

SUJET DE THESE AU LEESU – SCIENCES DE L’ENVIRONNEMENT

Relier la pression chimique à ses effets sur la biodiversité dans les milieux aquatiques en couplant l'analyse non ciblée avec l'ADN environnemental à l'aide d'outils numériques

Contexte

La qualité des milieux aquatiques et de la ressource en eau peut être appréhendée du point de vue de sa contamination chimique, en particulier par des contaminants émergents, ou de la biodiversité qu'elle accueille. La présence de ces contaminants est néfaste pour la santé des écosystèmes avec des effets sur les organismes aquatiques et une perte de biodiversité (Santos *et al.* 2010; Hamilton *et al.* 2016; Reid *et al.* 2019; Alderton *et al.* 2021). Afin de caractériser cette contamination chimique dans les milieux aquatiques, des méthodes d'analyse non ciblée par spectrométrie de masse haute résolution (HRMS) ont été développées (Aurich *et al.* 2023; Hollender *et al.* 2023). Au Leesu, deux thèses ont été menées sur cette thématique et ont permis d'une part de développer la méthode analytique (Huynh *et al.* 2021) et les outils informatiques de traitement des données (Sade *et al.* 2022), et d'autre part d'appliquer ces méthodes à différentes eaux urbaines afin d'évaluer la variabilité spatio-temporelle de la contamination. En parallèle, l'étude de l'ADN environnemental (ADNe), dans lequel les signaux des macro- et micro-organismes de l'environnement sont récupérés à partir d'échantillons d'eau est de plus en plus utilisée pour la surveillance de la biodiversité des milieux (Duarte *et al.* 2021; Altermatt *et al.* 2023). Ces méthodes permettent de détecter les espèces aussi bien, voire mieux, que les méthodes traditionnelles d'étude de biodiversité, en particulier dans l'environnement aquatique qui peut intégrer des informations sur la biodiversité de tout le bassin versant (Altermatt *et al.* 2023; Bunholi *et al.* 2023; Reji Chacko *et al.* 2023).

Les progrès techniques qui permettent aujourd'hui d'enregistrer des empreintes chimiques par des approches non ciblées, ou de mesurer l'ADN environnemental pour étudier la biodiversité, fournissent d'énormes quantités de données. Le défi consiste à mettre en place les outils d'analyse biologique et chimique multivariée pour faciliter l'utilisation de ces données (Slobodnik *et al.* 2019). L'utilisation d'outils numériques comme le machine learning se développe pour traiter et interpréter les spectres HRMS de contaminants dans les eaux (Sade *et al.* 2022; Arturi and Hollender 2023; Hollender *et al.* 2023), de même que dans le domaine de l'ADN environnemental pour extraire des indicateurs de biodiversité ou expliquer les variabilités spatio-temporelles (Cordier *et al.* 2018; Hu *et al.* 2023)

Pour l'instant ces deux approches n'ont pas été couplées. Or les mélanges complexes de produits chimiques doivent être pris en compte en même temps que leurs effets complexes et leurs impacts sur les écosystèmes (Slobodnik *et al.* 2019). Le couplage entre l'empreinte chimique des contaminants et l'ADNe qui mesure la biodiversité est un enjeu majeur pour la surveillance de la qualité des milieux récepteurs. Des essais de couplage entre ADNe et hydrodynamique ont été tentés (Warter *et al.* 2024) mais pour l'instant les études couplant spectres HRMS et ADNe sont rares (Reid *et al.* 2019; Sieber *et al.* 2023). L'objectif général de ce projet de thèse est donc d'évaluer la possibilité d'interpréter conjointement les empreintes

chimiques de contaminants dans les eaux avec les informations sur la biodiversité données par l'ADNe en développant et appliquant des méthodes numériques de type machine learning.

Cadre et partenaires du projet

Cette thèse s'insère dans les actions de recherche du Leesu en collaboration avec le programme MeSeine Innovation porté par le Syndicat Interdépartemental pour l'Assainissement de l'Agglomération Parisienne (SIAAP) dont un des objectifs est d'améliorer le suivi des masses d'eau en couplant différentes approches hydrologiques, chimiques et biologiques. Des campagnes de prélèvements sur 4 à 7 sites répartis entre la Seine, la Marne et l'Oise sont réalisées régulièrement (figure 1).

Les analyses de contaminants émergents s'appuient sur les instruments analytiques de la plateforme Prammics (OSU Efluve), en particulier les instruments de chromatographie en phase liquide (Waters Vion – UPLC-IMS-QTOF et Shimadzu HPLC Fraction collector). L'UPLC-IMS-QTOF est un instrument de HRMS équipé d'une séparation par mobilité ionique (IMS). L'ADNe sera mesuré par un prestataire.



Figure 1 : Réseau hydrographique et points de prélèvement de l'observatoire MeSeine

Objectifs de la thèse

Les objectifs de la thèse sont les suivants :

- Identifier pour chaque type de données (HRMS et ADNe) des traceurs ou indicateurs d'intérêt à partir de la littérature scientifique et automatiser leur traitement.
- Développer des méthodes de traitement des données pour coupler les données HRMS et d'ADNe en utilisant des outils statistiques avancés ou numériques de type machine learning.
- Proposer une stratégie d'échantillonnage pertinente pour acquérir les deux types de données (HRMS et ADNe) sur le réseau MeSeine et effectuer un suivi régulier des différents points d'échantillonnage.
- Suivre et interpréter le devenir spatio-temporel de molécules/marqueurs d'intérêt identifiés précédemment (suivi à différentes fréquences : saisonnières, hebdomadaires, journalières et le long des différents points de prélèvement en Seine)

Laboratoire d'accueil

La thèse se déroulera au Laboratoire Eau Environnement et Systèmes Urbains (LEESU - UMR MA102 (Université Paris-Est Créteil et École des Ponts ParisTech), principalement sur le site de la Maison des Sciences de l'Environnement à Créteil (61 avenue du Général de Gaulle, Créteil). Une partie de la thèse pourra être menée au sein des locaux du SIAAP à Colombes.

Profil recherché

Formation de niveau M2 ou ingénieur en bioinformatique/biostatistiques, chimie analytique ou chimie de l'environnement :

- Compétences en statistiques (régressions, classification automatique discriminante, arbres de décision...) et analyse de données, machine learning
- Compétences en programmation, utilisation de langages de programmation pour le traitement de données (data science) (R, python...)
- Compétences en chimie analytique (spectrométrie de masse, métabolomique) et/ou biologie moléculaire
- Des connaissances en sciences de l'environnement (polluants, qualité des eaux et notions de biodiversité) sont bienvenues
- Aisance dans la rédaction et bon niveau d'anglais

Contacts et modalités de candidature

Encadrement

- Adèle Bressy, chargée de recherche HDR de l'École des Ponts, chimie de l'environnement, hydrologie (adele.bressy@enpc.fr)
- Julien Le Roux, maître de conférences UPEC, chimie analytique, outils statistiques et numériques, chimie de l'environnement (julien.le-roux@u-pec.fr)
- My Dung Jusselme, maîtresse de conférence UPEC, biologie moléculaire, microbiologie environnementale
- Sabrina Guérin, responsable innovation au SIAAP

Candidature

Envoyer CV, lettre de motivation et contact de référent.e.s scientifiques à adele.bressy@enpc.fr et julien.le-roux@u-pec.fr.

Références bibliographiques

- Alderton I, Palmer BR, Heinemann JA, Patti I, Weaver L, Gutiérrez-Ginés MJ, Horswell J, Tremblay LA (2021). The role of emerging organic contaminants in the development of antimicrobial resistance. *Emerging Contaminants* 7, 160–171. doi:10.1016/j.emcon.2021.07.001
- Altermatt F, Carraro L, Antonetti M, Albouy C, Zhang Y, Lyet A, Zhang X, Pellissier L (2023). Quantifying biodiversity using eDNA from water bodies: General principles and recommendations for sampling designs. *Environmental DNA* 5, 671–682. doi:10.1002/edn3.430
- Arturi K, Hollender J (2023). Machine Learning-Based Hazard-Driven Prioritization of Features in Nontarget Screening of Environmental High-Resolution Mass Spectrometry Data. *Environmental Science & Technology* 57, 18067–18079. doi:10.1021/acs.est.3c00304
- Aurich D, Diderich P, Helmus R, Schymanski EL (2023). Non-target screening of surface water samples to identify exposome-related pollutants: a case study from Luxembourg. *Environmental Sciences Europe* 35. doi:10.1186/s12302-023-00805-5
- Bunholi IV, Foster NR, Casey JM (2023). Environmental DNA and RNA in aquatic community ecology: Toward methodological standardization. *Environmental DNA*. doi:10.1002/edn3.476
- Cordier T, Forster D, Dufresne Y, Martins CIM, Stoeck T, Pawlowski J (2018). Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources* 18, 1381–1391. doi:10.1111/1755-0998.12926
- Duarte S, Leite BR, Feio MJ, Costa FO, Filipe AF (2021). Integration of DNA-based approaches in aquatic ecological assessment using benthic macroinvertebrates. *Water (Switzerland)* 13. doi:10.3390/w13030331
- Hamilton PB, Cowx IG, Oleksiak MF, Griffiths AM, Grahn M, Stevens JR, Carvalho GR, Nicol E, Tyler CR (2016). Population-level consequences for wild fish exposed to sublethal concentrations of chemicals – a critical review. *Fish and Fisheries* 17, 545–566. doi:10.1111/faf.12125
- Hollender J, Schymanski EL, Ahrens L, Alygizakis N, Béen F, Bijlsma L, Brunner AM, Celma A, Fildier A, Fu Q, Gago-Ferrero P, Gil-Solsona R, Haglund P, Hansen M, Kaserzon S, Kruve A, Lamoree M, Margoum C, Meijer J, Merel S, Rauert C, Rostkowski P, Samanipour S, Schulze B, Schulze T, Singh RR, Slobodnik J, Steininger-Mairinger T, Thomaidis NS, Togola A, Vorkamp K, Vulliet E, Zhu L, Krauss M (2023). NORMAN guidance on suspect and non-target screening in environmental monitoring. *Environmental Sciences Europe* 35. doi:10.1186/s12302-023-00779-4
- Hu H, Wei X-Y, Liu L, Wang Y-B, Jia H-J, Bu L-K, Pei D-S (2023). Supervised machine learning improves general applicability of eDNA metabarcoding for reservoir health monitoring. *Water Research* 246. doi:10.1016/j.watres.2023.120686
- Huynh N, Caupos E, Soares Peirera C, Le Roux J, Bressy A, Moilleron R (2021). Evaluation of Sample Preparation Methods for Non-Target Screening of Organic Micropollutants in Urban Waters Using High-Resolution Mass Spectrometry. *Molecules* 26, 7064. doi:10.3390/molecules26237064
- Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke SJ (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews* 94, 849–873. doi:10.1111/brv.12480
- Reji Chacko M, Altermatt F, Fopp F, Guisan A, Keggin T, Lyet A, Rey P-L, Richards E, Valentini A, Waldock C, Pellissier L (2023). Catchment-based sampling of river eDNA integrates terrestrial and aquatic biodiversity of alpine landscapes. *Oecologia* 202, 699–713. doi:10.1007/s00442-023-05428-4
- Sade J, Le Roux J, Mottelet S, Guérin-Rechdaoui S, Moilleron R (2022). Caractérisation de la contamination des eaux de surface par le couplage d'analyses non-ciblées en spectrométrie de masse avec des analyses d'écotoxicologie - Rapport d'avancement du projet MeSeine Innovation.
- Santos LHMLM, Araújo AN, Fachini A, Pena A, Delerue-Matos C, Montenegro MCBSM (2010). Ecotoxicological aspects related to the presence of pharmaceuticals in the aquatic environment. *Journal of Hazardous Materials* 175, 45–95. doi:10.1016/j.jhazmat.2009.10.100

Sieber G, Drees F, Shah M, Stach TL, Hohrenk-Danzouma L, Bock C, Vosough M, Schumann M, Sures B, Probst AJ, Schmidt TC, Beisser D, Boenigk J (2023). Exploring the efficacy of metabarcoding and non-target screening for detecting treated wastewater. *Science of The Total Environment* 903, 167457. doi:10.1016/j.scitotenv.2023.167457

Slobodnik J, Hollender J, Schulze T, Schymanski EL, Brack W (2019). Establish data infrastructure to compile and exchange environmental screening data on a European scale. *Environmental Sciences Europe* 31, 65. doi:10.1186/s12302-019-0237-6

Warter MM, Tetzlaff D, Ring A-M, Christopher J, Kissener HL, Funke E, Sparmann S, Mbedi S, Soulsby C, Monaghan MT (2024). Environmental DNA, hydrochemistry and stable water isotopes as integrative tracers of urban ecohydrology. *Water Research* 250. doi:10.1016/j.watres.2023.121065

PhD THESIS PROPOSAL – ENVIRONMENTAL SCIENCES

Bridging the gap between biodiversity data and chemical pressure in aquatic environments by coupling non-targeted analysis with environmental DNA using digital tools

Context

The quality of aquatic environments and water resources can be assessed from the perspective of chemical contamination, especially by emerging contaminants, or the biodiversity they host. The presence of contaminants is harmful to ecosystems, with effects on aquatic organisms and a loss of biodiversity (Santos *et al.* 2010; Hamilton *et al.* 2016; Reid *et al.* 2019; Alderton *et al.* 2021). To characterize this chemical contamination in aquatic environments, non-targeted analysis methods using high-resolution mass spectrometry (HRMS) have been developed (Aurich *et al.* 2023; Hollender *et al.* 2023). At Leesu, two PhD theses have been conducted on this topic, leading to the development of an analytical method (Huynh *et al.* 2021) and data processing tools (Sade *et al.* 2022), as well as the application of these methods to different urban waters to assess spatiotemporal variability of contamination. Concurrently, the study of environmental DNA (eDNA), where signals from macro- and microorganisms in the environment are retrieved from water samples, is increasingly used for monitoring the biodiversity of environments (Duarte *et al.* 2021; Altermatt *et al.* 2023). These methods can detect species as well as or even better than traditional biodiversity study methods, especially in aquatic environments that can integrate information about the entire watershed's biodiversity (Altermatt *et al.* 2023; Bunholi *et al.* 2023; Reji Chacko *et al.* 2023).

Advances in technology that allow recording chemical fingerprints through non-targeted approaches or measuring eDNA for biodiversity studies provide enormous amounts of data. The challenge is to establish multivariate biological and chemical analysis tools to facilitate the use of this data (Slobodnik *et al.* 2019). Digital tools such as machine learning are increasingly used to process and interpret HRMS spectra of contaminants in water (Sade *et al.* 2022; Arturi and Hollender 2023; Hollender *et al.* 2023), as well as in the field of eDNA to extract biodiversity indicators or explain spatiotemporal variabilities (Cordier *et al.* 2018; Hu *et al.* 2023).

Currently, these two approaches have not been coupled. However, complex mixtures of chemicals must be considered along with their complex effects and impacts on ecosystems (Slobodnik *et al.* 2019). The coupling of the chemical footprint of contaminants and eDNA, which measures biodiversity, is a major challenge for monitoring the quality of receiving environments. Attempts to couple eDNA with hydrodynamics have been made (Warter *et al.* 2024) but studies coupling HRMS spectra and eDNA are currently rare (Reid *et al.* 2019; Sieber *et al.* 2023). The overall objective of this thesis project is to evaluate the possibility of jointly interpreting chemical fingerprints of contaminants in water with biodiversity information provided by eDNA by developing and applying numerical methods such as machine learning.

Project framework and partners

This PhD thesis is part of the Leesu's research activities in collaboration with the MeSeine Innovation program led by the Public Service of Parisian Sanitation (SIAAP), one of whose objectives is to improve the monitoring of water bodies by combining different hydrological, chemical, and biological approaches. Regular sampling campaigns are carried out at 4 to 7 sites distributed between the Seine, the Marne, and the Oise (Figure 1).

The analysis of emerging contaminants relies on analytical instruments from the Pramms platform (OSU Efluve), especially liquid chromatography instruments (Waters Vion – UPLC-IMS-QTOF and Shimadzu HPLC Fraction collector). The UPLC-IMS-QTOF is an HRMS instrument equipped with ion mobility separation (IMS). eDNA will be measured by a service provider.



Figure 1: Hydrographic network and sampling points of the MeSeine observatory

Objectives

The objectives of the PhD thesis are as follows:

- Identify for each type of data (HRMS and eDNA) tracers or indicators of interest from scientific literature and automate their processing.
- Develop data processing methods to couple HRMS and eDNA data using advanced statistical or numerical tools such as machine learning.
- Propose a relevant sampling strategy to acquire both types of data (HRMS and eDNA) on the MeSeine network and regularly monitor different sampling points.
- Monitor and interpret the spatiotemporal fate of previously identified molecules/markers of interest (monitoring at different frequencies: seasonal, weekly, daily, and at different sampling points along the Seine river).

Hosting laboratory

The thesis will be conducted at the Water, Environment and Urban Systems Laboratory (LEESU - UMR MA102, University of Paris-Est Créteil and École des Ponts ParisTech), mainly at the Maison des Sciences de l'Environnement in Créteil (61 avenue du Général de Gaulle, Créteil). Part of the thesis may be carried out at SIAAP (Colombes).

Profile

Master's degree or engineering degree in bioinformatics/biostatistics, analytical chemistry, or environmental chemistry:

- Skills in statistics (regression, discriminant automatic classification, decision trees...) and data analysis, machine learning
- Programming skills, use of programming languages for data processing (R, Python...)
- Skills in analytical chemistry (mass spectrometry, metabolomics) and/or molecular biology
- Knowledge in environmental sciences (pollutants, water quality, and biodiversity concepts) is welcome
- Proficiency in writing and a good level of English

Contacts and application details

Supervision:

- Adèle Bressy, HDR research associate at École des Ponts ParisTech, environmental chemistry, hydrology (adele.bressy@enpc.fr)
- Julien Le Roux, assistant professor at UPEC, analytical chemistry, statistical and numerical tools, environmental chemistry (julien.le-roux@u-pec.fr)
- My-Dung Jusselme, assistant professor at UPEC, molecular biology, environmental microbiology
- Sabrina Guérin, head of innovation at SIAAP

Application:

Send CV, cover letter, and scientific references to:

adele.bressy@enpc.fr and julien.le-roux@u-pec.fr

Literature

Alderton I, Palmer BR, Heinemann JA, Pattis I, Weaver L, Gutiérrez-Ginés MJ, Horswell J, Tremblay LA (2021). The role of emerging organic contaminants in the development of antimicrobial resistance. Emerging Contaminants 7, 160–171. doi:10.1016/j.emcon.2021.07.001

Altermatt F, Carraro L, Antonetti M, Albouy C, Zhang Y, Lyet A, Zhang X, Pellissier L (2023). Quantifying biodiversity using eDNA from water bodies: General principles and recommendations for sampling designs. Environmental DNA 5, 671–682. doi:10.1002/edn3.430

Arturi K, Hollender J (2023). Machine Learning-Based Hazard-Driven Prioritization of Features in Nontarget Screening of Environmental High-Resolution Mass Spectrometry Data. Environmental Science & Technology 57, 18067–18079. doi:10.1021/acs.est.3c00304

Aurich D, Diderich P, Helmus R, Schymanski EL (2023). Non-target screening of surface water samples to identify exposome-related pollutants: a case study from Luxembourg. Environmental Sciences Europe 35. doi:10.1186/s12302-023-00805-5

Bunholi IV, Foster NR, Casey JM (2023). Environmental DNA and RNA in aquatic community ecology: Toward methodological standardization. Environmental DNA. doi:10.1002/edn3.476

Cordier T, Forster D, Dufresne Y, Martins CIM, Stoeck T, Pawlowski J (2018). Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. Molecular Ecology Resources 18, 1381–1391. doi:10.1111/1755-0998.12926

- Duarte S, Leite BR, Feio MJ, Costa FO, Filipe AF (2021). Integration of DNA-based approaches in aquatic ecological assessment using benthic macroinvertebrates. *Water (Switzerland)* 13. doi:10.3390/w13030331
- Hamilton PB, Cowx IG, Oleksiak MF, Griffiths AM, Grahn M, Stevens JR, Carvalho GR, Nicol E, Tyler CR (2016). Population-level consequences for wild fish exposed to sublethal concentrations of chemicals – a critical review. *Fish and Fisheries* 17, 545–566. doi:10.1111/faf.12125
- Hollender J, Schymanski EL, Ahrens L, Alygizakis N, Béen F, Bijlsma L, Brunner AM, Celma A, Fildier A, Fu Q, Gago-Ferrero P, Gil-Solsona R, Haglund P, Hansen M, Kaserzon S, Kruve A, Lamoree M, Margoum C, Meijer J, Merel S, Rauert C, Rostkowski P, Samanipour S, Schulze B, Schulze T, Singh RR, Slobodnik J, Steininger-Mairinger T, Thomaidis NS, Togola A, Vorkamp K, Vulliet E, Zhu L, Krauss M (2023). NORMAN guidance on suspect and non-target screening in environmental monitoring. *Environmental Sciences Europe* 35. doi:10.1186/s12302-023-00779-4
- Hu H, Wei X-Y, Liu L, Wang Y-B, Jia H-J, Bu L-K, Pei D-S (2023). Supervised machine learning improves general applicability of eDNA metabarcoding for reservoir health monitoring. *Water Research* 246. doi:10.1016/j.watres.2023.120686
- Huynh N, Caupos E, Soares Peirera C, Le Roux J, Bressy A, Moilleron R (2021). Evaluation of Sample Preparation Methods for Non-Target Screening of Organic Micropollutants in Urban Waters Using High-Resolution Mass Spectrometry. *Molecules* 26, 7064. doi:10.3390/molecules26237064
- Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke SJ (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews* 94, 849–873. doi:10.1111/brv.12480
- Reji Chacko M, Altermatt F, Fopp F, Guisan A, Keggin T, Lyet A, Rey P-L, Richards E, Valentini A, Waldock C, Pellissier L (2023). Catchment-based sampling of river eDNA integrates terrestrial and aquatic biodiversity of alpine landscapes. *Oecologia* 202, 699–713. doi:10.1007/s00442-023-05428-4
- Sade J, Le Roux J, Mottelet S, Guérin-Rechdaoui S, Moilleron R (2022). Caractérisation de la contamination des eaux de surface par le couplage d'analyses non-ciblées en spectrométrie de masse avec des analyses d'écotoxicologie - Rapport d'avancement du projet MeSeine Innovation.
- Santos LHMLM, Araújo AN, Fachini A, Pena A, Delerue-Matos C, Montenegro MCBSM (2010). Ecotoxicological aspects related to the presence of pharmaceuticals in the aquatic environment. *Journal of Hazardous Materials* 175, 45–95. doi:10.1016/j.jhazmat.2009.10.100
- Sieber G, Drees F, Shah M, Stach TL, Hohrenk-Danzouma L, Bock C, Vosough M, Schumann M, Sures B, Probst AJ, Schmidt TC, Beisser D, Boenigk J (2023). Exploring the efficacy of metabarcoding and non-target screening for detecting treated wastewater. *Science of The Total Environment* 903, 167457. doi:10.1016/j.scitotenv.2023.167457
- Slobodnik J, Hollender J, Schulze T, Schymanski EL, Brack W (2019). Establish data infrastructure to compile and exchange environmental screening data on a European scale. *Environmental Sciences Europe* 31, 65. doi:10.1186/s12302-019-0237-6
- Warter MM, Tetzlaff D, Ring A-M, Christopher J, Kissener HL, Funke E, Sparmann S, Mbedi S, Soulsby C, Monaghan MT (2024). Environmental DNA, hydrochemistry and stable water isotopes as integrative tracers of urban ecohydrology. *Water Research* 250. doi:10.1016/j.watres.2023.121065