
UV-Visible	E3:E4	7.71	1.57	19.28	6.45	5.71	10.41	7.80	6.82	10.54	5.85	5.40	6.30	5.56	5.11	8.21	6.05	5.33	6.41
	Suva 254	2.25	1.36	3.17	2.47	1.85	2.67	1.68	0.86	5.49	2.71	2.32	2.91	3.33	0.70	3.44	2.29	2.04	2.87
	S1 (275_295)	0.019	0.016	0.022	0.013	0.010	0.014	0.018	0.016	0.022	0.013	0.013	0.014	0.020	0.019	0.023	0.014	0.012	0.018
EEM fluorescence	SR	0.81	0.63	0.96	0.69	0.43	0.76	0.70	0.63	0.89	0.72	0.71	0.74	1.02	1.00	1.08	0.85	0.77	0.99
	BIX	0.86	0.80	0.97	0.79	0.68	0.86	0.79	0.67	0.81	0.72	0.72	0.84	0.81	0.78	0.84	0.79	0.76	0.83
	FI	1.40	1.38	1.53	1.34	1.31	1.38	1.31	1.29	1.36	1.35	1.31	1.36	1.34	1.30	1.34	1.36	1.30	1.38
	YFI	0.74	0.63	0.75	0.63	0.58	0.71	0.61	0.53	0.70	0.59	0.58	0.66	0.72	0.70	0.75	0.72	0.68	0.76
	HIX	6.35	2.53	9.79	5.88	3.98	8.46	8.22	5.13	11.90	8.23	6.32	9.16	4.37	3.67	5.27	4.82	4.00	5.48
	C1	0.35	0.28	0.84	0.43	0.22	0.81	1.69	0.70	2.08	0.79	0.67	0.93	0.68	0.67	1.06	0.61	0.47	0.96
	C2	0.11	0.09	0.32	0.15	0.07	0.29	0.54	0.25	0.82	0.32	0.23	0.36	0.24	0.23	0.35	0.20	0.17	0.33
	C3	0.13	0.13	0.19	0.11	0.09	0.17	0.20	0.15	0.48	0.16	0.15	0.19	0.28	0.25	0.30	0.23	0.16	0.44
	C4_EN	0.28	0.21	0.53	0.31	0.18	0.54	0.94	0.36	0.97	0.47	0.41	0.54	0.35	0.34	0.52	0.39	0.28	0.54
	C5_EN	0.01	0.00	0.12	0.01	0.00	0.09	0.58	0.12	0.81	0.18	0.13	0.20	0.43	0.39	0.82	0.19	0.16	0.53
HPSEC UV-fluorescence	TSA_254	395.89	227.95	855.95	478.74	210.76	794.72	2021.55	792.86	2367.17	859.45	737.47	981.54	1129.04	1078.76	1625.80	674.69	474.52	1153.60
	TSA_280	302.34	180.30	606.34	358.05	156.44	581.41	1410.52	418.26	1736.06	628.38	515.13	727.28	811.37	763.64	1078.08	508.58	340.30	784.84
	TSF_315-443	1667.45	1259.32	4501.91	2005.47	976.36	3738.41	9120.25	3499.75	12214.64	3874.78	3129.96	4533.19	3378.74	3279.15	6208.89	2675.73	1924.12	4357.66
	TSF_275-354	688.80	595.26	1745.54	1016.75	338.38	1350.45	3502.74	1823.36	5371.09	1697.46	1496.74	1808.27	2258.87	2108.18	3907.96	2233.08	1565.79	3685.31
	TSF_280-330	743.36	588.39	1796.23	943.09	328.51	1624.59	3554.54	1806.08	5217.09	1707.98	1604.62	1913.76	2306.14	2098.22	3943.32	2401.54	1524.38	3655.79
	HMW TSA_254	1.18	0.18	46.56	0.91	0.48	1.34	2.24	0.91	5.03	1.27	0.28	7.63	2.44	1.62	3.41	0.77	0.45	2.37
	MMW TSA_254	87.71	33.70	98.02	85.07	67.88	88.67	88.77	87.12	98.39	88.52	80.26	94.01	80.63	80.35	91.87	86.38	82.68	88.46
LMW TSA_254	11.11	1.79	19.74	14.45	10.53	31.01	6.20	0.14	10.64	10.21	5.71	12.10	16.24	6.52	16.93	13.04	10.70	14.94	
	HMW TSA_280	1.83	0.81	57.22	3.60	1.53	20.83	2.30	1.61	3.51	1.30	2.07	2.59	2.06	4.62	0.60	0.00	2.21	
	MMW	62.31	41.02	84.03	79.14	45.90	83.43	91.22	81.75	96.87	86.40	83.01	92.44	81.69	76.65	96.66	81.94	75.88	95.26

Annexe 5-4 : Classement des groupes d'échantillons pour la construction de l'analyse linéaire discriminante

Groupe	Bassin versant	Echantillons	Groupe	Bassin versant	Echantillon
Groupe 1 Forêts/prairies période de hautes eaux	Ardières-Morcille	Ard-FP-HW-01 Ard-FP-HW-02 Ard-FP-HW-03 Ard-FP-HW-04 Ard-FP-HW-05	Groupe 4 Forêts/prairies période de basses eaux	Ardières-Morcille	Ard-FP-LW-01 Ard-FP-LW-02 Ard-FP-LW-03 Ard-FP-LW-04 Ard-FP-LW-05
	Yzeron	Yze-FP-HW-01 Yze-FP-HW-02 Yze-FP-HW-03 Yze-FP-HW-04 Yze-FP-HW-05		Yzeron	Yze-FP-LW-01 Yze-FP-LW-02 Yze-FP-LW-03 Yze-FP-LW-04 Yze-FP-LW-05
	Gier	Gie-FP-HW-01 Gie-FP-HW-02 Gie-FP-HW-03 Gie-FP-HW-04 Gie-FP-HW-05		Gier	Gie-FP-LW-01 Gie-FP-LW-02 Gie-FP-LW-03 Gie-FP-LW-04 Gie-FP-LW-05
Groupe 2 Agriculture période de hautes eaux	Ardières-Morcille	Ard-AG-HW-01 Ard-AG-HW-02 Ard-AG-HW-03 Ard-AG-HW-04 Ard-AG-HW-05	Groupe 5 Agriculture période de basses eaux	Ardières-Morcille	Ard-AG-LW-01 Ard-AG-LW-02 Ard-AG-LW-03 Ard-AG-LW-04 Ard-AG-LW-05
	Yzeron	Yze-AG-HW-01 Yze-AG-HW-02 Yze-AG-HW-03 Yze-AG-HW-04 Yze-AG-HW-05		Yzeron	Yze-AG-LW-01 Yze-AG-LW-02 Yze-AG-LW-03 Yze-AG-LW-04 Yze-AG-LW-05
	Gier	Gie-AG-HW-01 Gie-AG-HW-02 Gie-AG-HW-03 Gie-AG-HW-04 Gie-AG-HW-05		Gier	Gie-AG-LW-01 Gie-AG-LW-02 Gie-AG-LW-03 Gie-AG-LW-04 Gie-AG-LW-05
Groupe 3 Retenue d'eau période de hautes eaux	Ardières-Morcille	Ard-WR-HW-01 Ard-WR-HW-02 Ard-WR-HW-03 Ard-WR-HW-04	Groupe 6 Retenue d'eau période de basses eaux	Ardières-Morcille	Ard-WR-LW-01 Ard-WR-LW-02 Ard-WR-LW-03 Ard-WR-LW-04
	Yzeron	Yze-WR-HW-01 Yze-WR-HW-02 Yze-WR-HW-03 Yze-WR-HW-04		Yzeron	Yze-WR-LW-01 Yze-WR-LW-02 Yze-WR-LW-03 Yze-WR-LW-04
	Gier	Gie-WR-HW-01 Gie-WR-HW-02 Gie-WR-HW-03 Gie-WR-HW-04		Gier	Gie-WR-LW-01 Gie-WR-LW-02 Gie-WR-LW-03 Gie-WR-LW-04

Chapitre 6.

Utilisation de la MOD comme traceur de sources de pollution à l'échelle d'un bassin versant péri-urbain : approche optique et moléculaire non-ciblée.

Chapitre 6. Utilisation de la MOD comme traceur de sources de pollution à l'échelle d'un bassin versant péri-urbain : approche optique et moléculaire non-ciblée (Publication n°3, en préparation).

Avant-propos

Dans ce chapitre, nous étudions les propriétés optiques et moléculaires de la matière organique dissoute (MOD) provenant de sources diffuses. Les échantillons ont été prélevés selon la méthodologie présentée dans le chapitre 5, et de sources ponctuelles issues d'activités urbaines. Le but est d'identifier des marqueurs pertinents pour le traçage des sources via les chemins de l'eau. Les marqueurs comprennent à la fois une série d'indicateurs classiques UV-Visible et EEM de fluorescence et des indicateurs HPSEC/UV-fluorescence et LC-HRMS tels que développés dans le cadre de cette thèse et présentés dans le chapitre 4. Les résultats obtenus sont présentés sous la forme d'une publication scientifique en préparation.

6.1 Résumé de la publication

L'étude bibliographique (chapitre 1) a montré que l'utilisation de la MOD comme traceur des rejets de pollution par les voies d'eaux n'a été que rarement envisagée. Cela s'explique principalement par sa forte réactivité et ses transformations dont elle peut faire l'objet (biodégradation, photodégradation, précipitation, adsorption ...) (Søndergaard et Thomas, 2004), ainsi que par la multiplicité des apports potentiels de MOD dans le cours d'eau (production autochtone, apports souterrains, apports terrigènes et anthropiques). Néanmoins, des développements analytiques récents, ainsi qu'une exploitation plus approfondie des données optiques et moléculaires, ont ouvert de nouvelles perspectives permettant l'utilisation de la MOD comme traceur pertinent des sources de pollutions anthropiques (Romero González-Quijano et al., 2022).

Dans ce contexte, cette étude vise à évaluer le potentiel de la MOD en tant que traceur des rejets anthropiques depuis leurs sources jusqu'aux rivières en aval. Pour cela, une base de données regroupant les résultats d'analyses d'un ensemble d'échantillons issus de sources naturelles et anthropiques (e.g. forêts, ruissellement urbain, rejets de stations de traitement des eaux usées) a été mise en place (n = 85 échantillons). Chaque échantillon a été caractérisé par un large panel de techniques analytiques : le dosage du carbone organique dissous, la

spectroscopie UV-Visible et EEM de fluorescence, la chromatographie d'exclusion stérique couplée à la détection UV et fluorescence et la spectrométrie de masse haute résolution.

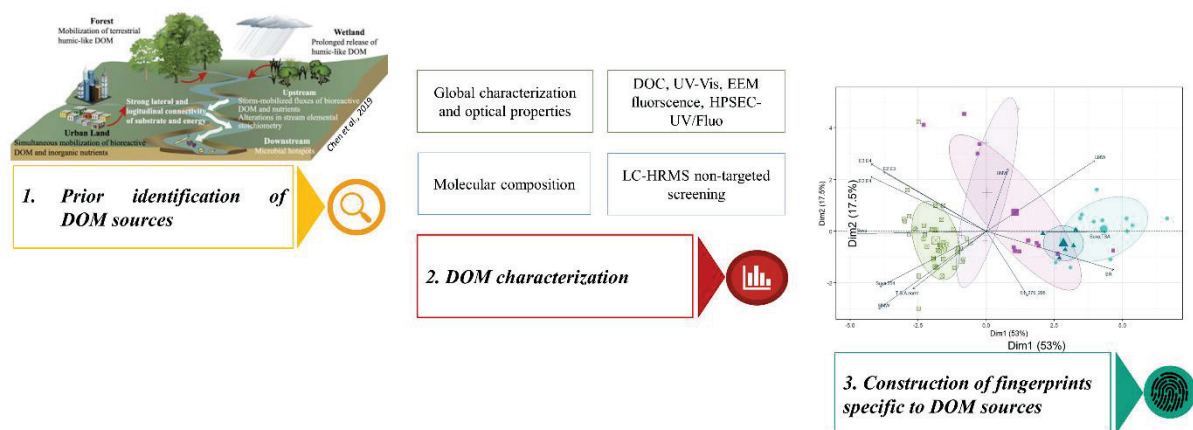


Figure 6-1. Résumé graphique de la publication scientifique: *Using dissolved organic matter as a tracer of land-use-inputs at the scale of a peri-urban watershed: optical and non-targeted molecular combined approaches.*

L'intégration de l'ensemble des indicateurs (optiques et moléculaires, n = 47) dans des analyses multivariées a permis de mettre en évidence des empreintes distinctes entre les sources de MOD (Figure 6-1). Ceci a permis de différencier les sources à différentes échelles, en fonction des usages de sols et des conditions hydrologiques dans les cours d'eau, et également d'identifier les contributions de ruissellement urbain et les rejets de STEU. Des différences d'empreintes ont également été observées entre les types de traitement appliqués dans les STEU (traitement intensif à base de boues activées et traitement extensif à base de filtres plantés de roseaux). A travers l'analyse LC-HRMS non ciblée, il a été possible d'identifier des composés moléculaires spécifiques de chaque type de source de MOD, ainsi que des composés d'intérêt commun à deux sources ou plusieurs sources. Ceci ouvre des perspectives d'application dans des modèles de traçage pour mieux identifier les sources de rejets et ainsi pour mieux caractériser et prévenir la contamination des environnements aquatiques.

Using dissolved organic matter as a tracer of land-use-inputs at the scale of a peri-urban watershed: optical and non-targeted molecular combined approaches

Amine Boukra¹, Matthieu Masson¹, Edith Parlanti², Cécile Miège¹, Marina Coquery¹

¹INRAE, UR RiverLy, 69625 Villeurbanne, France

²Univ. Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, F-33600 Pessac, France

Correspondence to: matthieu.masson@inrae.fr ; +33472208706; RiverLy, INRAE

5 rue de la Doua, Villeurbanne, France, 69625

Highlights:

Multiple source discrimination using optical and molecular data from DOM analyses.

Implementation of global LC-HRMS indicators based on mass and polarity distribution.

Identification of optical descriptors and molecular compounds specific to types of DOM sources.

Impact of land use and hydrologic conditions on DOM quality in downstream watercourses.

The fingerprint of planted reed filters discharges is closer to natural water DOM than activated sludges

Abstract

Tracing the sources of organic pollutions in waters is an essential issue to identify and address the causes of water contamination. In this paper, we studied the optical and molecular properties of dissolved organic matter (DOM) originating from more or less anthropogenic diffuse and point sources (dataset n=82), with the objective of identifying relevant markers for source tracing through water pathways. The markers correspond to a series of indicators that provide information on the characteristics of DOM (e.g. molecular weight, aromaticity, relative contribution of protein-type compounds). New indicators from non-targeted LC-HRMS data are introduced for the first time for this kind of application. They correspond to nine classes of mass and polarity distribution for molecular compounds; and the results showed their relevance to differentiate DOM sources. The integration of all these indicators (optical and molecular, n=47) in multivariate analyses allowed to highlight distinct fingerprints between the DOM sources and according to the land uses, hydrological conditions in the watercourses, runoff contributions from urban surfaces and type of treatment in the wastewater treatment plants. This multi-analytical approach has also shown the complementarity between the optical and molecular techniques. A further examination on a smaller dataset allowed identifying molecular compounds specific to each type of DOM source, and compounds of common interest to two or more sources. This opens perspectives for application in tracing models to better identify sources of discharges and prevent contamination of the aquatic environments.

Keywords: pollution sources tracking, dissolved organic matter, optical properties, high-resolution-mass spectrometry, land use, urban discharges, autochthonous production.

1. Introduction

At the watershed scale, many sources of discharge can contribute to the downstream water quality. Industrial activities, urban development, road transport or agricultural activities generate a supply of pollutants (e.g. pharmaceutical products, hydrocarbons, pesticides and transformation by-products) in the watercourses, with possible disastrous consequences for the ecosystem (Gómez-Regalado et al., 2023; Grgić et al., 2023; Rad et al., 2022). These sources can be classified into point or diffuse sources. Point sources are those that originate from a single, identifiable or channeled source such as sewage from wastewater treatment plant (WWTP) or industries (Lapworth et al., 2012). Diffuse sources are those that are spread out over a large area and have no single identifiable source such as runoff from agriculture activities (Tran et al., 2019). The identification of these pollution sources at the scale of a peri-urban catchment area is essential for the monitoring of water quality considering the impact of climate change (hydrolclimatic changes) and land use evolution. Dissolved organic matter (DOM) is ubiquitous ; it originates from these anthropogenic pollution sources and also from natural sources such as decomposition of plants and animal material, microbial activity and soil leaching (Butturini et al., 2022; da Silva et al., 2021; Morling et al., 2017). It is currently very difficult to precisely define the composition of DOM because it is highly dynamic and varies according to the origins mentioned above and other environmental factors (e.g. pH, temperature, hydrology, microbial activity) (Aiken, 2014). Nevertheless, it is possible to give an operational definition of DOM as being a wide range of organic compounds obtained after water filtration at 0.7 μm .

The high complexity of DOM composition as well as its close link to pollution sources have raised the interest of the scientific community. Thus, a large number of studies have been interested in studying the composition of DOM through its global (carbon determination), optical (UV-Visible, excitation-emission fluorescence, synchronous fluorescence) and molecular (nuclear magnetic resonance spectroscopy, size exclusion chromatography, liquid chromatography coupled to high-resolution mass spectrometry etc...) characteristics (Derrien et al., 2019). Among the panel of techniques used to characterize DOM according to its sources, UV-Visible spectroscopy and EEM fluorescence have shown relevance to characterize DOM chromophoric and fluorescent fractions. For example, the development of indicators, such as absorbance ratios (e.g. E2E3, E3.E4), SUVA, HIX and FI indicators, has allowed to

differentiate DOM originating from terrigenous vs autochthonous inputs (Murphy et al., 2008). These indicators provide information on the molecular weight, aromaticity and content of humic compounds in DOM samples. Fluorescence matrix decompositions have also allowed the identification of specific fluorophore compounds related to agricultural, industrial sources and wastewater discharges (e.g. pesticides, hormones) (Ferretto et al., 2014; Hassoun et al., 2017; Goffin et al., 2018). Other more recent indicators derived from size exclusion chromatography have made possible to differentiate diffuse sources of DOM (e.g. terrestrial with land use impact, marine) based on the molecular weight distribution of compounds contained in a sample (Asmala et al., 2021; Boukra et al., 2023). These indicators represent percentages of contribution of low, medium and high molecular compounds. Finally, the use of high-resolution mass spectrometry techniques (HRMS, FT-ICRMS) allowed to explore the molecular scale and to identify various groups of molecules specific to terrigenous sources (compounds such as lignins, lipids or some proteins) or to urban and industrial sources (e.g. Polycyclic Aromatic Hydrocarbons, phytosanitary drugs, pharmaceutical compounds) (Choi et al., 2021). In the end, each technique provides a partial information on DOM composition that can be linked to a type of pollution source. However, the use of DOM as a proxy of pollution discharges through water pathways was not always considered. This is mainly due to its high reactivity and then instability/transformation (by biodegradation, photodegradation, precipitation, adsorption) (Søndergaard and Thomas, 2004), as well as to the multiplicity of potential inputs in the watercourse (autochthonous production, underground inputs, terrigenous inputs...). Nevertheless, recent analytical developments, as well as deeper exploitation of optical and molecular data (chemometric tools) have opened new perspectives allowing the use of DOM as relevant tracers of anthropogenic pollutions sources (Romero González-Quijano et al., 2022).

To increase knowledge on the sources and pathways of organic pollutions and to improve strategies for organic pollution tracking in watercourses, we aim at assessing the potential of DOM as a tracer of anthropogenic discharges from their sources to the downstream rivers. More precisely, our objective is to determine if there is a link between DOM characteristics and types of sources encountered at the scale of a peri-urban watershed characterized by natural lands and more or less high anthropogenic pressure areas with agricultural and urban activities. To do this, a multi-analytical approach based on the use of optical (UV-Visible, EEM

fluorescence and HPSEC/UV-fluorescence) and molecular (LC-HRMS/QTOF) tools was applied on a sample set (n=85 samples) representative of some well-defined sources of DOM collected in the Yzeron watershed (Rhône, France).

2. Materials and methods

2.1. Sampling strategy

2.1.1. Yzeron watershed description

This study is based on the establishment of a sample database characteristic of a wide range of natural and anthropogenic sources. The Yzeron watershed (149 km²) located in eastern France (Figure 6-2A), is considered as a model example of peri-urban catchment areas and is monitored in the framework of long-term observatories (OTHU and OZCAR-RI).

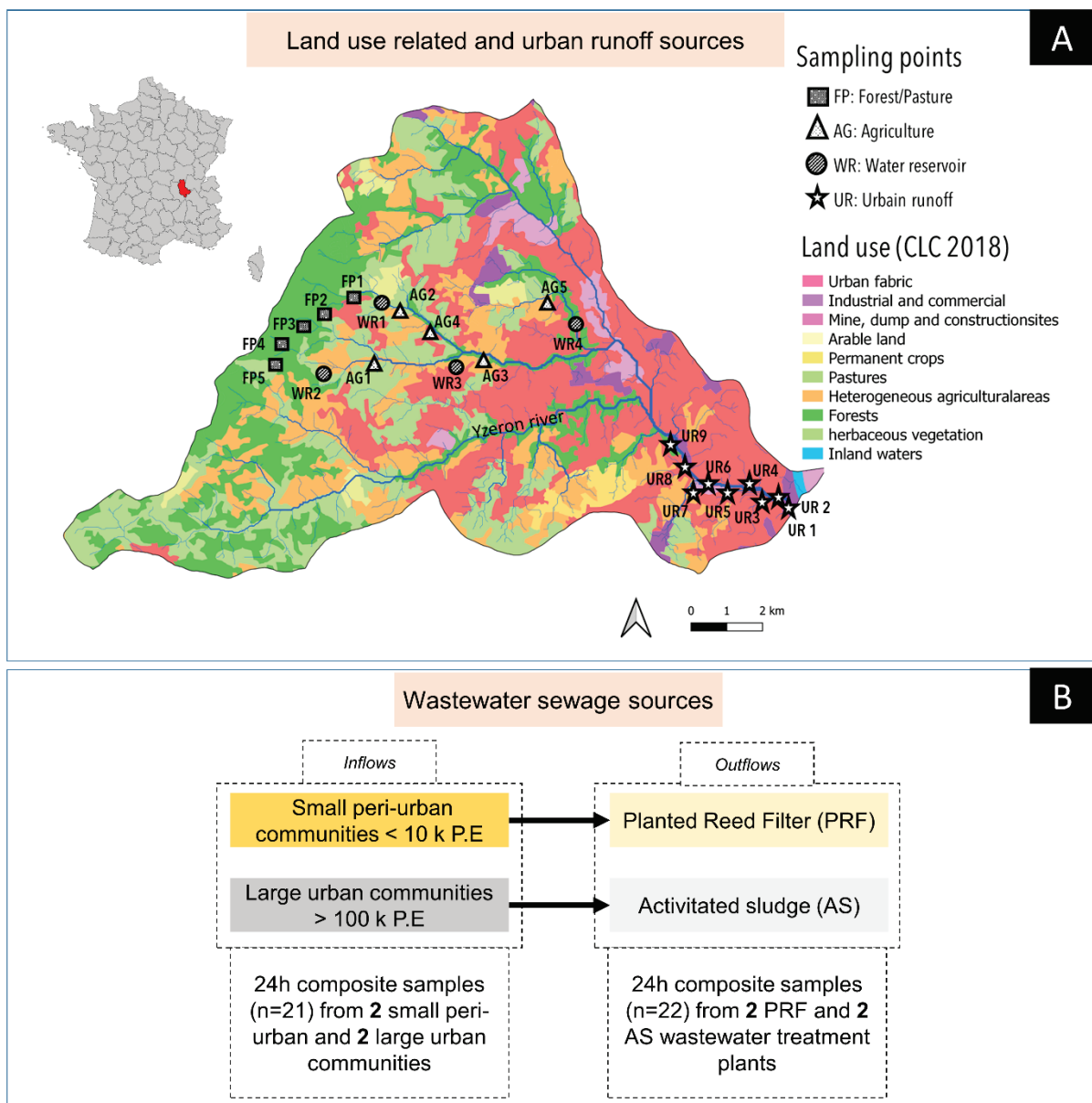


Figure 6-2. Sampling strategy of DOM sources with A/ the location of diffuse source related to land-use and urban runoff sources (Corine Land Cover CLC 2018) and B/ description of wastewater sewage (inflows and outflows) sources

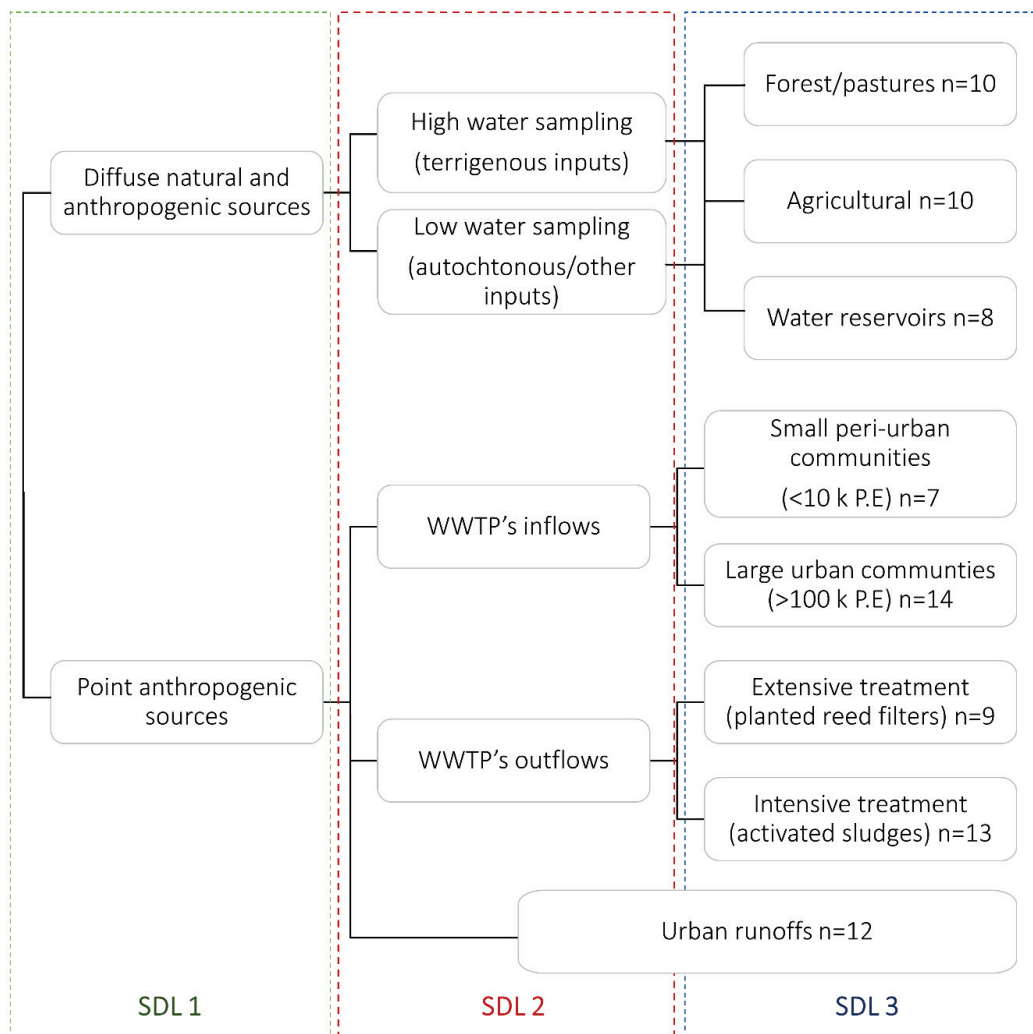
Yzeron is characterized by very heterogeneous land uses, with forests (deciduous and coniferous) and farms (grazing areas and various crops) covers in the upstream area. In the center, agricultural crops and grasslands represent the majority of land uses and high urban density is located in the downstream area. The climate of the site is semi-continental with maximum temperatures in summer at 17 °C and minimum temperatures in winter at 8 °C and average annual precipitations at 830 mm (Braud et al., 2013). The geology of the basin is largely dominated by crystalline and metamorphic rocks of the gneiss, granite, and schist type. A local quaternary alluvium on a crystalline and metamorphic base is observed in the east of the watershed (Gnouma, 2006; Labbas, 2017).

2.1.2. Classification of DOM sources and introduction to the concept of scale discrimination level

Through its particular configuration, the watercourses in Yzeron can receive diffuse sources (runoffs from rural soils influenced by land use) but also point sources of urban-channeled discharges during important rainy events, as well as discharges of WWTP. Other discharges of untreated wastewater have also been observed through overflowing sewage networks in downstream urban part. In this study, we classified the main sources encountered according to the nature of the discharge to the waterways (Figure 6-3). Thus, a first level of source discrimination (SDL1) separates the sources nature into two groups: a first group of samples from diffuse sources submitted more or less to anthropogenic pressure (in rural area) and a second group of samples from point high anthropogenic sources (in urban area).

The diffuse sources were classified into a second level (SDL 2) based on the hydrological conditions of the watercourses (high water and low water sampling). The purpose of the SDL2 level was to differentiate between DOM from terrigenous inputs taken during high water periods and locally produced DOM (autochthonous DOM) during low water periods when surface runoff is lower. For anthropogenic point sources, the SDL2 level separates the sample group into three categories: samples from raw wastewaters, samples from treated wastewaters and samples from urban impervious surface runoffs. Treated wastewaters and urban runoffs are discharged directly into the environment, while raw wastewaters are channeled to wastewater treatment plants. However, discharges of raw wastewaters into the environment can occur during heavy rainfall events or during accidents caused by poor

network connections. It was therefore relevant to have a typology of the DOM from these anthropogenic sources.



Complexity of source discrimination: necessity to have a complete information of DOM properties (global, optical and molecular)

Figure 6-3. Classification of the collected DOM sources according to the concept of source discrimination level (SDL). WWTP: wastewater treatment plants, P.E: population equivalent.

To go further, a third level of classification (SDL3) has been established. For the group of diffuse sources: the SDL3 level separates the samples according to the uses and occupations of soils upstream of the sampling points. These samples were taken according to the three main land uses found in the Yzeron basin (forest/pasture, agriculture and water reservoir). Regarding the point source discharges, the SDL3 level separates raw untreated wastewaters according to discharges from small communities (mainly domestic waters) versus discharges from large urban communities (domestic, urban, industrial waters). Treated wastewaters have

also been classified taking into account the type of treatment applied in the WWTP: planted reed filter treatment (PRF) for raw waters from small communities versus intensive activated sludge treatment (AS) for raw waters from large urban communities.

2.1.3. Diffuse sources of DOM related to land use: stations and sampling

Water samples were collected manually from watercourses downstream of homogeneous land uses of forest/grassland (FP1 to FP5), agricultural (AG1 to AG5), and water reservoirs (WR1 to WR4), see Figure 6-2A. Samples were collected during two campaigns, depending on the hydrological status of the watercourses: a first sampling campaign in January 2021 during high water period ($0.42 \text{ m}^3 \cdot \text{s}^{-1}$ at Craponne) and a second campaign in July 2021 during low water period ($0.20 \text{ m}^3 \cdot \text{s}^{-1}$ at Craponne). This sampling strategy is based on taking a sample as representative as possible of a diffuse source (Boukra et al., 2023). For DOC and optical analyses, the water samples were collected in 250 mL borosilicate glass bottles. For LC-HRMS analyses, the water samples were collected in 100 mL borosilicate amber glass bottles. All the glass bottles were previously cleaned using an automatic washer at 90°C , then calcined in an oven at 500°C . The samples were filtered in the field using polypropylene syringe and polyvinylidene fluoride (PVDF) filters with a porosity of $0.45 \mu\text{m}$ for DOC and optical analyses and with a porosity of $0.2 \mu\text{m}$ for LC-HRMS analysis, in order in order to guarantee the best preservation of the DOM by limiting the presence of bacteria as much as possible. During their transport to the laboratory, the samples were stored in a cooler (at $\pm 4^\circ\text{C}$) protected from light. In laboratory, the samples were stored in 100 mL borosilicate glass bottles at 4°C in the dark until DOC and optical analyses and stored in amber glass vials of 1.5 mL at -20°C for LC-HRMS analyses. Field blanks using ultrapure water (resistivity at $18.2 \text{ M}\Omega \cdot \text{cm}$) were prepared in laboratory using the same protocol as for the other samples, in order to identify possible contamination due to bottles, syringes or filters.

2.1.4. Point anthropogenic sources of DOM: stations and sampling

2.1.4.1. Urban runoff samples

Downstream part of the Yzeron watershed (Figure 6-2A), characterized by a very dense urban area, has been defined as the preferential sampling area for DOM sources related to urban runoff. For this purpose, nine points (Stormwater network outlets, UR1-9) were defined as point sources of discharge in the Yzeron River. The samples were taken during a rainy event

(15 mm over a period of 3 hours; <https://www.meteociel.fr/>). Points UR1, UR7 and UR9 were sampled twice at intervals of 30 minutes to evaluate the temporal variation. The samples were then filtered following the protocol described in Section 2.1.3 and stored at 4°C.

2.1.4.2. Wastewater samples

The wastewater sources of DOM were classified according to the discharges upstream wastewater WWTP (Figure 6-2B). Two WWTP (300 k and 120 k population equivalent-P.E) using activated sludge treatment (AS) and two WWTP (1 k P.E) with planted reed filters treatment (PRF) were considered. The WWTP with AS treatment, considered as an intensive treatment, use a biological treatment of free culture type associated with longitudinal decanters. The WWTP with PRF treatment use an extensive treatment based on soil natural remediation capacity and is widely used in France (Morvannou et al., 2015) .

The overflow of the sewerage network observed on the Yzeron led us to consider the collection of untreated raw waters in order to obtain the fingerprint of DOM associated with this type of discharge. Thus, for each WWTP, one untreated inflow water and one treated outflow water were taken over 24 h at regular intervals of 30 min using automatic samplers. The 24 h composite samples were then stored at 4°C in high-density polyethylene (HDPE) plastic bottles for transport to laboratory. Once in the laboratory, the samples were filtered and stored following the protocol described in Section 2.2.2. In total, there were 21 samples of WWTP inflow and 22 samples of WWTP outflows spread over a period of approximately 1 year.

2.2. Analytical strategy

2.2.1. Dissolved organic carbon (DOC) analyses

The measurement of DOC was performed using an Analytik Jena multi N/C® 3100 TOC analyzer based on high temperature catalytic combustion according to the standard NF EN 1484. The calibration, rinsing, quality controls solutions, as well as samples, were introduced in precombusted (at 500°C during 2 h) glass vials, where they were acidified to pH=1 with hydrochloric acid (HCl 37% RPE). The detection limit was 0.2 mgC.L⁻¹ and the analytical uncertainty was 12%. The accuracy of the analyses (105%) was routinely checked using a

certified standard of potassium hydrogen phthalate ($5 \text{ mg}\cdot\text{L}^{-1}$, Fluka[®]) different from that used for the calibration ($0.2\text{--}25 \text{ mg}\cdot\text{L}^{-1}$, Sigma-Aldrich[®]).

2.2.2. UV-Visible spectroscopic analysis

UV-Visible absorbance spectra were acquired using a Shimadzu UV 1900 spectrophotometer with a 1 cm path length quartz cell (Hellma Analytics[®]) between 200 and 800 nm with a resolution of 1 nm and a scan speed of $0.2 \text{ s}\cdot\text{nm}^{-1}$. Ultrapure water was used for baseline correction. A series of classical descriptive indicators (SUVA_{254} , absorbance ratios and spectral slopes) were calculated from the generated absorbance spectra (Supplementary Material 1).

2.2.3. EEM fluorescence spectroscopy and PARAFAC decomposition

Excitation-emission matrix (EEM) fluorescence analyses were performed using an Aqualog (Horiba Scientific) spectrofluorometer in a 1 cm path length quartz cuvette (Hellma Analytics[®]) thermostated at 20°C . EEM fluorescence detailed data acquisition are described in (Boukra et al., 2023). The generated spectra were treated using TreatEEM-V1 software developed by Dario Omanović (Ruđer Bošković Institute, Zagreb, Croatia; <https://sites.google.com/site/daromasoft/home/treateem>). The fluorescence indices HIX (Zsolnay), BIX, FI, YFI (Supplementary Material 1) were calculated for an excitation wavelength of 250, 310, 370 and 280 nm respectively (Heo et al., 2016; Huguet et al., 2009; Mcknight et al., 2001; Zsolnay et al., 1999)

The multi-way parallel factor analysis (PARAFAC) was applied on two sets of EEM spectra in order to decompose the fluorescence matrices into different contributions of fluorophores. The first set included 97 samples (28 samples corresponding to diffuse sources of DOM collected for this study, 12 urban runoff samples, 57 samples of diffuse sources collected for other purposes). The second set of EEM spectra included 43 samples, corresponding to WWTP raw inflows and treated outflows. The PARAFAC model was run, using the drEEM toolbox under MatLab R2020a, for 4 to 8 components with non-negativity constraints (Murphy et al., 2013). Five components were validated in each set of EEM spectra, based on split-half analysis (Supplementary Material 2). Three of the five components had similar excitation and emission (Ex-Em) wavelengths between the two sets. The remaining two components of each set were considered different as the difference between excitation and emission exceeded 20 nm. All

obtained components were compared to previously reported components using the OpenFluor fluorescence database.

2.2.4. Size exclusion chromatography HPSEC/UV-fluorescence

High-performance size-exclusion chromatography (HPSEC; LC-2030C 3D Plus, Shimadzu) coupled with photodiode array (Deterium D2 Lamp PDA) and fluorescence (RF-20A xs) detectors was used to determine the molecular size distribution of DOM chromophoric and fluorescent compounds. All the details on the analytical parameters and reference standards used are described in (Boukra et al., 2023). The chromatograms were generated for two UV wavelengths at 254 and 280 nm and three fluorescence Ex-Em wavelengths at 315-443 nm, 280-330 nm and 278-354 nm. These wavelengths are classically used to characterize DOM, humic-like and protein-like components. In order to decompose the size continuum obtained in each HPSEC chromatogram, a Gaussian decomposition was performed on the chromatograms using the toolbox «peakfit.m » (O’Haver, 2020) on MatLab R2020a.

Areas of the identified Gaussian components were summed in order to calculate the total sec absorbance (TSA) from chromatograms obtained with UV detection at 254 nm and 280 nm (Asmala et al., 2021). The total sec fluorescence (TSF) was calculated by summing Gaussian component area from chromatograms obtained with fluorescence detection at three Ex-Em wavelengths (nm): Ex: 315-Em: 443, Ex: 280-Em: 330 and Ex: 278-Em: 354. From the TSA and TSF values and based on a calibration curve of analytical standards (Boukra et al., 2023), three molecular fractions were calculated. Low molecular weight fraction (LMW) with an average molecular weight < 1 kDa, medium molecular weight fraction (MMW) with an average molecular weight ranging from 1 to 10 kDa, and high molecular weight (HMW) with an average molecular weight >10 kDa. The LMW, MMW and HMW fractions represent the relative percentages of respectively small, medium and large molecules detected at the measured wavelengths.

2.2.5. LC-HRMS analysis and data processing

LC-HRMS analysis were acquired on an ACQUITY UPLC H-Class system from Waters (Milford, MA, USA) coupled with a quaternary pump and interfaced to a QToF mass spectrometer (Xevo G2-S Q-TOF; Waters). The molecular features list was generated with xcms (WorkForMetabolomics platform) and CAMERA algorithm was used to annotate adducts and

isotopes (Kuhl et al., 2012; Giacomoni et al., 2015). Detailed analytical parameters, as well as data conversion, processing workflow, and clean-up steps are described in Boukra et al. (in preparation). The final list of molecular compounds was obtained based on filtering steps using quality controls (injections blanks, quality control pool sample, and randomized triplicate injections). This allowed us to retain the less variable compounds present at relatively high intensities in the samples.

After the clean-up steps for HRMS data filtering, the final list of data (i.e. a list of compounds) were processed using two different methods. The first method consisted in setting up global LC-HRMS indicators that allow the integration and comparison of data from several samples analyzed in the same or different analytical sequences. The principle of this first method consisted on splitting the distribution of molecular compounds, on a two-dimensional RT (min) and Mass (Da) plane (Figure 6-4), and into nine equal classes. Each class represents an indicator of a range of molecular masses and retention times of compounds in a given sample. The RT parameter reflects the polarity of the molecular compounds (polarity decreases while RT increases), as reverse phase chromatography is used in the separation stage (Bride et al., 2021). For each sample, the sum of the intensities of the compounds included in each class was calculated, and then normalized by the DOC concentration of the sample as follows:

$$RT_MW = \frac{\sum_{i=1}^{nF} intensity (cps)}{[DOC]}$$

where *RT* and *MW* are the retention time and mass ranges and *F* is the number of molecular compounds contained in this range. For example, the indicator *RT0.6_MW500.1000* corresponds to the sum of intensities of all the molecular compounds that eluted before 6 minutes of acquisition (highly polar compounds), with masses ranging between 500 and 1000 Da.

The second method to analyze the LC-HRMS final list of compounds consisted to identify the specific molecular compounds relevant to mark each type of source (forests/pastures, agriculture, water reservoirs, WWTP raw and treated sewages, and urban runoffs). For this, and to avoid possible inter-sequence analytical variability we proceeded with a single analytical sequence and then with a reduced number of samples (n=3) representative of each type of source. An intersection diagram (venn diagram with more than three intersections)

was performed to identify the specific compounds present in only one type of source, and also the compounds of common interest between two or more sources.

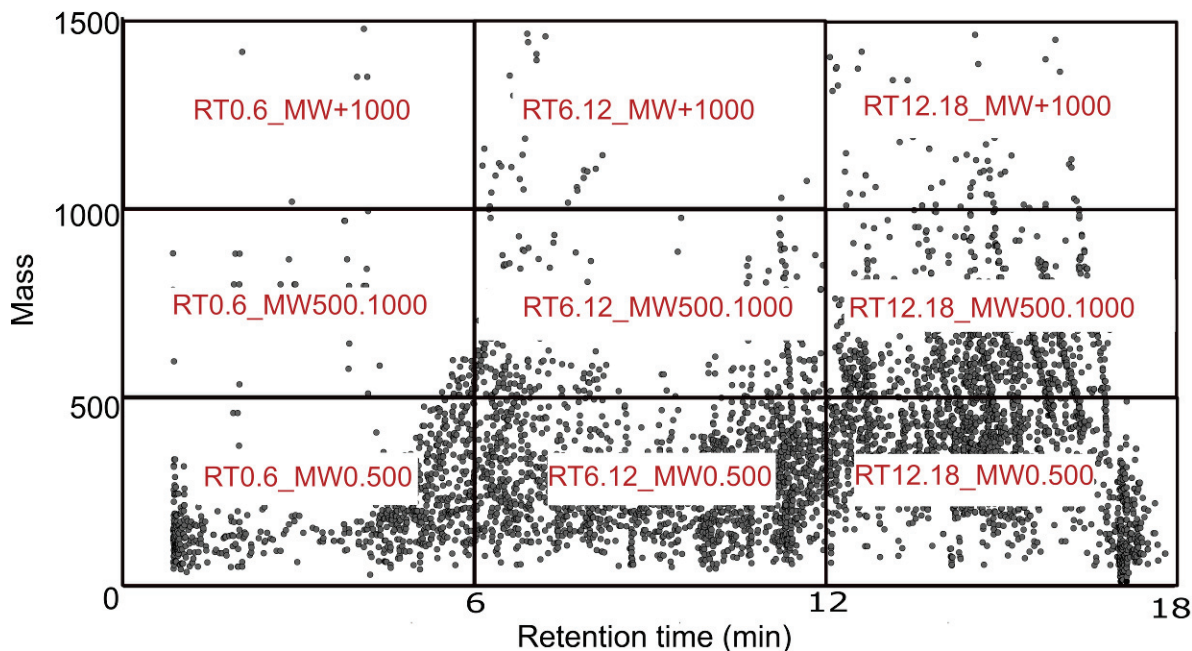


Figure 6-4. Splitting the distribution of molecular compounds according to their retention time and mass. e.g. RT0.6_MW0.500 represents the sum of the intensities of the compounds which elute between 0 and 6 min and whose neutral mass is lower than 500 Da. The sum of the intensities is then normalized by the DOC concentration of the sample.

2.3. Mathematical and statistical analyses

Anova analysis coupled with a Tukey post-hoc test (packages stats and rstatix; R 4.0.2 software) were used to compare indicators values across the types of DOM source. Principal component analysis (PCA; package factoextra, R 4.0.2 software) was performed to integrate all indicators ($n=47$) and to visualize the samples distribution. The identification of molecular compounds of interest was performed using the intersection and non-intersection barplot (package UpSetR, R 4.0.2 software) (Lex et al., 2014)

3. Results and discussion

3.1. DOC content, optical and molecular indicators: a first overview

This section is intended to provide a general description of some of the indicators used to characterize the DOM. The median, maximum and minimum values of all the indicators (n=47) are reported in Table 1. Multivariate data analyses that integrate all the data according the different levels of sources are presented in Sections 3.2, 3.3 and 3.3.

For diffuse sources (runoffs from rural soils) sampled during high and low water periods, the median DOC concentrations were respectively 8.3 and 6.6 mg.L⁻¹. The maximum DOC values observed during high water were 19.7 versus 12 mg.L⁻¹ during low water periods. The minimum values were 5.3 mg.L⁻¹ during high water versus 3.6 mg.L⁻¹ during low water. The range of DOC concentrations (Table 1) are comparable to those observed in studies on stream waters from mixed crop sub-watersheds (Smith et al., 2021; Williams et al., 2010). The increase in carbon concentrations observed between low and high water periods can be explained by a greater contribution of terrigenous DOM sources during high water periods. Indeed, several studies have already reported an increase in carbon during wet periods due to greater terrigenous inputs with leaching from adjacent soils (Wilson et al., 2013). Sources representative of urban runoff had the lowest median DOC concentration (3.05 mg.L⁻¹), with concentrations ranging between 1.8 and 4.7 mg.L⁻¹. Previous studies have reported highly variable concentrations of DOC in similar samples (between 4 and 17 mg.L⁻¹), explained by the rainfall rate and the conditions before the first flush (more or less long dry period favoring organic compounds deposition) (Wise et al., 2019; Gasperi et al., 2022). In rainfalls, DOC concentration are usually low (average DOC concentration at 1 mg.L⁻¹) (Li et al., 2016). The higher values of DOC observed in this study can be explained by a contribution of a multitude of organic compounds during the rain washing of the urban surfaces (e.g. pesticides, flame retardants, polycyclic aromatic hydrocarbons, corrosion inhibitors) (Burant et al., 2018). Finally, the median DOC concentrations in the wastewater inflows and outflows were respectively at 57.4 mg.L⁻¹ and 12.1 mg.L⁻¹. The minimum value of WWTP inflows was 25.6 mg.L⁻¹ and the maximum value was 146 mg.L⁻¹. The minimum value of WWTP outflows was 6.5 mg.L⁻¹ and the maximum value was 41.2 mg.L⁻¹. The DOC concentrations observed in these sources are consistent with those observed in previous studies (Katsoyiannis and Samara, 2007; Goffin et al., 2018; Gasperi et al., 2022). Carbon decay between inflows and

outflows is due to intense degradation of labile organic compounds (e.g. carbohydrates, proteins, lipids) during the multi-step treatment processes. Some persistent dissolved organic compounds (e.g. polycyclic aromatic hydrocarbons PAHs, polychlorinated biphenyls PCBs, organochlorine pesticides OCPs) can nevertheless not be degraded during these wastewaters treatments (Katsoyiannis and Samara, 2005).

The indicators from UV-Visible (E2.E3, E2.E4, E3.E4, SUVA₂₅₄, S1 and SR) and fluorescence (BIX, FI, YFI, HIX) measured in each type of sources (SDL2) are reported in Table 1. Overall, the UV-Visible results highlight a significant higher aromatic content of DOM in samples representative of diffuse sources (runoffs from rural soils) and urban runoff samples ($\sim 2 \text{ L.mg.C}^{-1}.\text{m}^{-1}$) compared to samples representative of WWTP's inflows (median value at $0.95 \text{ L.mg.C}^{-1}.\text{m}^{-1}$) and outflows (median value at $1.69 \text{ L.mg.C}^{-1}.\text{m}^{-1}$). Slope ratio SR, which is negatively correlated to the molecular weight of DOM, is negatively correlated to SUVA₂₅₄ in samples representative of urban runoff, WWTP's inflows and outflows. This indicates that low molecular weight compounds characterizing these samples are less aromatic than high molecular weight ones. Concerning diffuse sources only, the E2.E3, E2.E4 and E3.E4 indicators, which are negatively correlated to DOM molecular weight, humification and terrigenous contribution, reveal DOM with relatively lower molecular weight and lower terrigenous contribution during low water periods compared to high water periods. This may indicate the presence of an additional, more autochthonous source related to the activity of microorganisms in streams (Guarch-Ribot and Butturini, 2016).

The EEM fluorescence indicators BIX, HIX and FI show humified and terrestrial signatures for samples representative of diffuse sources (BIX ~ 0.7 , HIX ~ 7 and FI ~ 1.3) compared to WWTP and urban runoffs samples (BIX ~ 1 , HIX ~ 2 and FI ~ 1.7). Overall, the values of UV-Visible and EEM fluorescence indicators obtained in our study are similar to those observed in studies on the same types of sources (Chen et al., 2021; Park et al., 2009; Pernet-Coudrier et al., 2007; Yuan et al., 2019).

The PARAFAC decomposition resulted in several components obtained on two sets of EEM spectra. The first set included all samples from diffuse sources and urban runoff and the second one, all the samples corresponding to WWTP inflows and outflows. Among the five components obtained in each set, three of them were common to both datasets

(Supplementary Material 2). All the components were compared with the literature using the OpenFluor database with similarity scores above 0.97. Components C1, C2 and C3 who exhibited excitation and emission wavelengths at Ex240-350 Em440 nm, Ex260-Em505 nm, and Ex280-Em330 nm represent the compounds in common between the two datasets. They were found in all samples, i.e. in water samples from soil land use, wastewater and urban runoff samples. The component C1 has been described as terrestrial DOM degraded from lignin (Peleato et al., 2018), typical of soils and terrestrial sources in aquatic environments (Garcia et al., 2018; Yamashita et al., 2021). This component has further been attributed to bioavailable organic matter, microbially derived and/or sensitive to photodegradation (Cawley et al., 2012; Eder et al., 2022). The component C2 have been reported in previous studies as photochemical degraded humic-like compounds (Coulson et al., 2022). Component C3 has been reported as protein-like compounds (Wünsch and Murphy, 2021). Component C4_EN and C5_EN, who exhibited excitation and emission at Ex325-Em395 nm and Ex<240-Em390, were only detected in water samples impacted by land use and urban runoffs. The component C4_EN was previously reported as humic-like compounds obtained by biological activity (Yamashita et al., 2010), while other studies reported component C4_EN as microbial transformed compounds (Du et al., 2021). The component C5_EN has been reported as terrestrial and plant origin compounds having high molecular weight and aromatic content (D'Andrilli et al., 2017). Other studies reported C5_EN as Polycyclic Aromatic Hydrocarbons in industrial or municipal wastewater treatment plants (Carstea et al., 2020). This hypothesis can easily be excluded since this component has also been detected in runoff waters characteristic of forest land use. Finally component C4_ER and C5_ER, who exhibited excitation and emission at Ex<240-300-Em 380 nm and Ex<240-290-Em445, were only detected in wastewater samples. . The component C4_ER, detected specifically in wastewater samples, was attributed to microbial transformed compounds and the component C5_ER was reported as wastewater or nutrient enrichment tracers (Murphy et al., 2011).

Indicators from HPSEC/UV-fluorescence show higher fraction of small molecules (< 1 kDa detected at UV 254 nm) in samples taken from WWTP's inflows, outflows and in urban runoffs (~30 %) compared to runoffs from rural soils (~9 % in high and low water periods samples). This indicates that DOM sources from dense anthropogenic activities, as in urban areas, contain greatest amounts of relatively small molecules. The differences of LMW between

inflows and outflow of WWTPS (~17 % for inflows and ~45 % for outflows), show that during the treatment process, there is a production of low molecular weight compounds (by degradation of high molecular weight compounds). The correlation of Low Molecular Weight fraction (LMW) with BIX and FI indicators indicates the probable biological autochthonous origin of these compounds. For the indicator LMW, at fluorescence Ex-Em wavelengths specific to protein-like compounds (280-330 nm and 275-354 nm), the results show different percentages according to the type of sources considered. Indeed, the samples collected during high water periods had lower LMW (~14 %) than those collected during low water periods (~22 %).

For the global LC-HRMS indicators, which represent the sum of the intensities of molecular compounds included in molecular weight/polarity classes, the results show that compounds with mass higher than 500 Da are either missing or have very low intensities in samples collected during high and low water periods. Classes that represent the non-polar compounds (RT_12-18) had the highest values in urban runoff samples. Finally, polar compounds with mass ranging between 500 Da and 1 kDa were most present in WWTP's inflows and outflows.

Tableau 6-1. Dissolved organic carbon, DOM optical and molecular indicators: median, maximum and minimum values

Techniques	Indicators	High water n=14			Low water n=14			Urban runoff n=12			WWTP inflow n=21			WWTP outflow n=22		
		Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum
DOC	[DOC]	8.30	5.30	19.70	6.60	3.60	12.00	3.05	1.80	4.70	57.40	25.60	146.00	12.05	6.50	41.20
	E2:E3	6.19	5.75	7.00	6.96	5.51	8.24	4.91	4.44	7.12	4.50	3.97	6.93	4.82	3.21	8.33
	E2:E4	21.11	17.92	28.45	27.29	14.99	38.78	14.28	12.05	24.93	8.28	5.29	13.41	12.41	4.93	38.00
	E3:E4	6.00	5.15	7.04	7.15	5.08	8.63	4.47	3.76	5.88	3.35	2.51	5.26	4.35	2.59	12.00
UV-Visible	Suva 254	2.56	1.06	5.32	2.33	1.98	2.84	2.01	1.50	2.39	0.95	0.38	1.61	1.69	0.18	2.50
	S1 (275_295)	0.016	0.014	0.018	0.015	0.014	0.016	0.015	0.014	0.040	0.016	0.011	0.018	0.013	0.008	0.016
	SR	0.85	0.74	1.10	0.72	0.67	0.91	1.00	0.69	1.19	1.60	0.88	2.24	0.98	0.41	1.58
	BIX	0.74	0.68	0.78	0.75	0.66	0.83	0.87	0.79	0.93	1.01	0.75	1.20	1.03	0.89	1.12
EEM fluorescence	FI	1.33	1.25	1.39	1.35	1.27	1.45	1.23	1.07	1.31	1.78	1.62	1.98	1.70	1.57	1.84
	YFI	0.61	0.56	0.71	0.62	0.57	0.68	0.97	0.89	1.17	1.33	0.92	1.70	0.98	0.84	1.14
	HIX	7.04	3.75	10.29	6.43	3.38	9.43	2.76	1.84	3.32	1.42	0.81	2.76	2.86	2.39	4.54
	C1	1.44	0.69	2.77	1.33	0.54	2.52	0.28	0.17	0.57	2.86	1.22	6.36	2.12	1.28	3.77
	C2	0.51	0.24	0.91	0.46	0.19	0.90	0.15	0.13	0.27	4.18	1.71	10.89	1.36	0.54	3.11
	C3	0.26	0.19	0.38	0.30	0.11	0.56	0.35	0.20	0.55	0.64	0.28	0.97	0.55	0.26	0.86
	C4_EN	0.80	0.35	1.54	0.83	0.36	1.37	0.09	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00
	C5_EN	0.36	0.00	1.56	0.18	0.00	0.88	0.50	0.37	1.23	0.00	0.00	0.00	0.00	0.00	0.00
	C4_ER	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.09	1.91	5.38	2.54	1.64	3.82
	C5_ER	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.53	1.98	5.43	2.79	1.54	5.38
HPSEC/UV-fluorescence	TSA_254	216.02	47.63	289.79	189.43	149.25	242.04	154.95	78.80	323.07	27.85	10.59	69.44	103.38	7.66	193.05
	TSA_280	149.78	31.78	202.21	131.21	101.75	170.96	272.07	52.91	533.64	23.43	6.60	51.94	80.62	2.52	162.29
	TSF_315-443	949.64	175.78	1492.02	894.48	620.35	1097.67	354.36	266.34	515.73	243.79	82.78	468.08	818.23	8.09	1348.19
	TSF_275-354	404.01	95.18	671.41	392.53	233.06	471.40	197.03	147.81	330.48	128.11	85.06	285.97	282.99	34.37	452.27
	TSF_280-330	401.74	104.58	666.81	391.91	243.44	490.32	183.87	21.11	268.84	268.48	103.62	607.01	490.91	29.60	1524.81
	HMW_TSA_254	1.31	0.08	50.26	0.63	0.13	3.62	12.96	0.50	64.82	0.00	0.00	4.89	0.86	0.00	15.51
	MMW_TSA_254	87.65	40.90	92.97	89.54	86.28	92.29	43.23	11.09	97.69	63.24	40.20	88.88	70.21	16.50	96.89
	LMW_TSA_254	9.16	5.79	13.91	9.69	5.76	13.59	30.79	0.00	68.04	36.12	10.87	58.09	27.38	3.11	76.45
HMW_TSA_280	1.12	0.16	46.97	0.45	0.00	3.25	27.45	0.00	67.89	0.14	0.00	7.06	0.88	0.00	12.37	

MMW_TSA_280	88.17	43.56	93.53	87.37	82.31	91.80	39.63	0.00	100.00	75.93	44.76	93.50	71.05	27.70	93.76
LMW_TSA_280	8.94	2.00	11.92	11.69	7.41	17.39	6.47	0.00	72.06	23.99	0.00	55.24	28.49	4.91	61.23
HMW_TSF_315-443	0.68	0.08	22.47	0.63	0.00	1.95	0.01	0.00	0.96	0.00	0.00	0.56	0.00	0.00	95.31
MMW_TSF_315-443	85.78	58.70	92.61	82.20	75.31	92.53	70.93	50.95	82.54	53.16	15.36	64.90	49.52	4.69	80.59
LMW_TSF_315-443	13.18	6.73	18.83	16.85	7.47	24.17	28.59	17.46	49.02	46.61	35.00	84.64	48.66	0.00	79.25
HMW_TSF_275-354	1.82	0.00	41.58	0.67	0.03	40.35	14.22	0.00	26.72	38.33	14.70	59.54	34.83	5.42	92.18
MMW_TSF_275-354	82.64	49.01	94.09	76.51	41.34	84.53	45.81	22.70	75.62	41.53	14.07	61.47	16.06	0.00	31.91
LMW_TSF_275-354	14.40	5.41	26.19	21.05	13.77	41.01	38.35	9.32	63.52	19.79	2.03	59.05	49.28	7.82	86.67
HMW_TSF_280-330	0.85	0.00	17.87	0.39	0.00	13.24	32.57	1.50	42.06	33.58	13.93	49.03	34.20	4.54	66.14
MMW_TSF_280-330	81.63	69.10	91.97	75.36	35.36	85.49	35.63	10.26	59.23	51.73	15.54	68.64	20.72	0.09	46.80
LMW_TSF_280-330	14.43	8.03	19.02	23.75	14.49	64.49	34.03	10.97	88.24	15.49	0.00	44.53	43.93	20.06	82.40
RT0-6_MW0-500	5.742	5.643	5.958	5.889	5.765	6.254	5.666	5.441	5.831	5.041	4.699	6.629	5.426	5.292	6.607
RT6-12_MW0-500	5.606	5.325	5.747	5.714	5.457	5.939	6.503	6.043	6.767	5.650	4.774	6.639	5.938	4.844	6.940
RT12-18_MW0-500	6.863	6.565	6.977	6.962	6.677	7.244	6.825	6.591	8.003	5.924	5.222	6.388	6.474	5.583	6.648
RT0-6_MW500-1000	0	0	0	0	0	0	0	0	5	4.002	2.974	6.029	3.202	2.892	5.577
RT6-12_MW500-1000	3.211	2.650	3.574	2.647	2.431	2.997	4.929	4.563	6.531	5.454	4.277	6.487	4.975	3.645	6.102
RT12-18_MW500-1000	4.565	4.138	4.772	4.599	4.367	4.861	7.037	6.803	7.318	4.687	3.889	6.514	4.442	3.617	6.260
RT0-6_MW+1000	0	0	0	0	0	0	0	0	0	2.532	0.000	2.742	1.957	0	2.648
RT6-12_MW+1000	0	0	0	0	0	0	5	5	6	4.242	3.289	5.549	3.603	2.346	5.160
RT12-18_MW+1000	0	0	0	0	0	0	6	6	7	3.919	3.488	5.518	3.267	2.486	5.341

LC-HRMS
(LOG10)

3.2. Evidence of DOM-characteristics groups according to anthropogenic pressures

Integrating all the optical and molecular indicators (listed in Table 1) into a multivariate analysis allowed to investigate DOM combined characteristics and to identify source-specific signatures. The principal component analysis (PCA) build with the 47 indicators measured in the 82 samples shows the variability of the DOM characteristics for each type of sources. The PCA shows clearly the distinction between point sources related to wastewater sewages (inflows and outflows), urban runoffs samples and diffuse sources (SDL2, Figure 6-3). An evolution of the DOM quality according to the anthropogenic pressure (from the lower anthropogenic pressure at the right to the higher anthropogenic pressure at the left, in Figure 6-5-up) is carried mainly by axis 1 (37.1%). The raw wastewater samples are characterized by DOM that contain high contributions of both small and large molecules of protein-type nature, highlighted by the high values of HMW and LMW indicators relative to the wavelengths (EX280_EM330, EX275_EM354). These samples also show high values of SR indicator, indicating the presence of low molecular weight DOM; and their negative correlation with SUVA₂₅₄ highlight a low aromatic and hydrophilic content. These compounds come from domestic, urban and industrial discharges (e.g. pharmaceuticals and personal care products and many other compounds such as plasticizers, surfactants, pesticides, nanoparticles) (Michael-Kordatou et al., 2015).

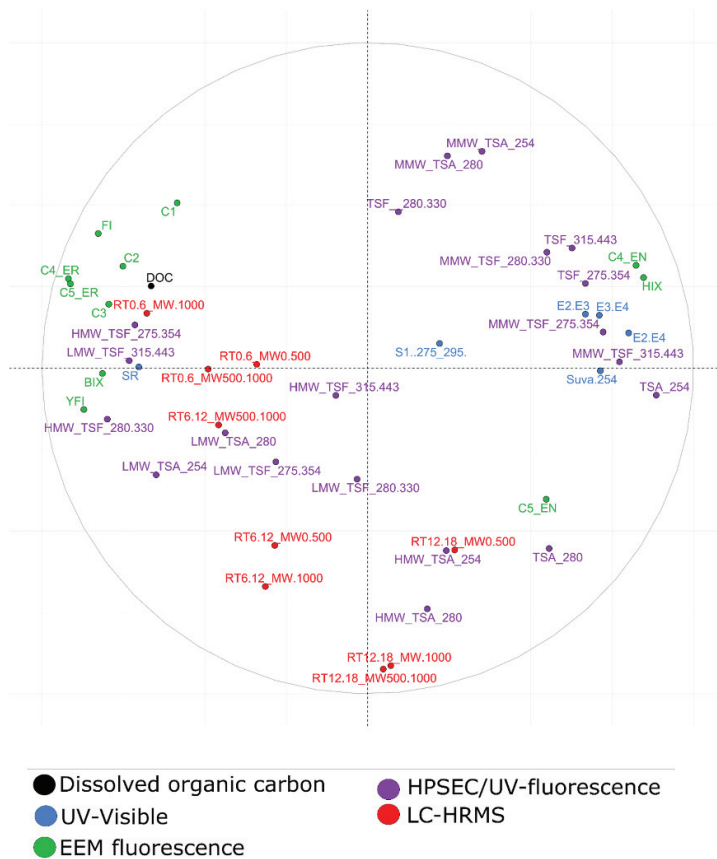
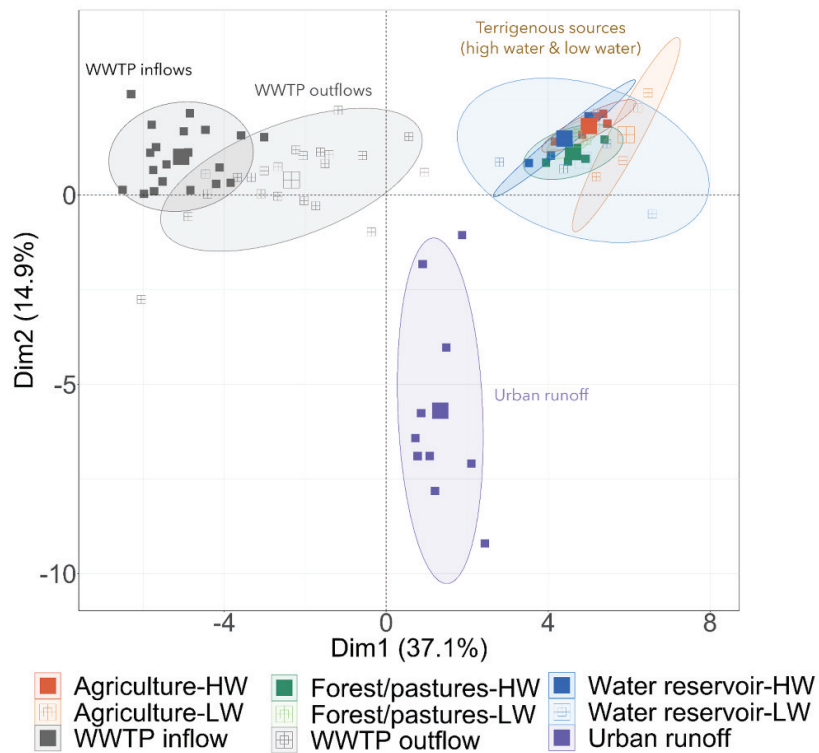


Figure 6-5. First factorial plane of the principal component analysis performed on DOM indicators (UV-Visible, EEM fluorescence, HPSEC/UV-fluorescence and LC-HRMS) measured on wastewaters (WWTP's inflows and outflows), urban runoffs and waters influenced by various land uses and hydrological conditions.

Synthetic compounds (i.e. from industrial synthetic chemistry) that have fluorescence properties are generally detected in the same range of wavelengths as the natural compounds (Tang et al., 2020). The fluorescence indicators BIX and FI reveal the presence of a high biological activity with little terrigenous contribution in these wastewater samples. Finally, the PARAFAC components C4_ER and C5_ER associated mainly with microbial decomposition by-products and anthropogenic compounds confirm these results and seem to be relevant markers of wastewater discharges. Wastewaters after treatment (outflows) seem more dispersed on axis 1 than raw wastewaters (inflows). This can be due to the fact that two types of treatment was studied (activated sludge and planted reed filters). The DOM fingerprints of WWTP outflows tend toward those of the natural water samples. Indeed, a gradient is observed in the evolution of DOM fingerprints, expressed by the relative increase of the molecules size, as well as their aromaticity. This is probably due to the decomposition of labile aliphatic compounds during treatment and then an increase of the proportion of more resistant aromatic compounds (e.g. Goffin et al., 2018).

The urban runoff samples form a group apart from the other types of sources, as illustrated in Figure 6-5-up. DOM contained in these samples are primarily marked by the presence of non-polar compounds. This is highlighted by the HRMS indicators (RT12.18_MW0.500, RT12.18_MW500.1000 and RT12.18_MW1000). The PARAFAC C5_EN component confirms the polycondensed and aromatic character of these samples. These compounds could correspond to a mixture of natural (e.g. degraded DOM from leaves) and synthesized compounds (e.g. PAHs, pesticides, pharmaceuticals) transported during the leaching of the asphalt during rain events. Other organic compounds of relatively high molecular weight (>10 kDa) and absorbing at UV wavelengths 254 and 280 nm, characterize the DOM of urban runoffs. We hypothesis that it could be molecules linked to the domestic animal waste directly transferred to the watercourses (Galfi et al., 2016).

The samples collected in watercourses downstream of land uses (forest/grassland, agricultural and water reservoir) appear difficult to distinguish from each other. Indeed, the differences between WWTP discharges and urban runoffs are much more significant than the differences between the above-mentioned land-use terrigenous sources. Nevertheless, the particularities of these land-use sources lie in the presence of high-aromatic medium-molecular-weight molecules with a high degree of humification. These refractory compounds like lignins and

tannins (Seidel et al., 2015) originate mainly from terrigenous inputs with signatures very close in comparison to the other non-terrigenous sources studied in this paper.

3.3. Influence of land uses and hydrological conditions on DOM characteristics in the watercourses

To better understand the impact of diffuse sources on DOM characteristics, samples collected in watercourses at the outlet of homogenous land uses during high and low water periods (n=28) were further studied without considering urban runoffs nor WWTP in- and outflows, and using the combined set of UV-Visible, EEM fluorescence, HPSEC/UV-fluorescence, and LC-HRMS indicators (Figure 6-6). For the samples taken during high water periods, two first distinct groups are observed: a first group of samples taken downstream of forests and pastures (FP1_HW to FP5_HW) and a second group of samples taken downstream of agricultural land uses (AG1_HW to AG5_HW). The samples collected in the high water reservoirs (WR1_HW to WR4_HW) form a third group with a very large and heterogeneous signature including those of the first two groups. However, an influence of land uses is also observed on the quality of DOM in these water reservoirs. Indeed, sample WR1_HW, close to the forest/pasture group in Figure 6-6, was taken from the same sub-watershed as the forest and pasture samples; same for samples WR2_HW and WR3_H3 taken from the same sub-watersheds as the agricultural samples. Sample WR4_HW, taken downstream of the basin, is far from the other sample groups in Figure 6-6; probably because a multiplicity of various inputs in this sample (presence of an urban area, as well as an industrial pharmaceutical group nearby).

When considering the samples collected during low water periods, the differences between land uses are less pronounced (Figure 6-6). Indeed, the forests/pastures samples group is closer to the agricultural samples group and the dispersion inside each group is lower. The water reservoir samples show a more heterogeneous signature (with an higher dispersion inside this group), as during the high water period. As for high water periods, these water reservoir samples are influenced by land use.

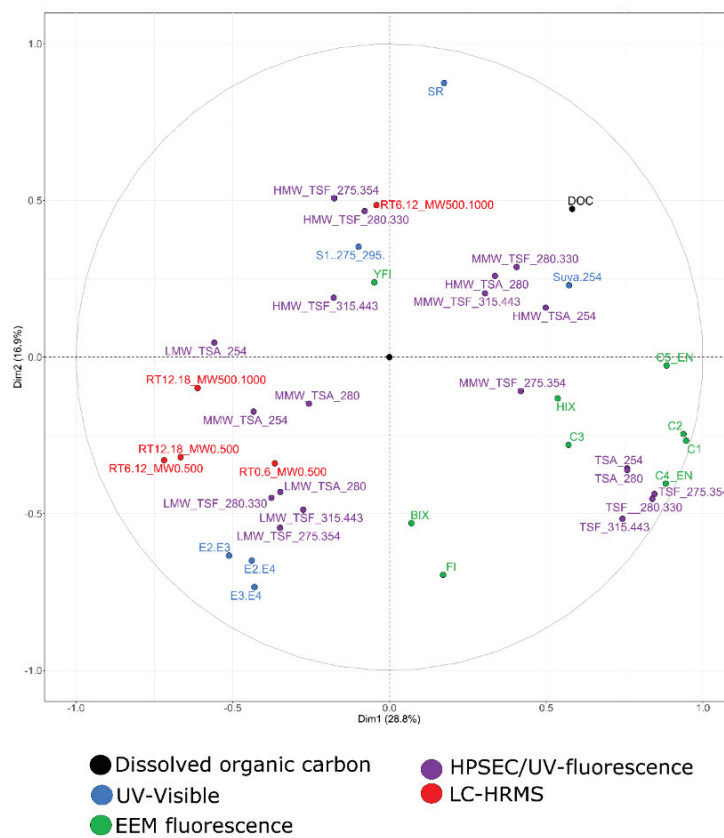
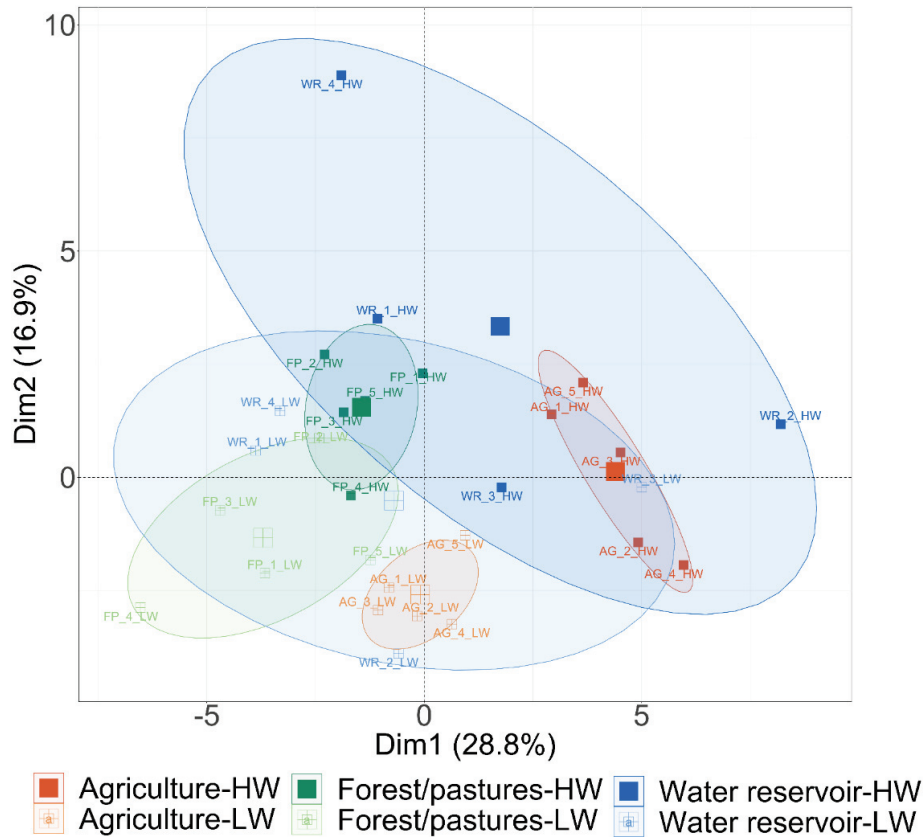


Figure 6-6. First factorial plane of the principal component analysis performed on DOM indicators measured on diffuse sources influenced by various land uses (forests/pastures, agriculture and water reservoirs) and hydrological conditions (high and low water flows

The difference between DOM in the watercourses collected during high and low water periods, is mainly due to an higher contribution of medium/large and strongly aromatic molecules during high water periods. Indeed, the HMW and MMW indicators (> 1kDa), as well as the SUVA₂₅₄ indicator show higher values for samples taken in high waters. This result can be explained by a higher contribution of a terrigenous DOM having undergone degradation processes (mainly labile non-aromatic groups). The SR indicator, inversely correlated to the molecular weight, reveals a DOM with low molecular weight compounds during high water periods, which is in contradiction with information from the UV-Visible indicator E₂. E₃ and the HMW (high molecular weight; >10 kDa). This demonstrates the limit of interpretation of certain optical indicators like SR. Regarding the water samples taken during low water periods, DOM is characterized by an higher contribution of small molecules (<1 kDa) of amino acids and protein types. The high values of the absorbance ratios (E₂.E₃, E₃.E₄ and E₂.E₄) confirm the presence of low molecular weight compounds with little humification and low aromatic content in these samples compared to those from high water periods. The LC-HRMS indicators show compounds with a wide range of polarity during low water periods and the BIX and FI indicators reveal the influence of a high microbial activity. To conclude, the characteristics of DOM sampled during low water periods result mainly from autochthonous contributions with also a detectable influence of the surrounding terrigenous inputs.

3.4. Evolution of DOM characteristics across different wastewater treatment

DOM samples representative of the anthropogenic discharge sources (i.e. wastewater sewages and urban runoffs) were also studied according to SDL3 level. The PCA analysis (Figure 6-7) allows to discuss DOM signatures for the 5 groups of samples. The first two groups, composed of untreated wastewaters from large urban communities (domestic and industrial discharges) and small communities (mainly domestic discharges), have very similar DOM fingerprints. Whereas, the two other groups on treated wastewaters, show fingerprints varying according to the type of wastewaters treatment. Indeed, the samples taken at the outlet of sludge activated WWTP are closer to the untreated water than to the outlet of reed filter WWTP. Finally, the last group composed of urban runoff samples is gathered far from the other groups of samples, revealing very specific DOM fingerprints. DOM present in raw wastewaters (from small and large communities) is mainly characterized by a mixture of low molecular weight, low aromatic and very polar compounds. The component C2 strongly visible

in these waters indicates the presence of degraded humic compounds (Coulson et al., 2022) and the indicator YFI reflects an high microbial activity, specific of wastewaters (Heo et al., 2016). At the outlets of sludge activated WWTP, the DOM has an high molecular weight of humic-type compounds, contrary to the outlets of planted reed filters (PRF) where DOM is more diversified in terms of sizes (LMW and MMW) and natures (protein type, amino acid type and humic type). The degree of humification and aromaticity of DOM at the PRF outlet is close to that of waters impacted by terrigenous land uses inputs. Finally, urban runoffs are marked by the presence of DOM consisting of both small non-polar and aromatic molecules, as well as large molecules of a protein nature. As previously discussed, these heterogeneous organic compounds could result from the leaching of roads and artificial surfaces (e.g. PAH's, synthetic biopolymers etc.).

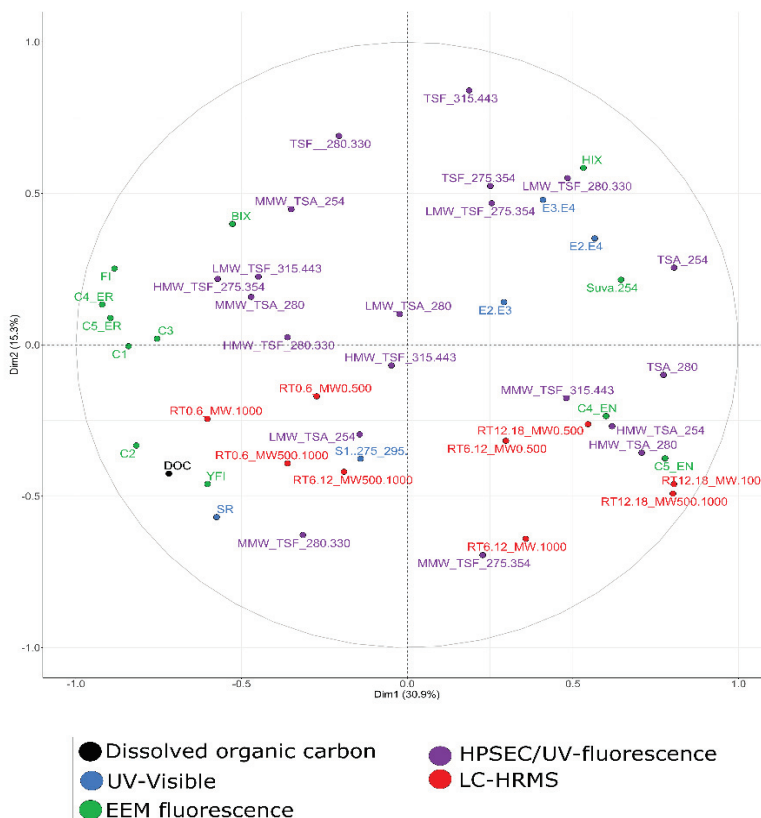
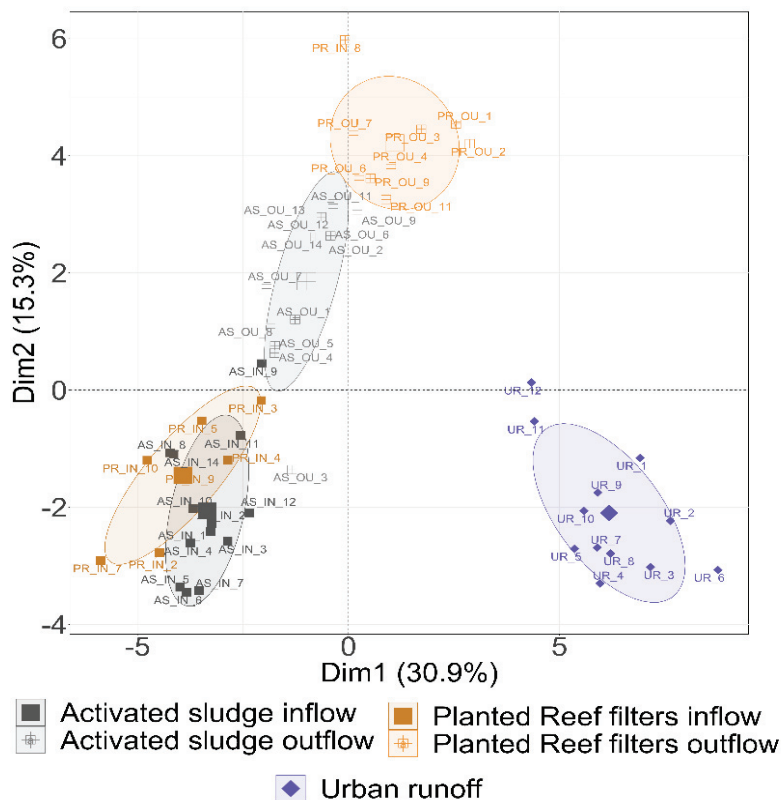


Figure 6-7. First factorial plane of the principal component analysis performed on DOM indicators (UV-Visible, EEM fluorescence, HPSEC/UV-fluorescence and LC-HRMS) measured on point high-anthropogenic sources (PRF and AS WWTP's inflows and outflows and urban runoffs).

3.5. Identification of molecular markers specific to DOM sources

The LC-HRMS global indicators presented in this study (Section 2.3.6) have shown relevance in the characterization and discrimination of DOM sources. Indeed, the nine indicators were integrated with the other optical indicators to provide information on the molecular weight distribution, complementary to the HPSEC/UV-fluorescence indicators. LC-HRMS indicators give also information on the polarity of organic compounds. These LC-HRMS global indicators can be used for the identification of molecular compounds of interest specifically present in the predefined types of DOM sources. For this purpose, we used a single analytical LC-HRMS sequence in which three samples of each source type were injected in randomized triplicates. This was done in order to minimize the variability related to the instrumentation. After the data cleaning steps with the use of quality controls (injection blanks, Quality controls pool and injection triplicates; Boukra et al. in preparation), intersection diagrams were performed based on the presence/absence of molecular compounds. We used the following rule: If the intensity of a compound exceeded the intensity of the HRMS threshold set at 10000 counts per second, then, the compound was considered present in the sample, otherwise it was considered absent.

Figures 6-8.A, 6-8.B and 6-8.C represent a barplot of intersections (molecular compounds in common between two or more sources) and non-intersections (source specific molecular compounds) between DOM sources impacted by : i) land use and urban runoff, ii) land use and PRF WWTP's inflows and outflows and iii) land use and AS WWTP's inflows and outflows. The result of Figure 6-8.A shows that water samples collected downstream of forests/pastures DOM sources contain 16 exclusive compounds, compared to agricultural and water reservoir DOM sources, which contain 32 and 51 exclusive compounds respectively. Water reservoirs, which signatures were very heterogeneous are probably impacted by both terrigenous contribution but also autochthonous inputs through microbial activities. The samples representative of urban runoffs show a higher number of exclusive compounds (n=456), originating from urban surfaces (buildings, traffic roads, public and individual gardens ...). Previous studies have highlighted the presence of PCBs, PAHs, phthalates, plasticizers, pesticides in such urban runoffs. (Gasperi et al., 2022; Wicke et al., 2021). Figure 6-8.A also shows the presence of 252 compounds common to the waters derived from forest-pasture-agricultural and reservoirs land uses. These compounds could be good markers of less

anthropogenic DOM, in rural area, as they are not present in waters leached from urban area. The largest number of compounds (n= 802) remain those in common between all DOM sources (ubiquitous in all these types of waters). This was the subject of a study that showed that the development of DOM is likely the result of a series of breakdowns or synthetic processes that lead in the creation of an universal background (Zark and Dittmar, 2018).

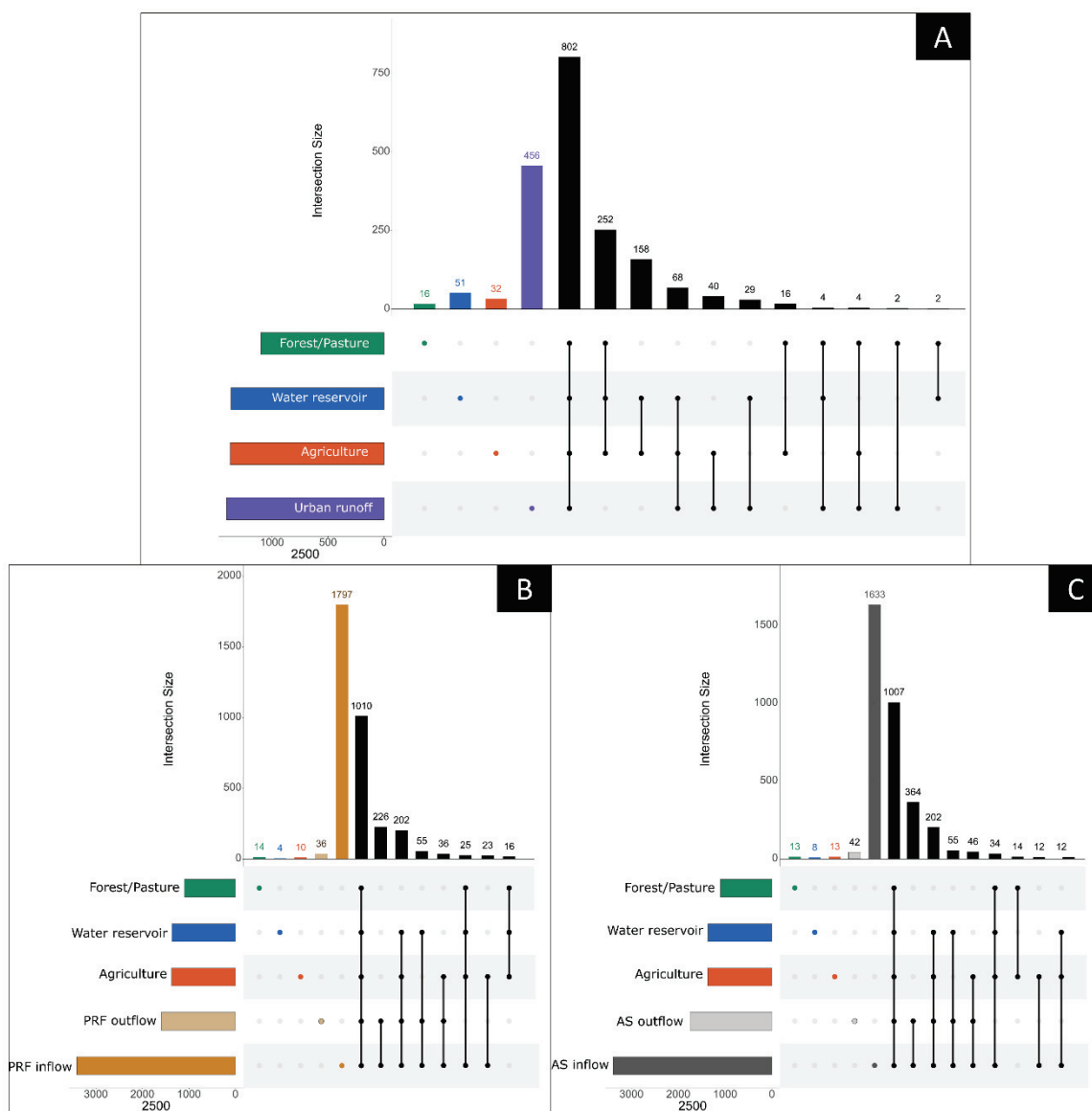


Figure 6-8. Barplot highlighting the number of intersections and non-intersections (specific or common molecular compounds) between A: DOM sources influenced by various land uses with DOM sources from urban runoff. B: DOM sources influenced by various land uses with DOM sources from planted reed filter WWTP. C: DOM sources influenced by various land uses with DOM sources from activated sludge WWTP.

Figures 6-8.B and 7.C show quite similar distribution of intersections across DOM sources. The greatest number of compounds was detected in the WWTP's inflows with no significant difference considering the size of the upstream communities. Indeed, 1797 compounds were detected specifically in raw waters from small communities and 1633 in raw waters from large urban and industrial communities. At the outlet of AS and PRF WWTP, the number of specific compounds is 36 and 42, respectively. These compounds, which are specific to treated wastewater, should be degradation products of synthetic molecules (e.g. from pesticides, pharmaceuticals, personal care products...) and could represent good indicators of WWTP discharges to streams. The intersection diagrams also show the presence of a significant number of compounds present in both inflows and outflows of the WWTP (226 compounds for the PRF and 364 compounds for the AS). These compounds are refractory to the PRF and AS treatments. Several refractory compounds are reported in the literature, such as phthalates, bisphenol A, per- and polyfluoroalkyl substances (PFAS) and polychlorinated biphenyls (PCBs) for their non-degradable character and their toxicity towards bacterial growth in WWTP (Cretin and Lesage, 2015). A large number of molecular compounds are common to all types of sources (an highly complex pool of non-specific organic compounds; Figure 6-8.B and 6.C). The final comparison between highly anthropogenic point sources (treated and untreated wastewater and urban runoffs; Figure 6-9) confirms that the majority of compounds observed in raw wastewaters are common to discharges from small and large communities. This result confirms the PCA result (Figure 6-7) which showed a group of raw wastewater samples that were not discriminated in terms of signature. The intersection plot in Figure 6-9 also shows that there are 3 times more compounds in common between raw wastewater (AS and PRF combined) and AS outflows ($n=127$) than PRF outflows ($n=38$). This result suggests that degradation processes are more or less efficient to degrade these organic compounds present in the raw waters.

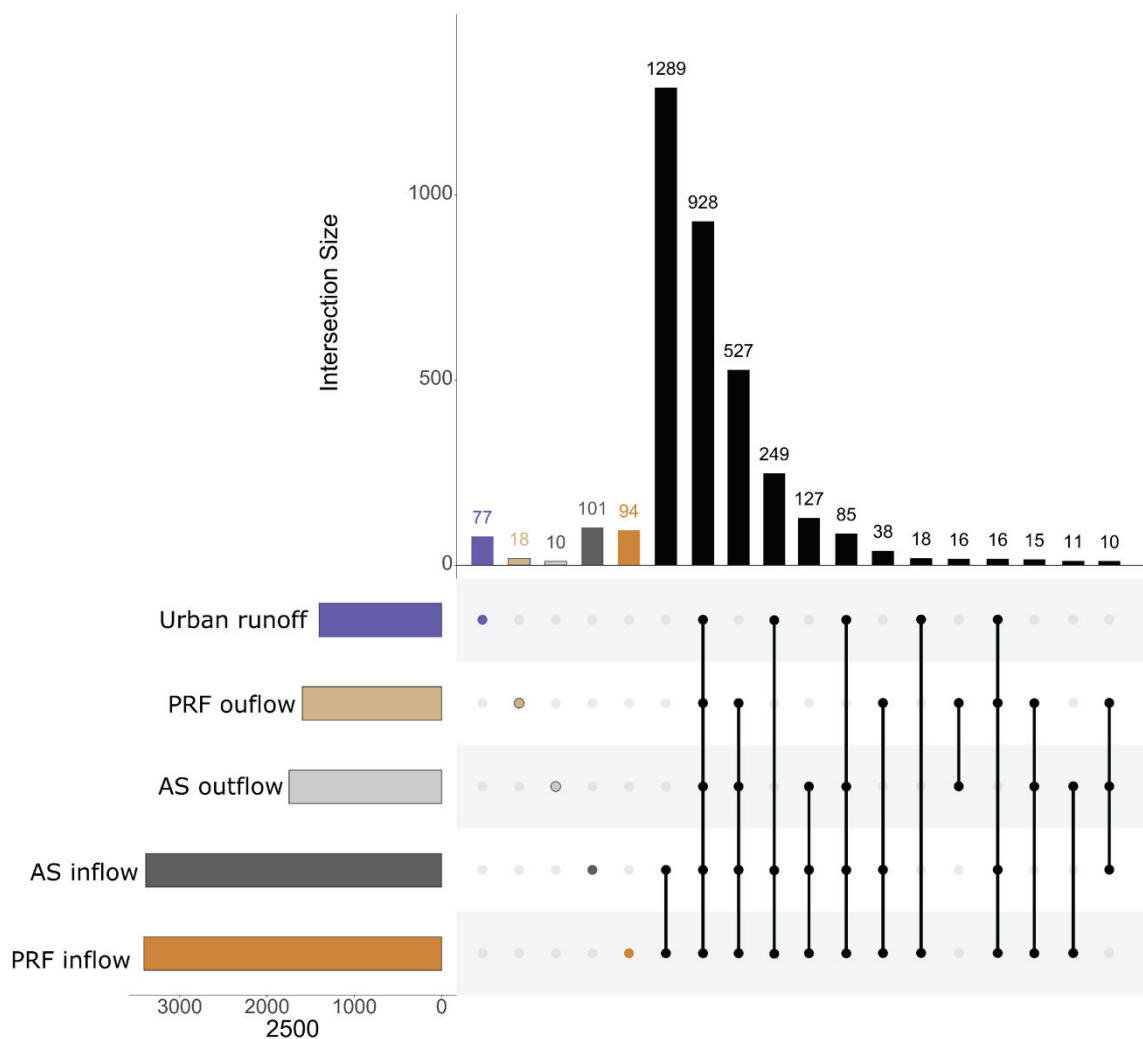


Figure 6-9. Barplot highlighting the number of intersections and non-intersections (specific or common molecular compounds) between DOM sources from urban runoff, DOM sources from planted reed filter WWTP and DOM sources from activated sludge WWTP

Overall, the exploitation of these LC-HRMS data, through the intersection and non-intersection plots, revealed the presence of source-specific compounds as well as compounds of interest present in two or more sources (regardless of the multiplicity of inputs, as well as the numerous environmental processes that can alter the quality of the DOM). Specific molecular compounds (e.g. to forest, agricultural, urban runoff) may be relevant source markers. Compounds common to several sources may also discriminate types of DOM sources on a more global spatial scale (e.g. difference between low anthropogenic impacted waters and wastewaters). From the point of view of tracing sources of pollutions in downstream rivers, the establishment of a database compiling specific compounds could make it possible to better trace the sources of discharges on a watershed scale.

4. Conclusion

In the context of improving the monitoring of anthropogenic discharges into watercourses, the objective of this study was to determine whether DOM could be a relevant tracer of various anthropogenic inputs at a watershed scale. The overall optical and molecular properties of DOM revealed an aromatic character with the presence of high molecular weight compounds in the rural diffuse sources (in forest/pasture and agricultural areas) in contrast to the more anthropogenic impacted waters (as wastewaters and urban runoffs). In the diffuse rural sources, DOM was mainly influenced by the hydrological conditions, with a terrigenous contribution highlighted by an impact of land uses during high water periods. During low water periods, DOM mainly resulted from autochthonous contributions with a lower influence of the surrounding terrigenous inputs. DOM sampled in WWTP's inflows showed no clear differences according to the upstream discharges, but differences were observed in the outflows depending on the type of treatment applied in the WWTP. In PRF outflows, DOM fingerprint was closer to that in natural watercourses compared to AS outflows. Finally, urban runoff sources exhibited very specific signatures with a DOM consisting mainly of highly aromatic and non-polar low molecular weight compounds.

The processing of non-targeted LC-HRMS data has allowed us to establish nine global indicators based on molecular weight and polarity classes. These indicators showed relevance to characterize DOM from typical sources. A more detailed exploitation of LC-HRMS data revealed the presence of a large number of molecular compounds with no DOM source-specificity (universal molecular background), but also source-specific, which number depended on the type of DOM source. These compounds could constitute good markers of anthropogenic pollution sources, as they provide additional information to the optical indicators. This study suggests that DOM in rivers could be used as a tracer of water origin and associated pollutants by differentiating between inputs from natural areas, agricultural areas, treated and untreated WWTP discharges and urban streams. A further approach would be to use some of these tracers (i.e. additive and conservative ones) in water mixing models.

5. Acknowledgements

This study was part of the “CHYPSTER” project supported by the French National Research Agency (ANR-21-CE34-0013-01) and “IDESOC” project granted by the ZABR – Rhone Basin LTSER within the Water Agency RMC – RB LTSER funding agreement, and performed within the framework of the EUR H2O’Lyon (ANR-17-EURE-0018) of Université de Lyon (UdL), within the program “Investissements d’Avenir” operated by the French National Research Agency (ANR). We would like to thank our colleagues Christelle Margoum and Kevin Rocco for the discussions on the non-targeted screening analysis and data processing. We are also grateful to our colleagues Corinne Brosse, Loïc Richard, Alexandra Gruat and Mahaut Sourzac for their assistance in field sampling, samples preparation and chemical analysis. Finally, we would like to thank our colleague Olivier Garcia and Guthi Raja Sekhar from REVERSAL research unit for their assistance in sampling wastewater.

6. References

- Aiken, G.R., 2014. 1.11 - Dissolved Organic Matter in Aquatic Systems, in: Ahuja, S. (Ed.), *Comprehensive Water Quality and Purification*. Elsevier, Waltham, pp. 205–220. <https://doi.org/10.1016/B978-0-12-382182-9.00014-1>
- Asmala, E., Massicotte, P., Carstensen, J., 2021. Identification of dissolved organic matter size components in freshwater and marine environments. *Limnol. Oceanogr.* 66, 1381–1393. <https://doi.org/10.1002/lno.11692>
- Boukra, A., Masson, M., Brosse, C., Sourzac, M., Parlanti, E., Miège, C., 2023. Sampling terrigenous diffuse sources in watercourse: Influence of land use and hydrological conditions on dissolved organic matter characteristics. *Sci. Total Environ.* 162104. <https://doi.org/10.1016/j.scitotenv.2023.162104>
- Braud, I., Breil, P., Thollet, F., Lagouy, M., Branger, F., Jacqueminet, C., Kermadi, S., Michel, K., 2013. Evidence of the impact of urbanization on the hydrological regime of a medium-sized periurban catchment in France. *J. Hydrol.* 485, 5–23. <https://doi.org/10.1016/j.jhydrol.2012.04.049>
- Bride, E., Heinisch, S., Bonnefille, B., Guillemain, C., Margoum, C., 2021. Suspect screening of environmental contaminants by UHPLC-HRMS and transposable Quantitative Structure-Retention Relationship modelling. *J. Hazard. Mater.* 409, 124652. <https://doi.org/10.1016/j.jhazmat.2020.124652>
- Burant, A., Selbig, W., Furlong, E.T., Higgins, C.P., 2018. Trace organic contaminants in urban runoff: Associations with urban land-use. *Environ. Pollut. Barking Essex 1987* 242, 2068–2077. <https://doi.org/10.1016/j.envpol.2018.06.066>
- Butturini, A., Herzsprung, P., Lechtenfeld, O.J., Alcorlo, P., Benaiges-Fernandez, R., Berlanga, M., Boadella, J., Freixinos Campillo, Z., Gomez, R.M., Sanchez-Montoya, M.M., Urmeneta, J., Romani, A.M., 2022. Origin, accumulation and fate of dissolved organic matter in an extreme hypersaline shallow lake. *Water Res.* 221, 118727. <https://doi.org/10.1016/j.watres.2022.118727>
- Carstea, E.M., Popa, C.L., Baker, A., Bridgeman, J., 2020. In situ fluorescence measurements of dissolved organic matter: A review. *Sci. Total Environ.* 699, 134361. <https://doi.org/10.1016/j.scitotenv.2019.134361>
- Cawley, K.M., Ding, Y., Fourqurean, J., Jaffé, R., 2012. Characterising the sources and fate of dissolved organic matter in Shark Bay, Australia: a preliminary study using optical properties and stable carbon isotopes. *Mar. Freshw. Res.* 63, 1098. <https://doi.org/10.1071/MF12028>
- Chen, S., Du, Y., Das, P., Lamore, A.F., Dimova, N.T., Elliott, M., Broadbent, E.N., Roebuck, J.A., Jaffé, R., Lu, Y., 2021. Agricultural land use changes stream dissolved organic matter via altering soil inputs to streams. *Sci. Total Environ.* 796, 148968. <https://doi.org/10.1016/j.scitotenv.2021.148968>
- Choi, Y., Lee, J.-H., Kim, K., Mun, H., Park, N., Jeon, J., 2021. Identification, quantification, and prioritization of new emerging pollutants in domestic and industrial effluents, Korea: Application of LC-HRMS based suspect and non-target screening. *J. Hazard. Mater.* 402, 123706. <https://doi.org/10.1016/j.jhazmat.2020.123706>
- Claret, F., Schäfer, T., Bauer, A., Buckau, G., 2003. Generation of humic and fulvic acid from Callovo-Oxfordian clay under high alkaline conditions. *Sci. Total Environ.* 317, 189–200. [https://doi.org/10.1016/S0048-9697\(03\)00337-1](https://doi.org/10.1016/S0048-9697(03)00337-1)
- Coulson, L.E., Weigelhofer, G., Gill, S., Hein, T., Griebler, C., Schelker, J., 2022. Small rain events during drought alter sediment dissolved organic carbon leaching and respiration in intermittent stream sediments. *Biogeochemistry* 159, 159–178. <https://doi.org/10.1007/s10533-022-00919-7>
- Cretin, M., Lesage, G., 2015. Biorefractory Pollutants, in: Drioli, E., Giorno, L. (Eds.), *Encyclopedia of Membranes*. Springer, Berlin, Heidelberg, pp. 1–1. https://doi.org/10.1007/978-3-642-40872-4_2041-1

- da Silva, M.P., Blaurock, K., Beudert, B., Fleckenstein, J.H., Hopp, L., Peiffer, S., Reemtsma, T., Lechtenfeld, O.J., 2021. Delineating Source Contributions to Stream Dissolved Organic Matter Composition Under Baseflow Conditions in Forested Headwater Catchments. *J. Geophys. Res. Biogeosciences* 126, e2021JG006425. <https://doi.org/10.1029/2021JG006425>
- D'Andrilli, J., Foreman, C.M., Sigl, M., Priscu, J.C., McConnell, J.R., 2017. A 21 000-year record of fluorescent organic matter markers in the WAIS Divide ice core. *Clim. Past* 13, 533–544. <https://doi.org/10.5194/cp-13-533-2017>
- Derrien, M., Brogi, S.R., Gonçalves-Araujo, R., 2019. Characterization of aquatic organic matter: Assessment, perspectives and research priorities. *Water Res.* 163, 114908. <https://doi.org/10.1016/j.watres.2019.114908>
- Du, Y., Lu, Y., Roebuck, J.A., Liu, D., Chen, F., Zeng, Q., Xiao, K., He, H., Liu, Z., Zhang, Y., Jaffé, R., 2021. Direct versus indirect effects of human activities on dissolved organic matter in highly impacted lakes. *Sci. Total Environ.* 752, 141839. <https://doi.org/10.1016/j.scitotenv.2020.141839>
- Eder, A., Weigelhofer, G., Pucher, M., Tiefenbacher, A., Strauss, P., Brandl, M., Blöschl, G., 2022. Pathways and composition of dissolved organic carbon in a small agricultural catchment during base flow conditions. *Ecohydrol. Hydrobiol.* 22, 96–112. <https://doi.org/10.1016/j.ecohyd.2021.07.012>
- Ferretto, N., Tedetti, M., Guigue, C., Mounier, S., Redon, R., Goutx, M., 2014. Identification and quantification of known polycyclic aromatic hydrocarbons and pesticides in complex mixtures using fluorescence excitation–emission matrices and parallel factor analysis. *Chemosphere* 107, 344–353. <https://doi.org/10.1016/j.chemosphere.2013.12.087>
- Galfi, H., Österlund, H., Marsalek, J., Viklander, M., 2016. Indicator bacteria and associated water quality constituents in stormwater and snowmelt from four urban catchments. *J. Hydrol.* 539, 125–140. <https://doi.org/10.1016/j.jhydrol.2016.05.006>
- Garcia, R.D., Diéguez, M. del C., Gereá, M., Garcia, P.E., Reissig, M., 2018. Characterisation and reactivity continuum of dissolved organic matter in forested headwater catchments of Andean Patagonia. *Freshw. Biol.* 63, 1049–1062. <https://doi.org/10.1111/fwb.13114>
- Gasperi, J., Le Roux, J., Deshayes, S., Ayrault, S., Bordier, L., Boudahmane, L., Budzinski, H., Caupos, E., Caubrière, N., Flanagan, K., Guillon, M., Huynh, N., Labadie, P., Meffray, L., Neveu, P., Partibane, C., Paupardin, J., Saad, M., Varnede, L., Gromaire, M.-C., 2022. Micropollutants in Urban Runoff from Traffic Areas: Target and Non-Target Screening on Four Contrasted Sites. *Water* 14, 394. <https://doi.org/10.3390/w14030394>
- Giacomoni, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., Duperier, C., Tremblay-Franco, M., Martin, J.-F., Jacob, D., Goullitquer, S., Thévenot, E.A., Caron, C., 2015. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinforma. Oxf. Engl.* 31, 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Goffin, A., Guérin, S., Rocher, V., Varrault, G., 2018. Towards a better control of the wastewater treatment process: excitation-emission matrix fluorescence spectroscopy of dissolved organic matter as a predictive tool of soluble BOD5 in influents of six Parisian wastewater treatment plants. *Environ. Sci. Pollut. Res.* 25, 8765–8776. <https://doi.org/10.1007/s11356-018-1205-1>
- Gómez-Regalado, M. del C., Martín, J., Santos, J.L., Aparicio, I., Alonso, E., Zafra-Gómez, A., 2023. Bioaccumulation/bioconcentration of pharmaceutical active compounds in aquatic organisms: Assessment and factors database. *Sci. Total Environ.* 861, 160638. <https://doi.org/10.1016/j.scitotenv.2022.160638>
- Grgić, I., Cetinić, K.A., Karačić, Z., Previšić, A., Rožman, M., 2023. Fate and effects of microplastics in combination with pharmaceuticals and endocrine disruptors in freshwaters: Insights from a microcosm experiment. *Sci. Total Environ.* 859, 160387. <https://doi.org/10.1016/j.scitotenv.2022.160387>

- Guarch-Ribot, A., Butturini, A., 2016. Hydrological conditions regulate dissolved organic matter quality in an intermittent headwater stream. From drought to storm analysis. *Sci. Total Environ.* 571, 1358–1369. <https://doi.org/10.1016/j.scitotenv.2016.07.060>
- Hassoun, H., Lamhasni, T., Foudeil, S., El Bakkali, A., Ait Lyazidi, S., Haddad, M., Choukrad, M., Hnach, M., 2017. Total Fluorescence Fingerprinting of Pesticides: A Reliable Approach for Continuous Monitoring of Soils and Waters. *J. Fluoresc.* 27, 1633–1642. <https://doi.org/10.1007/s10895-017-2100-8>
- Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnol. Oceanogr.* 53, 955–969. <https://doi.org/10.4319/lo.2008.53.3.0955>
- Heo, J., Yoon, Y., Kim, D.-H., Lee, H., Lee, D., Her, N., 2016. A new fluorescence index with a fluorescence excitation-emission matrix for dissolved organic matter (DOM) characterization. *Desalination Water Treat.* 57, 20270–20282. <https://doi.org/10.1080/19443994.2015.1110719>
- Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J.-M., Parlanti, E., 2009. Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Org. Geochem.* 40, 706–719. <https://doi.org/10.1016/j.orggeochem.2009.03.002>
- Jaffé, R., Boyer, J.N., Lu, X., Maie, N., Yang, C., Scully, N.M., Mock, S., 2004. Source characterization of dissolved organic matter in a subtropical mangrove-dominated estuary by fluorescence analysis. *Mar. Chem.* 84, 195–210. <https://doi.org/10.1016/j.marchem.2003.08.001>
- 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. *Environ. Sci. Pollut. Res. Int.* 14, 284–292. <https://doi.org/10.1065/espr2006.05.302>
- Katsoyiannis, A., Samara, C., 2005. Persistent organic pollutants (POPs) in the conventional activated sludge treatment process: fate and mass balance. *Environ. Res.* 97, 245–257. <https://doi.org/10.1016/j.envres.2004.09.001>
- Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T.R., Neumann, S., 2012. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* 84, 283–289. <https://doi.org/10.1021/ac202450g>
- Lapworth, D.J., Baran, N., Stuart, M.E., Ward, R.S., 2012. Emerging organic contaminants in groundwater: A review of sources, fate and occurrence. *Environ. Pollut.* 163, 287–303. <https://doi.org/10.1016/j.envpol.2011.12.034>
- Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., Pfister, H., 2014. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- Li, C., Yan, F., Kang, S., Chen, P., Qu, B., Hu, Z., Sillanpää, M., 2016. Concentration, sources, and flux of dissolved organic carbon of precipitation at Lhasa city, the Tibetan Plateau. *Environ. Sci. Pollut. Res.* 23, 12915–12921. <https://doi.org/10.1007/s11356-016-6455-1>
- Li, P., Hur, J., 2017. Utilization of UV-Vis spectroscopy and related data analyses for dissolved organic matter (DOM) studies: A review. *Crit. Rev. Environ. Sci. Technol.* 47. <https://doi.org/10.1080/10643389.2017.1309186>
- Mcknight, D., Boyer, E., Westerhoff, P., Doran, P., Kulbe, T., Andersen, D., 2001. Spectrofluorometric Characterization of Dissolved Organic Matter for Indication of Precursor Organic Material and Aromaticity. *Limnol. Oceanogr.* 46, 38–48. <https://doi.org/10.4319/lo.2001.46.1.0038>
- Michael-Kordatou, I., Michael, C., Duan, X., He, X., Dionysiou, D.D., Mills, M.A., Fatta-Kassinos, D., 2015. Dissolved effluent organic matter: Characteristics and potential implications in wastewater treatment and reuse applications. *Water Res.* 77, 213–248. <https://doi.org/10.1016/j.watres.2015.03.011>
- Morling, K., Herzsprung, P., Kamjunke, N., 2017. Discharge determines production of, decomposition of and quality changes in dissolved organic carbon in pre-dams of drinking water reservoirs. *Sci. Total Environ.* 577, 329–339. <https://doi.org/10.1016/j.scitotenv.2016.10.192>

- Murphy, K.R., Hambly, A., Singh, S., Henderson, R.K., Baker, A., Stuetz, R., Khan, S.J., 2011. Organic Matter Fluorescence in Municipal Water Recycling Schemes: Toward a Unified PARAFAC Model. *Environ. Sci. Technol.* 45, 2909–2916. <https://doi.org/10.1021/es103015e>
- Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Anal. Methods* 5, 6557–6566. <https://doi.org/10.1039/C3AY41160E>
- Murphy, K.R., Stedmon, C.A., Waite, T.D., Ruiz, G.M., 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Mar. Chem.* 108, 40–58. <https://doi.org/10.1016/j.marchem.2007.10.003>
- Park, M.-H., Lee, T.-H., Lee, B.-M., Hur, J., Park, D.-H., 2009. Spectroscopic and Chromatographic Characterization of Wastewater Organic Matter from a Biological Treatment Plant. *Sensors* 10, 254–265. <https://doi.org/10.3390/s100100254>
- Peleato, N.M., Legge, R.L., Andrews, R.C., 2018. Neural networks for dimensionality reduction of fluorescence spectra and prediction of drinking water disinfection by-products. *Water Res.* 136, 84–94. <https://doi.org/10.1016/j.watres.2018.02.052>
- Pernet-Coudrier, B., Varrault, G., Clouzot, L., Rousselot, O., Tusseau-Vuillemin, M.-H., Mouchel, J.-M., 2007. Caractérisation de la matière organique dissoute des rejets d'une importante station d'épuration et influence sur la biodisponibilité du cuivre. *Tech. Sci. Méthodes* 35–42. <https://doi.org/10.1051/tsm/200711035>
- Rad, S.M., Ray, A.K., Barghi, S., 2022. Water Pollution and Agriculture Pesticide. *Clean Technol.* 4, 1088–1102. <https://doi.org/10.3390/cleantechnol4040066>
- Romero González-Quijano, C., Herrero Ortega, S., Casper, P., Gessner, M.O., Singer, G.A., 2022. Dissolved organic matter signatures in urban surface waters: spatio-temporal patterns and drivers. *Biogeosciences* 19, 2841–2853. <https://doi.org/10.5194/bg-19-2841-2022>
- Seidel, M., Yager, P.L., Ward, N.D., Carpenter, E.J., Gomes, H.R., Krusche, A.V., Richey, J.E., Dittmar, T., Medeiros, P.M., 2015. Molecular-level changes of dissolved organic matter along the Amazon River-to-ocean continuum. *Mar. Chem., Biogeochemistry of dissolved organic matter* 177, 218–231. <https://doi.org/10.1016/j.marchem.2015.06.019>
- Smith, M.A., Kominoski, J.S., Gaiser, E.E., Price, R.M., Troxler, T.G., 2021. Stormwater Runoff and Tidal Flooding Transform Dissolved Organic Matter Composition and Increase Bioavailability in Urban Coastal Ecosystems. *J. Geophys. Res. Biogeosciences* 126, e2020JG006146. <https://doi.org/10.1029/2020JG006146>
- Søndergaard, M., Thomas, D., 2004. Dissolved Organic Matter (DOM) in Aquatic Ecosystems: A Study of European Catchments and Coastal Waters.
- Stedmon, C.A., Markager, S., Kaas, H., 2000. Optical Properties and Signatures of Chromophoric Dissolved Organic Matter (CDOM) in Danish Coastal Waters. *Estuar. Coast. Shelf Sci.* 51, 267–278. <https://doi.org/10.1006/ecss.2000.0645>
- Tang, K., Ooi, G.T.H., Spiliotopoulou, A., Kaarsholm, K.M.S., Sundmark, K., Florian, B., Kragelund, C., Bester, K., Andersen, H.R., 2020. Removal of Pharmaceuticals, Toxicity and Natural Fluorescence by Ozonation in Biologically Pre-Treated Municipal Wastewater, in Comparison to Subsequent Polishing Biofilm Reactors. *Water* 12, 1059. <https://doi.org/10.3390/w12041059>
- 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. *Sci. Total Environ.* 676, 252–267. <https://doi.org/10.1016/j.scitotenv.2019.04.160>
- Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Fram, M.S., Fujii, R., Mopper, K., 2003. Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. *Environ. Sci. Technol.* 37, 4702–4708. <https://doi.org/10.1021/es030360x>
- Wicke, D., Matzinger, A., Sonnenberg, H., Caradot, N., Schubert, R.-L., Dick, R., Heinzmann, B., Dünnbier, U., von Seggern, D., Rouault, P., 2021. Micropollutants in Urban Stormwater Runoff of Different Land Uses. *Water* 13, 1312. <https://doi.org/10.3390/w13091312>

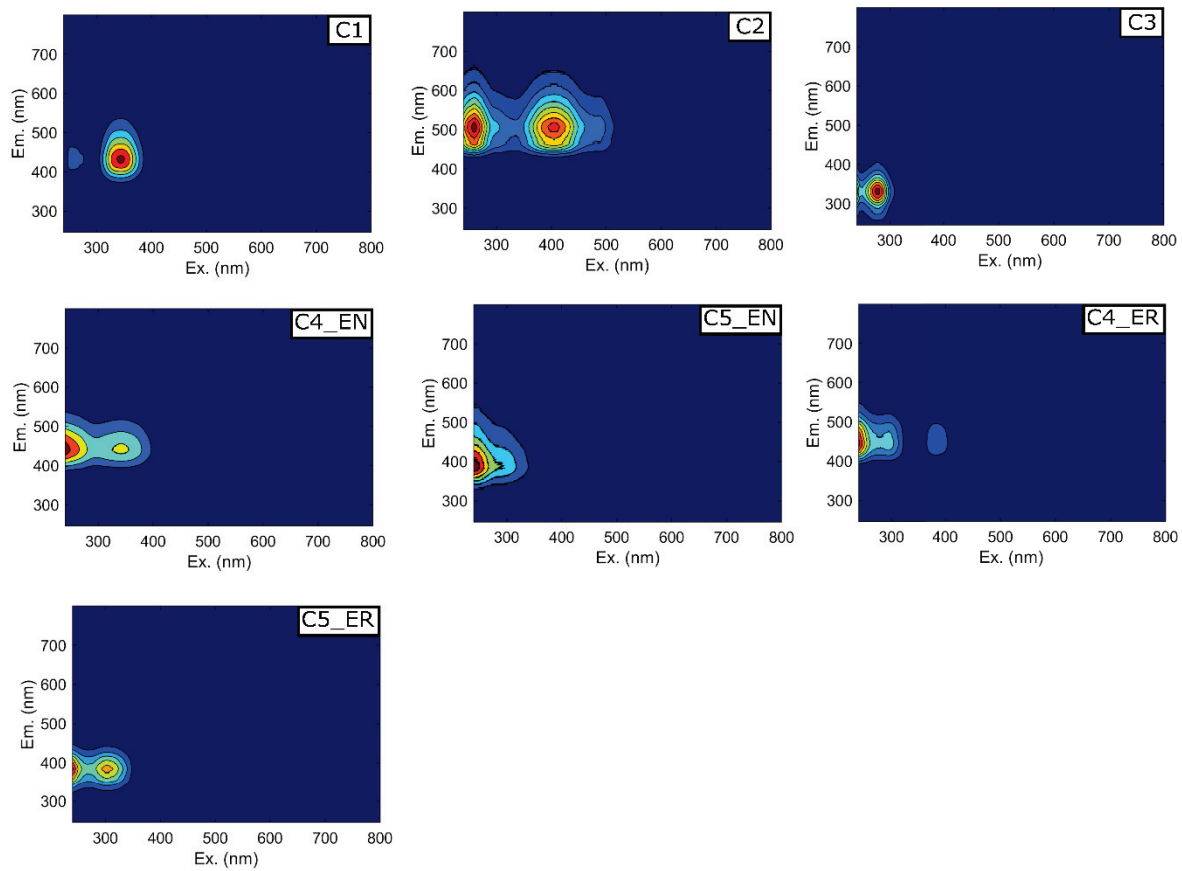
- Williams, C.J., Yamashita, Y., Wilson, H.F., Jaffé, R., Xenopoulos, M.A., 2010. Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. *Limnol. Oceanogr.* 55, 1159–1171. <https://doi.org/10.4319/lo.2010.55.3.1159>
- Wilson, H.F., Saiers, J.E., Raymond, P.A., Sobczak, W.V., 2013. Hydrologic Drivers and Seasonality of Dissolved Organic Carbon Concentration, Nitrogen Content, Bioavailability, and Export in a Forested New England Stream. *Ecosystems* 16, 604–616. <https://doi.org/10.1007/s10021-013-9635-6>
- Wise, J.L., Van Horn, D.J., Diefendorf, A.F., Regier, P.J., Lowell, T.V., Dahm, C.N., 2019. Dissolved organic matter dynamics in storm water runoff in a dryland urban region. *J. Arid Environ.* 165, 55–63. <https://doi.org/10.1016/j.jaridenv.2019.03.003>
- Wünsch, U.J., Murphy, K., 2021. A simple method to isolate fluorescence spectra from small dissolved organic matter datasets. *Water Res.* 190, 116730. <https://doi.org/10.1016/j.watres.2020.116730>
- Yamashita, Y., Cory, R.M., Nishioka, J., Kuma, K., Tanoue, E., Jaffé, R., 2010. Fluorescence characteristics of dissolved organic matter in the deep waters of the Okhotsk Sea and the northwestern North Pacific Ocean. *Deep Sea Res. Part II Top. Stud. Oceanogr., Ecological and Biogeochemical Interactions in the Dark Ocean* 57, 1478–1485. <https://doi.org/10.1016/j.dsr2.2010.02.016>
- Yamashita, Y., Kojima, D., Yoshida, N., Shibata, H., 2021. Relationships between dissolved black carbon and dissolved organic matter in streams. *Chemosphere* 271, 129824. <https://doi.org/10.1016/j.chemosphere.2021.129824>
- Yuan, D., An, Y., Wang, J., Chu, S., Lim, B., Chen, B., Xiong, Y., Kou, Y., Li, J., 2019. Dissolved organic matter characteristics of urban stormwater runoff from different functional regions during grassy swale treatment. *Ecol. Indic.* 107, 105667. <https://doi.org/10.1016/j.ecolind.2019.105667>
- Zark, M., Dittmar, T., 2018. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* 9, 1–8. <https://doi.org/10.1038/s41467-018-05665-9>
- Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere* 38, 45–50. [https://doi.org/10.1016/S0045-6535\(98\)00166-0](https://doi.org/10.1016/S0045-6535(98)00166-0)

7. Supplementary Material

Supplementary Material 1. Classical UV-Visible and fluorescence indicators to characterize dissolved organic matter.

Indicator	Calculation	Interpretation	References
SUVA ₂₅₄	$\frac{Abs\ 254\ nm}{[DOC]}$	Correlated with the aromaticity of DOM.	Weishaar et al., 2003
E2.E3	$\frac{Abs\ 250\ nm}{Abs\ 365\ nm}$	Negatively correlated to organic molecules size	Helms et al., 2008
E2.E4	$\frac{Abs\ 254\ nm}{Abs\ 436\ nm}$	Autochthonous versus terrestrial DOM	Jaffé et al., 2004; Li and Hur, 2017
E3.E4	$\frac{Abs\ 300\ nm}{Abs\ 400\ nm}$	Negatively correlated to humification degree	Claret et al., 2003; Li and Hur, 2017
S ₁ (275-295 nm)	$\alpha_{\lambda} = \alpha_{\lambda_0} e^{S(\lambda_0 - \lambda)}$ α_{λ} : absorption coefficient at specific wavelength λ_0 : reference wavelength S: exponential slope	Negatively correlated to aromatic carbon content	Stedmon et al., 2000 Helms et al., 2008
S ₂ (350-400 nm)			
SR	$\frac{S_1}{S_2}$	Negatively correlated to molecular weight	Helms et al., 2008
HIX	$\frac{\sum IF\ 435-480}{\sum IF\ 300-445}$ at 255 nm	Correlated to the aromatic carbon content/humification degree of DOM	Zsolnay et al., 1999
BIX	$\frac{IF\ 380}{IF\ 430}$ at 310 nm	Indicator of recent autochthonous production of DOM due to biological activity	Huguet et al., 2009
FI	$\frac{IF\ 450}{IF\ 500}$ at 370 nm	Terrigenous vs microbial DOM	Mcknight et al., 2001
YFI	$\frac{FEI\ 350-400}{FEI\ 400-450}$ at 280 nm FEI : average fluorescence intensities	Increases the resolution of FI indicator	Heo et al., 2016

Supplementary material 2. PARAFAC component excitation-emission matrices.



Chapitre 7.

Conclusion et perspectives

Chapitre 7. Conclusion et perspectives

Dans un contexte d'amélioration des stratégies de suivi de la pollution organique dans les cours d'eau, l'objectif général de cette thèse a été de tracer l'origine de la MOD en mettant en évidence des marqueurs ou empreintes physico-chimiques propres à différents types de sources (diffus vs. ponctuel, urbain vs. agricole). Pour cela, une première étape a été d'identifier et d'échantillonner les sources *a priori* de la MOD à l'échelle d'un bassin versant soumis à des pressions anthropiques, ce qui a requis la construction d'une base de données de plus de 140 échantillons. Dans un second temps, la MOD de chaque échantillon a été caractérisée avec une approche intégrée comprenant l'exploitation simultanée de l'ensemble des données générées par plusieurs techniques analytiques. Le dosage du DOC, la spectroscopie UV-Visible et l'EEM de fluorescence, la chromatographie HPSEC/UV-fluorescence et la spectrométrie de masse LC-HRMS sont utilisés pour étudier les caractéristiques et les propriétés physico-chimiques de la MOD. Dans une dernière étape, des approches statistiques univariées, multivariées et supervisées ont permis d'exploiter cette base de données originale en mettant en évidence des empreintes spécifiques aux sources sélectionnées *a priori*. Ces empreintes correspondent à une série d'indicateurs ou descripteurs de la MOD (e.g. aromaticité, humification, taille de molécules, polarité) calculés à partir des données générées par chaque technique.

Suite à une étude bibliographique portée sur les généralités de la MOD, les techniques de caractérisation et les sources de MOD à l'échelle de bassins versants, trois questions principales ont été identifiées :

1. Est-il possible de mettre en évidence de nouveaux indicateurs optiques et moléculaires de caractérisation de la MOD ?
2. Quelle méthodologie adopter pour échantillonner des sources diffuses terrigènes de MOD ?
3. L'approche intégrée multi-analytique et multi-sources permet-elle d'utiliser la MOD comme traceur de sources de pollutions anthropiques dans les cours d'eaux ?

Les réponses à ces trois questions ont été présentées dans les chapitres 4, 5 et 6. Ce dernier chapitre vise à donner une vue d'ensemble des principaux résultats et à suggérer des perspectives possibles pour les recherches futures.

7.1 Nouveaux indicateurs de caractérisation de la MOD issue de différentes sources

Parmi les techniques utilisées dans cette thèse, la chromatographie HPSEC/UV-fluorescence et la spectrométrie de masse LC-HRMS présentent des avantages significatifs. Elles apportent en effet des informations supplémentaires (par rapport aux techniques de spectroscopie UV-Visible et EEM de fluorescence) sur la structure et la composition moléculaire de la MOD. Cependant, la complexité de ces données, ainsi que les nombreux outils de retraitement nécessaires pour les exploiter, peuvent parfois rendre leur exploitation difficile. Dans ce contexte, ce travail a permis d'améliorer et de fiabiliser le traitement de ces données, et d'explorer leur potentiel pour construire des empreintes de MOD typiques de sources naturelles et anthropiques.

Indicateurs issus du couplage HPSEC/UV-fluorescence :

L'exploitation des chromatogrammes HPSEC/UV-fluorescence a été réalisée à l'aide d'une décomposition gaussienne, ce qui a permis d'identifier entre 4 et 8 composantes en fonction des échantillons. La somme des aires composantes a permis de calculer le *Total Sec Absorbance* (TSA) et le *Total Sec Fluorescence* (TSF), respectivement issus des chromatogrammes générés par détection UV (254 et 280 nm) et fluorescence (314-443, 280-330 et 278-354 nm). Les longueurs d'onde ont été sélectionnées de sorte à cibler des composés chromophoriques et fluorescents d'intérêt pour la caractérisation de la MOD (e.g. composés type-humiques, type-protéiques). A partir des TSA et TSF, il a été possible de mettre en place des indicateurs basés sur le fractionnement en poids moléculaire de la MOD. Ces indicateurs correspondent à des pourcentages de contribution en petites (LMW ; < 1 kDa), moyennes (MMW ; 1-10 kDa) et grosses molécules (HMW ; > 10 kDa). Ils ont permis de différencier des types de MOD issues de sources diffuses, et une étude de corrélation avec des indicateurs classiques UV-Visible et EEM de fluorescence a montré une complémentarité, mais également la limite d'interprétation de certains indicateurs classiques, comme l'indicateur SR (inversement corrélé au poids moléculaire) qui présentait des incohérences dans l'interprétation des résultats de ces indicateurs classiques.

Cette approche ouvre la voie à plusieurs perspectives d'amélioration. La première, purement analytique, concerne la courbe d'étalonnage avec l'usage des standards analytiques de type polystyrene sulfonate (PSS). Cette étape est primordiale pour le traitement et l'interprétation

des données, car elle permet d'attribuer un poids moléculaire à chaque pic chromatographique. L'élution des composés en HPSEC ne dépend toutefois pas uniquement de leur poids moléculaire mais également de leur conformation. Il est donc important de prendre en compte ce paramètre dans l'étude de la composition de la MOD, qui est très hétérogène en termes de taille et de structure moléculaire. Cela peut se faire en diversifiant les types de standards analytiques en incluant, par exemple, des standards certifiés de MOD (e.g. standards IHSS). La deuxième perspective d'amélioration de l'utilisation de l'HPSEC concerne le domaine d'exclusion. En effet, pour une grande partie des échantillons de MOD (échantillons prélevés dans les cours d'eau naturels, sorties de STEU et rejets des eaux pluviales et du ruissellement urbain), l'élution se fait dans le domaine compris entre 0.1 et 10 kDa. Il serait donc pertinent à l'avenir de restreindre le domaine d'exclusion chromatographique à cet intervalle (0.1 et 10 kDa), au lieu des 0.1 – 100 kDa, pour réduire le temps d'analyse (passer de 25 min d'acquisition à 10 min) et faciliter le traitement des chromatogrammes. En effet, l'usage de colonnes chromatographiques sur un domaine de taille plus restreint permettra d'améliorer la séparation des composés de petites tailles qui présentent un réel intérêt dans la caractérisation des sources de MOD.

En ce qui concerne la décomposition gaussienne et le calcul des indicateurs, plusieurs pistes d'améliorations sont également envisageables. En effet, dans le cadre de cette thèse, seules deux longueurs d'onde UV et trois longueurs d'onde d'excitation-émission de fluorescence ont été utilisées pour le calcul des indicateurs. Pour de futures applications, il paraît important d'exploiter l'ensemble des données UV-vis (200 à 450 nm) en travaillant sur les matrices 3D de temps de rétention, longueur d'onde et absorbance. Ainsi, il peut être envisageable d'exploiter de manière inédite cette gamme étendue de données UV en faisant une décomposition multidimensionnelle (type PARAFAC), comme réalisé en EEM de fluorescence, et d'identifier de nouvelles composantes chromophoriques d'intérêt dans l'étude de la MOD.

Enfin, une dernière perspective de recherche serait d'analyser les échantillons en sortie de colonne (après fractionnement HPSEC) en utilisant la spectroscopie EEM de fluorescence. Les matrices de fluorescence de chaque fraction d'échantillon peuvent être utilisées pour déterminer la taille et la distribution des composés fluorescents.

Les indicateurs issus du couplage LC-HRMS :

L'exploitation des données LC-HRMS acquises en approche non ciblée a nécessité plusieurs étapes de traitements préalables, à commencer par la génération d'une liste préliminaire de *features* (couples de m/z et temps de rétention des pics extraits). Plusieurs logiciels permettant de passer des données brutes acquises à cette liste de *features* (extraction des pics, regroupement des pics, intégration des pics et correction de la dérive intra-séquence). Plusieurs logiciels d'acquisition des données HRMS sont mis à disposition en libre accès et des études ont montré qu'il pouvait y avoir des différences notables entre les listes de *features* générées par ces différents logiciels. Dans une étude comparative, les logiciels MZmine et le package XCMS, sous leurs nouvelles versions (2.5 et 3.5), ont ainsi été comparés. Les critères de comparaison étaient basés sur le nombre total de *features* générés, la détection des composés des contrôles qualité (les traceurs d'injections et les composés du QC mix) et la cohérence de répartition des contrôles qualités après analyses statistiques multivariées (sur des données issues d'échantillons naturels et de contrôles qualité). Ces comparaisons ont montré que le nombre de *features* générés par XCMS était plus élevé (de 11 à 35 %), tout comme le nombre de composés du QC mix détectés, et la répartition des contrôles qualité était plus cohérente qu'avec le logiciel MZmine. En conséquence, le package XCMS a été utilisé dans l'ensemble de la thèse pour générer les listes préliminaires de *features* HRMS.

La deuxième étape de traitement des données LC-HRMS a consisté d'optimiser puis d'appliquer un *workflow* de retraitement qui permet de réduire le nombre pics parasites ou faux positifs en se basant principalement sur les intensités des composés. A travers une série de filtres qualité (seuils d'intensités et coefficient de variabilité évalués à partir des blancs d'injections, des QC *pools* et des triplicats d'injection randomisés), il a été possible de réduire drastiquement le nombre de composés tout en gardant une information fiable et pertinente pour une étude des sources de MOD.

La liste finale de composés moléculaires a permis de proposer deux exploitations possibles de données LC-HRMS. La première utilise une série de neuf indicateurs LC-HRMS globaux basés sur la répartition en masse et polarité des composés moléculaires. Ces indicateurs LC-HRMS ont montré une pertinence dans l'étude des sources de MOD et une complémentarité avec les indicateurs optiques sur les fractions chromophoriques et fluorescentes de la MOD. L'avantage de cette approche globale est qu'elle permet d'utiliser les listes de composés de

séquences analytiques différentes sans correction de dérive inter-séquence (variabilité des temps de rétention chromatographique (RT) et masses entre différentes séquences). Cela ouvre des perspectives intéressantes dans l'utilisation de l'approche non-ciblée dans l'étude moléculaire de la MOD. Néanmoins, cette approche se base principalement sur les intensités des composés : il est donc nécessaire de corriger la dérive instrumentale à l'aide d'outils de modélisation avant de procéder au calcul des indicateurs HRMS.

Une autre exploitation des données a été d'identifier des composés moléculaires spécifiques à différents types de sources de MOD. Pour un ensemble d'échantillons analysés dans une séquence, et sur la base de la présence/absence des composés (définie à l'aide d'un seuil d'intensité), il a été possible d'identifier des composés par type de source. Ces composés moléculaires peuvent être considérés comme marqueurs pertinents dès lors qu'ils sont présents spécifiquement dans une source et pas dans une autre. Toutefois, cette approche présente le principal inconvénient de nécessiter une correction de la dérive des temps de rétention et de masses entre les séquences. Néanmoins, des corrections de RT inter-séquences restent possibles et des outils de comparaison de liste de composés tels que *MetabCombiner* permettent de compenser cette variabilité.

La question du retraitement des données HRMS suscite encore aujourd'hui de nombreuses réflexions qui témoignent des nombreux travaux qui restent à accomplir dans ce domaine, notamment en ce qui concerne la normalisation des protocoles de traitement des données inter-laboratoires et le développement de *workflows* permettant la comparaison inter-séquences. Ainsi, il n'existe actuellement aucun consensus sur les meilleures pratiques en matière de traitement des données HRMS dans le domaine de la chimie environnementale, et cela peut poser question sur la comparabilité des résultats publiés dans la littérature scientifique. Il s'agit d'un verrou pour les études environnementales sur les tendances temporelles (sur de longues durées). Ce manque d'homogénéité et de formalisme dans le traitement des données HRMS représente un enjeu fort, un défi à relever qui mobilise actuellement la communauté scientifique. Par exemple, le réseau NORMAN (*Network of reference laboratories for monitoring of emerging environmental substances*, [www.https://www.norman-network.net/](https://www.norman-network.net/)) œuvre au développement de protocoles communs pour l'acquisition et le traitement des données HRMS ainsi qu'à la création de plateformes de

partage de données (bases de données). Ces avancées à venir dans le domaine bénéficieront, à terme, à la caractérisation des MOD et à l'établissement de typologies des sources de MOD.

Conclusion sur les indicateurs innovants :

Les indicateurs et marqueurs moléculaires proposés dans cette thèse peuvent contribuer significativement aux recherches sur la MOD. La complémentarité entre les indicateurs issus de la LC-HRMS et ceux issus des techniques optiques permettra d'améliorer les connaissances sur la composition moléculaire de la MOD. Cette approche innovante pour caractériser des sources de MOD devrait ainsi permettre une meilleure compréhension des processus qui contrôlent sa dynamique dans les cours d'eau. A terme, ces indicateurs pourraient être utilisés à large échelle, comme par exemple pour la surveillance des rejets agricoles, industriels et urbains.

7.2 Validation d'une méthodologie de prélèvement des sources diffuses

Parmi la diversité des sources potentielles de MOD à l'échelle des bassins versants, les sources diffuses représentent une part non-négligeable de MOD dans les cours d'eau. Cependant, leur localisation et leur échantillonnage reste très difficile, du fait de la multiplicité des apports dans les cours d'eau, de la complexité de la MOD ainsi que de l'occurrence d'évènements non contrôlés à l'origine de ces sources. Le chapitre 5 avait comme objectif de déterminer la représentativité de l'échantillonnage d'une source diffuse terrigène de MOD dans les cours d'eau. Pour cela, le lien entre la MOD extraite de sols, caractérisés par une occupation particulière (forêts/prairies et vignes), et la MOD collectée dans les cours d'eau adjacents aux sols a été étudié sur le bassin de l'Ardières-Morcille. Les caractéristiques de la MOD extraite de sols ont ainsi été comparées à celles de la MOD provenant de cours d'eau adjacents, prélevée en période de hautes eaux et de basses eaux.

Les résultats ont montré une empreinte spécifique à la MOD extraite des sols forestiers et viticoles puis une évolution de l'empreinte au cours d'une. Une expérience d'incubation en laboratoire de 28 jours a ensuite mis en évidence une évolution de l'empreinte. Les indicateurs issus des techniques optiques ont révélé que l'évolution de la MOD s'est principalement produite au cours des trois premiers jours après l'incubation. Cela a entraîné la dégradation de molécules de type protéique, à la fois de faible poids moléculaire (<1 kDa) et de poids moléculaire élevé (>10 kDa). Or, cette fraction labile de la MOD n'est pas celle que l'on

retrouve dans les cours d'eau récepteurs. La recherche d'indicateurs de qualité des MOD selon les occupations des sols a donc été focalisée sur la fraction réfractaire, après incubation des extraits de MOD du sol. Les processus de transformation de la MOD pendant l'incubation (expérience réalisée en l'absence de lumière, comme dans les sols) s'expliquent principalement par l'activité des microorganismes. En perspective, afin de mieux expliquer les différences observées entre les caractéristiques de la MOD dans les cours d'eau récepteurs selon l'occupation des sols en amont, il serait intéressant de compléter cette expérience d'incubation avec une analyse des communautés microbiennes natives (initiales) des sols et d'étudier le lien avec la fraction réfractaire obtenue. L'approche intégrée employée ici permettrait ainsi d'améliorer la compréhension de la composition de la MOD et de son rôle dans la structuration des communautés microbiennes dans différents environnements (sol, eau, sédiments, biofilm) et inversement, d'étudier plus en détail comment les communautés microbiennes modifient la MOD par des processus tels que la dégradation ou la transformation.

Les empreintes optiques de MOD dans les cours d'eau adjacents étaient différenciées et principalement influencées par les conditions hydrologiques. Ainsi, les échantillons de MOD prélevés pendant les périodes de hautes eaux présentaient des contributions plus élevées des fractions LMW et HMW (indiquant un apport terrigène majoritaire spécifique aux usages de sols) par rapport aux échantillons de MOD prélevés pendant les périodes de basses eaux (mélange terrigène et autochtone). Les analyses linéaires discriminantes (LDA) ont révélé que 9 indicateurs sont nécessaires pour différencier les empreintes entre la MOD forestière et la MOD viticole. Ces indicateurs sont issus des trois techniques utilisées, ce qui démontre la complémentarité de ces techniques. Il s'agit des indicateurs UV-visible SR et SUVA, des indicateurs d'EEM de fluorescence HIX, FI, YFI et la composante PARAFAC C1 (Ex<240-320_Em 420 nm ; associé à la MOD terrigène), ainsi que les indicateurs HPSEC/UV-fluorescence LMW-TSA 254, MMW TSF 278-354 et MMW TSF 280-330. Pour finir, les résultats ont montré que la méthodologie d'échantillonnage, directement dans la rivière, d'une source diffuse de MOD doit tenir compte à la fois de l'homogénéité en termes d'usage des sols des sous-bassins versants en amont et des conditions hydrologiques (contribution plus ou moins importante de MOD autochtone aux apports terrigènes). Cette méthodologie a ensuite été appliquée à d'autres bassins versants (l'Yzeron et le Gier). Les résultats ont montré l'existence d'un effet

bassin versant visible dans les échantillons prélevés en aval des sources forestières et agricoles et ce en période de hautes eaux et de basses eaux. Les échantillons prélevés dans les retenues d'eau présentaient ainsi des empreintes plus homogènes, sans doute associées à une MOD autochtone produite localement. Cependant, il a été possible de s'affranchir de l'effet de bassin versant en identifiant une série d'indicateurs par analyse discriminante linéaire. La combinaison d'indicateurs optiques et moléculaires a permis de classer les différentes sources diffuses liées aux occupations de sols (principalement celles prélevées en hautes eaux) avec des erreurs relativement réduites (< 30 %).

Cette méthodologie d'échantillonnage des sources diffuses permet de mieux comprendre l'impact des sources diffuses sur la qualité de la matière organique dissoute des rivières. Elle peut également être utile pour comprendre l'influence de l'occupation des sols dans le cycle biogéochimique du carbone. Mieux identifier les sources diffuses à l'échelle d'un bassin versant peut alors donner des informations précieuses sur les chemins de l'eau et l'impact des pratiques d'usage des sols sur les milieux aquatiques. Ces connaissances permettront une meilleure gestion du territoire et des pratiques (e.g. agricoles, urbaines) pour réduire les apports de contaminants associés à chaque usage du sol.

7.3 Démonstration de la pertinence de la MOD comme traceur de sources de pollution dans les cours d'eau à l'échelle d'un bassin versant : application de l'approche intégrée.

La MOD ne provient pas uniquement des sources diffuses étudiées, il existe bien entendu de nombreuses autres sources de rejets anthropiques qui représentent des apports considérables de MOD et qui contribuent, par la même occasion, à la qualité de l'eau dans les milieux récepteurs. Les activités industrielles, le développement urbain ou le transport routier génèrent un apport considérable de polluants (e.g. produits pharmaceutiques, hydrocarbures, pesticides et sous-produits de transformation) dans les cours d'eau. Ces apports sont canalisés (il s'agit donc de sources ponctuelles) vers des stations de traitement des eaux usées (STEU), mais il arrive qu'ils se retrouvent directement dans le milieu naturel via les rejets urbains d'eau pluviale ou les déversoirs d'orage ou à cause de dysfonctionnements (mauvais raccordements du réseau d'assainissement ou débordement du réseau en période de fortes pluies). Malgré la grande complexité de la composition de la MOD, son lien étroit avec les sources de pollution

permet d'envisager son utilisation comme traceur de pollution dans les cours d'eau à l'échelle des bassins versants.

Afin de déterminer si l'on pouvait mettre en évidence le lien entre les caractéristiques de la MOD et l'ensemble des sources prélevées à l'échelle d'un bassin versant soumis à des rejets anthropiques, une approche intégrée basée sur l'exploitation simultanée des données UV-Visible, EEM de fluorescence, HPSEC/UV-Visible et LC-HRMS a été appliquée sur une banque d'échantillons de sources diffuses provenant de sols forestiers, agricoles et de retenues d'eau, d'apports autochtones et de sources ponctuelles de pollution en zone urbaine (entrées et sortie de STEU, ruissellement urbain) .

Les propriétés optiques et moléculaires globales de la MOD ont révélé un caractère aromatique, avec la présence de composés de poids moléculaire élevé dans les eaux les moins impactées par les activités humaines (dans la zone forêt-pâturage-agriculture), qui sont moins présents dans les eaux plus impactées par les activités humaines (comme les eaux usées et les rejets de pluie et de ruissellement urbain). Dans les cours d'eau naturels en zone rurale, la MOD est principalement influencée par les conditions hydrologiques, avec un impact des usages des sols pendant les périodes de hautes eaux. La MOD prélevée en aval des sols forestiers est constituée de composés de poids moléculaire relativement faible, avec de faibles degrés d'humification et une contribution plus élevée de molécules de type protéique inférieures à 1 kDa. A l'inverse, la MOD agricole est caractérisée par un poids moléculaire plus élevé, un degré d'humification plus élevé et une contribution plus faible de petites molécules de type protéique.

La MOD échantillonnée dans les eaux brutes (entrées) des STEU n'a pas montré de différence significative en fonction des rejets en amont. En effet, les eaux brutes issues de grosses collectivités sont assez similaires à celles des petites collectivités au regard de l'empreinte de MOD. Celle-ci est principalement caractérisée par un mélange de composés de faible poids moléculaire, peu aromatiques et très polaires. La composante C2, fortement visible dans ces eaux, indique la présence de composés humiques dégradés, et l'indicateur YFI reflète une activité microbienne élevée. En revanche, des différences ont été observées dans les rejets (sorties) de STEU, en fonction du type de traitement appliqué. La MOD des rejets de STEU qui utilisent un traitement intensif (boues activées) est caractérisée par un poids moléculaire

élevé de composés de type humique, contrairement aux sorties des filtres plantés de roseaux (PRF) où la MOD est plus diversifiée au regard de la taille (fractions LMW et MMW) et de la nature (type de protéines, type d'acides aminés et type d'humus) de ses composés. Le degré d'humification et d'aromaticité de la MOD à la sortie des PRF est ainsi plus proche de celui des eaux naturelles rurales évoquées précédemment. Enfin, les sources de rejet des eaux pluviales et de ruissellement urbain présentent des signatures très spécifiques, avec une MOD constituée principalement de composés hautement aromatiques, non polaires et de faible poids moléculaire (potentiellement un indicateur de composés hydrocarbures aromatiques polycycliques).

Le traitement des données LC-HRMS selon une approche non ciblée nous a permis de définir neuf indicateurs globaux basés sur les classes de poids moléculaire et de polarité des molécules. Ces indicateurs se sont révélés pertinents pour différencier les MOD selon des types de sources. Une exploitation plus détaillée des données LC-HRMS a révélé la présence d'un grand nombre de composés moléculaires non spécifiques à une source de MOD (fond moléculaire universel), et d'autres spécifiques à un type de source, et dont le nombre varie selon le type de source. Ces composés pourraient constituer de bons marqueurs des sources de pollutions anthropiques dans les milieux récepteurs, car ils fournissent des informations supplémentaires sur la composition moléculaire par rapport aux indicateurs optiques.

Une des perspectives ouverte par ces travaux de thèse concerne l'utilisation de ces marqueurs dans des modèles mathématiques de mélange de sources. Pour cela, il faut s'assurer du caractère conservatif et additif des marqueurs optiques et moléculaires, et tester leur pertinence dans des modèles de mélanges (e.g. *End Member Mixing analysis*, *Bayesian Mixing model*). Cela devrait permettre d'estimer les contributions des sources de MOD, et donc de contaminants associés, dans les cours d'eau à l'échelle des bassins versants et dans les zones de mélange (e.g. aux exutoires des bassins versants). Ces nouvelles connaissances permettront de prendre des mesures appropriées pour contrôler leurs rejets dans l'environnement. Ce type d'outil d'aide à la décision peut ainsi servir en appui aux politiques publiques et à la mise en place de réglementations environnementales pour protéger les écosystèmes aquatiques. Mieux identifier les sources de pollution revient également à mieux appréhender les risques potentiels pour la santé, risques associés à l'exposition à certains polluants ; et contribue à mettre en place des stratégies pour minimiser leurs effets potentiels.

Par exemple, l'identification de composés moléculaires spécifiques dans les eaux usées peut aider à prendre des mesures appropriées pour contrôler leurs rejets dans les rivières réceptrices (utilisation de ces molécules comme traceurs de rejets de STEU dans les rivières, voire aussi pour tracer des dysfonctionnements dans les procédés de traitement). L'identification de séries de marqueurs et d'empreintes spécifiques dans les eaux usées peut également être utile pour aider à optimiser les procédés de traitement, à contrôler les dysfonctionnements et à réduire l'impact des rejets de STEU sur les milieux aquatiques. Enfin, les marqueurs identifiés pourraient aussi servir pour la surveillance de la qualité de l'eau destinée à la consommation (*i.e.* surveillance des zones de captage et de l'efficacité des procédés de potabilisation) et contribuer ainsi à la préservation de nos ressources en eau.

Cette caractérisation intégrée de la MOD offre une approche prometteuse pour étudier les liens entre les contaminants et la MOD, ainsi qu'entre la MOD et les microorganismes dans divers environnements tels que les sols, les eaux et les sédiments. L'originalité de ce travail réside dans la création d'une vaste base de données sur les caractéristiques de la MOD provenant de diverses sources. Cette approche a le potentiel de globaliser notre connaissance de la MOD et de son impact sur différents environnements. Pour continuer à construire cette base de données, il serait nécessaire d'harmoniser certaines pratiques de caractérisation de la MOD afin de collecter des données auprès d'autres laboratoires. Cela permettrait de faire progresser notre compréhension de la MOD et de son rôle dans le fonctionnement de l'écosystème.