# Dynamics of dissolved and particulate organic matter along the land-sea continuum of the Seine Estuary (France)

Zhe-Xuan Zhang

▶ **To cite this version:**

Zhe-Xuan Zhang. Dynamics of dissolved and particulate organic matter along the land-sea continuum of the Seine Estuary (France). Earth Sciences. Sorbonne Université, 2023. English. NNT : 2023SORUS268 . tel-04264777

**HAL Id: tel-04264777**

**https://theses.hal.science/tel-04264777v1**

Submitted on 30 Oct 2023

**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.

Sorbonne Université

Ecole doctorale 398 Géosciences Ressources Naturelles et Environnement

*UMR 7619 METIS*

# Dynamics of dissolved and particulate organic matter along the land-sea continuum of the Seine Estuary (France)

Par Zhe-Xuan ZHANG

Thèse de doctorat en biogéosciences

Dirigée par Arnaud HUGUET et Edith PARLANTI

Présentée et soutenue publiquement le 12 octobre 2023

Devant un jury composé de :

| | |
|---|---|
| M. Piotr KOWALCZUK, Professeur | Rapporteur |
| M. Vincent GROSSI, Directeur de recherche CNRS | Rapporteur |
| Mme Laurence LESTEL, Directrice de recherche CNRS | Examinateur |
| Mme Anniet LAVERMAN, Directrice de recherche CNRS | Président du jury |
| M. Arnaud HUGUET, Chargé de recherche CNRS | Directeur de thèse |
| Mme Edith PARLANTI, Chargé de recherche CNRS | Directrice de thèse |

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

This thesis is dedicated to my family,

to Paris and Bordeaux,

to the Seine,

and finally,

to Han, with love

*"All the rivers run into the sea;*

*yet the sea is not full;*

*unto the place from whence the rivers come, thither they return again."*

Ecclesiastes

“水流心不競，雲在意俱遲”

唐·杜甫 《江亭》

# Acknowledgements

I am grateful for the experiences, growth, and knowledge I have gained in France during these busy and beautiful years. To begin, I want to sincerely thank my PhD supervisors, Arnaud and Edith. I am very thankful for their suggestions on experiments, posters, oral presentations, figures, manuscripts, future plans, and more. They are very busy, but they always manage to make time for me. Together, we have had numerous meetings ($n$=39), each often spanning an entire morning or afternoon. I am very lucky to have them as my PhD supervisors.

During my stay in Bordeaux, Mahaut helps a lot with the PARAFAC modeling and the use of Aqualog and Jasco. Thank you for your guidance.

I had a wonderful time in the city of Rouen and all these charming small towns like Les Andelys during my first sampling campaign. I appreciate the assistance of SARTRE members throughout the sampling process. A special thank you goes to Robert, who drove me to the 'sea', passing through the marvelous Pont de Tancarville. Our destination, Le Havre, became a city I returned to multiple times during some weekends.

Thank you, SARTRE and RUNTIME project. Thank you, CSC. Thank you for making this PhD thesis possible.

I would like to thank these awesome PhDs in Metis (Louis, Hugo, and Alienor). Thanks for the artwork: Ceci n’est pas une P.O.M. I will put it on my working desk wherever I go in the future. It is not a ‘pomme’ and not a ‘POM’ with a pore size of 0.5 μm (although I use 0.7 μm to separate DOM and POM in this thesis). It reminds me that appearances might not always reflect the truth, and I'll carry this philosophy with me.

Thank you so much Louis, my office mate and dear friend. I seriously cannot imagine my PhD life without you. Thank you for correcting my French emails and offering help with various registration/administrative tasks and many other things.

Thank you Christelle. Without your help with all these naughty instruments (GC-MS and LC-MS), I cannot finish my PhD. You are my super hero!

Many thanks to my other colleagues in corridor 56/66, Thanh-Thuy, Katell, Emmanuel, Maryse, Marie, Diane, Zheng, and others. I enjoy these lunch parties and the great food you prepare. I was not a fan of dessert before. But after living in France for several years, I started to try different kinds of dessert (and cheese). It's a new world!

Thanks to Valérie, Jean-Marie, and Montse in METIS for taking care of all the administrative matters.

Thank you, my lunchmates and friends, Shuaitao, Haoliang, and Xingcheng. I had great time with you guys!

Thank you, Mia, my first roommate and first good friend in Paris. I always miss the time we spent together to cook very healthy vegetarian food during the lockdown days. We talked a lot about art, films, and photography in that small kitchen. Thank you for your company!

Thank you, Yuye. Thanks for all the madeleines during many hungry afternoons. They will be my Proustian moment!

Thank you, Mom, for always supporting my decisions with endless love. While supporting my choice to study abroad, you continue to worry about me. Don't worry mom! I actually enjoy my PhD life very much!

I would also like to thank Le Grand Action, Ecoles Cinéma Club, La Filmothèque du Quartier Latin, Christine Cinéma Club, Reflet Médicis, Le Champo, UGC Ciné Cité Les Halles, and MK2 Beaubourg. These cinemas are very important part of my life during the three years.

Thank you, John Cassavetes, Michelangelo Antonioni, Stanley Kubrick, Martin Scorsese, Kelly Reichardt, Wes Anderson, Andrei Tarkovsky, Ingmar Bergman, Joanna Hogg, and Federico Fellini. These directors let me know that you must enjoy and love what you choose to do in order to produce valuable and memorable work. I hope I can pursue my career in the same way.

Finally, if you happen to see here, Han. Thank you so much for your support, encouragement and understanding during these years. I hope I can spend more time with you in the future. You are my river, my Seine, flowing through my life, calm and turbulent.

Zhe-Xuan Zhang

Paris, France

20/07/2023

# Table of contents

# General Introduction

Estuaries are highly dynamic and productive ecosystems, linking continents and oceans. Organic Matter (OM) represents a complex mixture of organic compounds, which is channeled through the estuaries into the costal seas. During transit, OM from different sources can be removed and/or processed within estuaries, and they can have a significant impact on the microbial loop as well as biogeochemical cycles (Brankovits et al., 2017; Canuel and Hardison, 2016; McCallister et al., 2006). Such processes are closely linked to OM composition (Derrien et al., 2019). Characterizing OM is thus a major environmental concern, which is important for monitoring and controlling the estuarine water quality.

Categorized by size, OM can be operationally divided into Dissolved Organic Matter (DOM) and Particulate Organic Matter (POM). Both of these carbon pools play important roles in nutrient dynamics and biogeochemical processes in estuaries (Bianchi, 2007; Bianchi and Canuel, 2011; Canuel and Hardison, 2016; Derrien et al., 2019). However, most studies investigated dynamics of estuarine DOM and POM separately. Simultaneous characterization of DOM and POM in estuaries is still lacking and should be prioritized, as each OM pool has its own properties (Thibault et al., 2019). The compositions of POM and DOM within estuaries can undergo spatial and temporal fluctuations due to factors such as the mixing of water masses, seasonality, and complex transformation processes (Bittar et al., 2016; Guo et al., 2014; Xie et al., 2018; Chupakova et al., 2018; Guo et al., 2019). Such variability makes the characterization of estuarine POM and DOM especially challenging. This further complicates our understanding of the ecological functioning of estuaries, notably how they regulate the different types of DOM and POM.

Studying the biogeochemical functioning of human-impacted estuaries may facilitate sustainable management of these essential ecosystems. This is particularly relevant to the Seine Estuary (France), which is a human-impacted estuary and is important from ecological,

economical, and biogeochemical points of view (Romero et al., 2019). Before undertaking costly restoration operations, it is critical to assess the functioning of this estuary, particularly in terms of DOM and POM dynamics, which influence global biogeochemical cycles.

Within the framework of the SARTRE (GIP Seine-Aval) and RUNTIME (EC2CO CNRS/INSU/OFB) projects, the aim of this PhD thesis is to evaluate the ecological role of the Seine estuary in the regulation of different types of DOM and POM, and assess the impact of natural (i.e. hydroclimate conditions) and anthropogenic (i.e. land use characteristics) changes on this role. This PhD manuscript consists of six chapters:

Chapter 1 provides a literature review on estuaries and the dynamics of estuarine POM and DOM. Different approaches to characterize POM and DOM are also reviewed, including elemental and isotopic analysis, lipid biomarkers, as well as absorbance and fluorescence spectroscopy.

Chapter 2 presents the sampling strategy and the different analytical and statistical techniques used in this thesis.

Chapter 3 investigates the POM dynamics across the Seine River basin at the bulk and molecular scales, through elemental and isotopic analyses as well as lipid biomarkers (i.e. bacterial tetraethers). A novel proxy, Riverine IndeX (RIX) is proposed in this chapter to trace riverine organic matter inputs, which can be broadly applicable in modern samples as well as in paleorecords. This chapter is organized for peer–reviewed publication and will be submitted to *Biogeosciences* in September 2023.

Chapter 4 explores the fate of different types of POM along the estuary and investigates the relationships between POM composition and hydroclimate conditions/land use using bulk analysis and complementary lipid biomarkers (i.e. sterols, stanols, fatty acids, and *n*-alkanes). The Seine estuary is further categorized into 3 zones based on variations of distinct types of POM. This chapter is in preparation for submission to *Chemical Geology*.

Chapter 5 investigates the sources, transformations, and fate of DOM in the Seine estuary using absorbance and fluorescence spectroscopy. DOM composition along this human-impacted estuary is further disentangled by unsupervised and supervised machine learning as well as explainable artificial intelligence. A machine learning model (light Gradient Boosting Machine classification for DOM, GBM_DOM) is developed in this chapter to classify estuarine zonation and identify main DOM characteristics within each zone. This chapter is organized as an article for submission to *Science of The Total Environment*.

Chapter 6 presents conclusions of this PhD thesis and synthesizes a conceptual model to assess the functioning of estuarine ecosystems in terms of DOM and POM dynamics across various land use types under high and low-flow scenarios.

# Chapter 1:

State of the art

## 1.1 Estuary and the carbon cycling

According to Fairbridge's definition, an estuary is an inlet of the sea reaching a river valley as far as the upper limit of tidal rise (Fairbridge, 1980). It represents the major boundary that link the continents to oceans (Figure 1-1), which is a highly dynamic and productive zone (Canuel et al., 2012). Large estuaries are common in low relief coastal regions (i.e. the east coast of North America and broad coastal plains of Europe), whereas they are much less common in uplifted coastlines (i.e. the Pacific edge of South and North America) (Day Jr. et al., 2012).

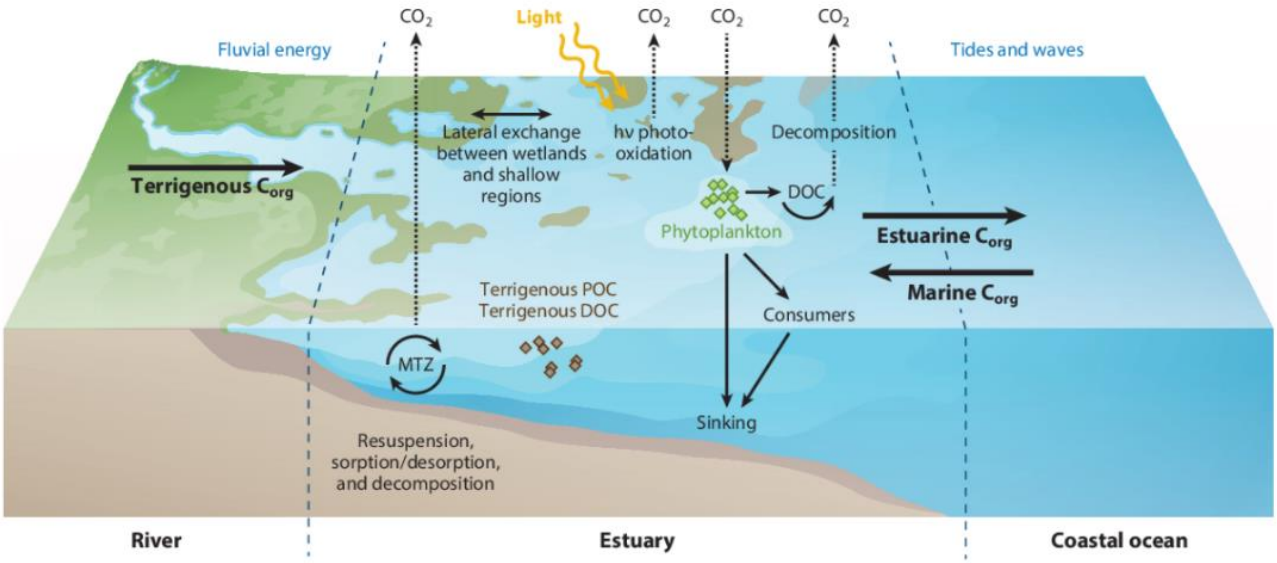



**Figure 1-1.** Carbon cycling in estuarine systems (Canuel and Hardison, 2016)

Estuaries are important zones from ecological, economical, and biogeochemical points of view. They play an essential role in the production and transformation of Organic Matter (OM) (Bianchi, 2007; Cai, 2011; Canuel and Hardison, 2016). OM exported from land is channelled through the estuaries into the coastal seas. During transit, OM is modified through abiotic (Aufdenkampe et al., 2001) and biotic (Sobczak et al., 2002) processes prior its export to the sea.

In addition, continental shelves are locations of terrestrial OM degradation/ sedimentation, but they also serve as a $CO_2$ sink (0.25±0.25 Pg C $y^{-1}$) (Cai, 2011). At a global inventory of~662±32 Pg C, marine dissolved organic carbon (DOC) exceeds the carbon inventory of ocean organic biomass by a factor of 200, making it one of the ocean's greatest bioreactive carbon pools (Hansell et al., 2009, 2012). In addition, carbon burial rates in estuarine and oceanic systems equal 237.6 ± 45.4 TgC (Nellemann and Corcoran, 2009). Estuaries have the potential to contribute as much as 81 Tg of carbon per year to organic carbon burial rates, corresponding to approximately 64% of the organic carbon burial rates found in the coastal ocean (Canuel et al., 2012). Hence, estuaries are important carbon storage and exchange sites between the land, ocean and atmosphere, and they play an important part in the global carbon cycle (Bianchi, 2007; Canuel et al., 2012; Cai, 2011). However, the magnitudes and mechanisms of the modifications of organic matter within estuaries remain poorly understood (Bianchi, 2007; Derrien et al., 2019). Global carbon budgets and accurate climate models for climate change prediction need an improved understanding of carbon cycling within estuary systems (Cai, 2011; Canuel and Hardison, 2016).

## 1.2 Organic matter in estuaries

Organic matter is a complex and heterogeneous mixture of organic compounds, which is made up of a large number of molecules from different classes (e.g. proteins, tannins, lignin, aromatic compounds, carbohydrates and polysaccharides, saturated and unsaturated hydrocarbons) with distinct size, bioavailability, and polarity (hydrophilic, hydrophobic, and transphilic) (Antony et al., 2017; Derrien et al., 2019; Leenheer and Croué, 2003; Volkman and Tanoue, 2002).

Estuarine waters are highly dynamic and productive systems, which receive significant external (allochthonous) material as well as local (autochthonous) production. The allochthonous

OM has different natural sources, from land (i.e. soils, terrestrial plants, and leaves) and the atmosphere (i.e. dust storms) (Derrien et al., 2018; Hamza, 2021; Lee et al., 2020; Osburn et al., 2015; Wei et al., 2009). Anthropogenic activities are also important allochthonous OM sources, including untreated and/or treated sewage (He et al., 2018; Meng et al., 2013), industrial wastewaters (Hao et al., 2021), and oil spills (Zhou et al., 2013). On the other hand, the autochthonous OM is formed within water bodies and processed in the aquatic food web. Autochthonous OM can be derived from aquatic biota (e.g. algae and bacteria) as well as viruses *via* viral lysis (Castillo et al., 2010; Kuhlisch et al., 2021; Patriarca et al., 2021). OM from distinct sources can be further removed and/or processed within estuaries, and they can have a significant impact on the biogeochemical cycle and microbial loop (Brankovits et al., 2017; Canuel and Hardison, 2016; McCallister et al., 2006).

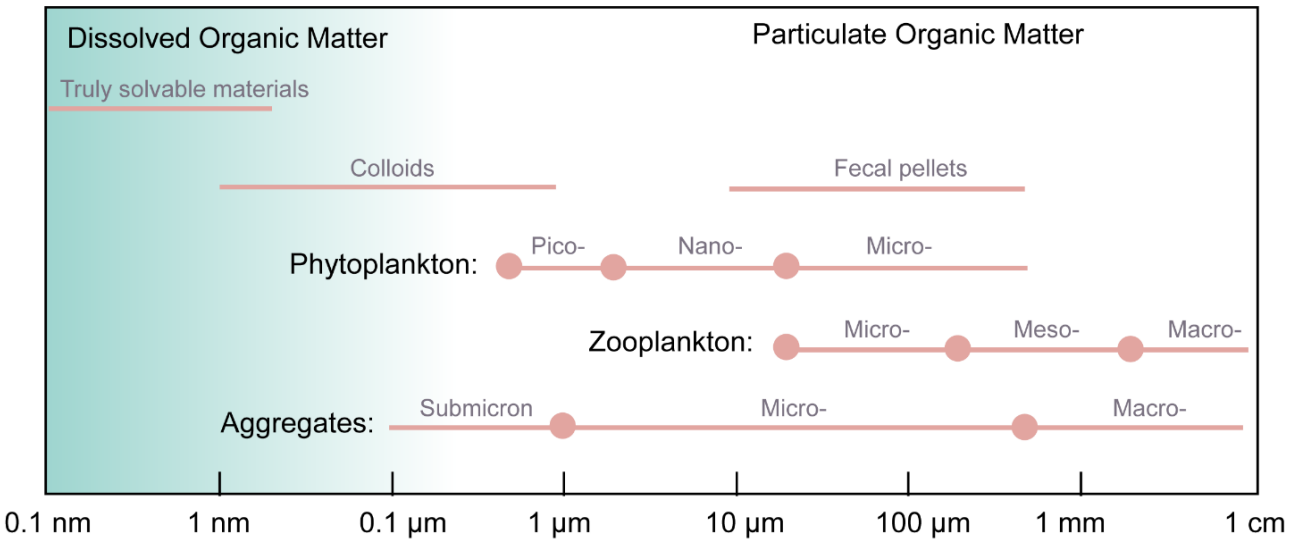



**Figure 1-2**. Diagram showing the size distribution of natural organic matter in aquatic systems adapted after Monroy et al. (2017)

OM in aquatic systems can be divided into Dissolved Organic Matter (DOM) and Particulate Organic Matter (POM) based on the size of the OM (Figure 1-2). In addition, there is a continuum between DOM and POM, termed as colloids (Figure 1-2). DOM (including most

colloids) and POM have been separated using the relatively arbitrary criteria of filter pore sizes ranging from 0.1 to 0.7 μm (Asmala et al., 2013). In marine studies, the Whatman GF/F glass fiber filters with the pore size of 0.7 μm are commonly used for filtration according to practical considerations, as these filters are non-contaminating, easy to clean and have great flow characteristics (Repeta, 2015).

Exchanges between DOM and POM are regulated by processes including flocculation, aggregation, sorption, solubilization, and microbial degradation (Figure 1-3) (Derrien et al., 2019; He et al., 2016). Additionally, DOM and POM are subject to photochemical processes, thus mediating the production of Reactive Oxygen Species (ROS), carbon monoxide (CO), and carbon dioxide ($CO_2$) (Figure 1-3) (Porcal et al., 2015; Vione et al., 2014; Wolf et al., 2018). Such processes are controlled by distinct environmental conditions (i.e. temperature, salinity, pH, and solar irradiance) (He et al., 2016; Porcal et al., 2013).

Most studies performed until now investigated estuarine DOM and POM separately. Simultaneous DOM and POM characterization in estuaries is still lacking and should be prioritized, as each OM pool has its own properties and dynamics (Thibault et al., 2019). Hence, characterizing DOM and POM simultaneously is important for understanding the exchange processes and overall dynamics of OM in such complex systems (Derrien et al., 2019; He et al., 2016).

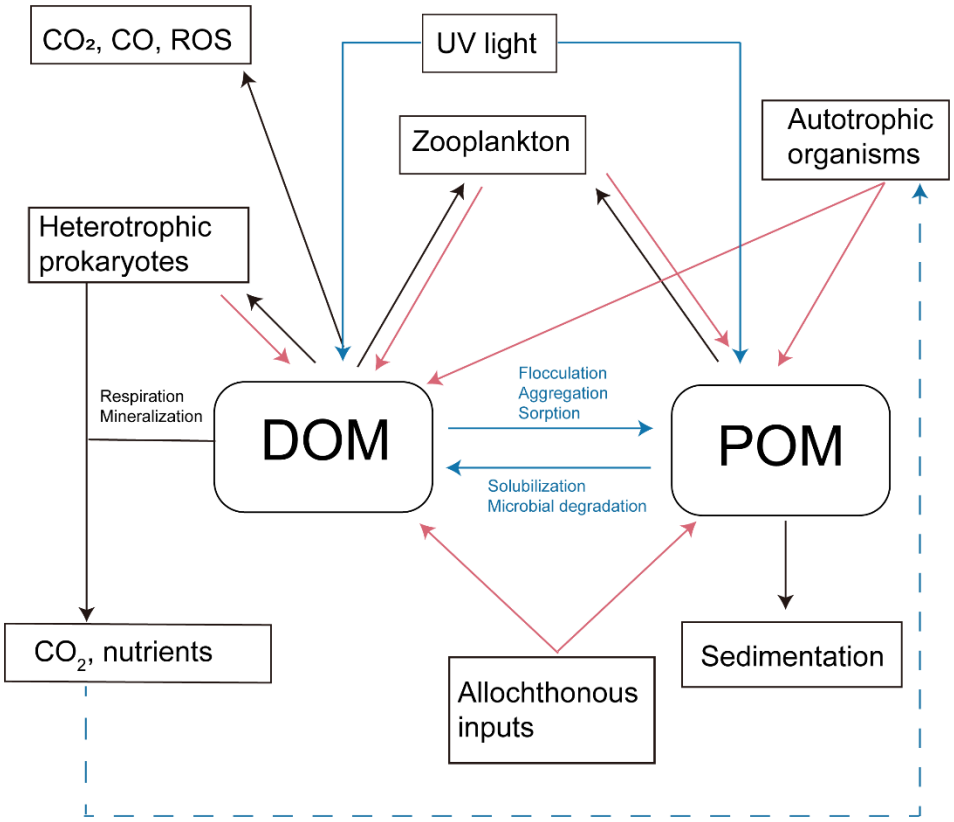



**Figure 1-3.** Controlling factors of DOM and POM in aquatic environment adapted after Derrien et al. (2019). Sources, sinks, and transformation processes are indicated by red arrows, black arrows, and blue arrows, respectively. The dotted line indicates the recycled inorganic nutrient pathway.

## 1.3 POM characterization

Understanding the composition, transport, and fate of POM in estuaries is essential for assessing carbon budget, nutrient dynamics, and biogeochemical processes (Bianchi, 2007; Bianchi and Canuel, 2011; Canuel et al., 2012). The characterization of POM can be determined using a combination of bulk geochemical analysis and lipid biomarkers (Bianchi, 2007; Bianchi and Canuel, 2011).

Bulk geochemical analysis involves the measurement of various chemical properties of POM samples, including total organic carbon content, total nitrogen content, and their stable

isotopic ratios. These analyses provide information about the global characteristics of POM (Bianchi and Canuel, 2011).

On the other hand, lipid biomarkers are more specific organic molecules, which are defined as compounds synthesized by organisms that are extractable by organic solvents, but insoluble in water. They are derived from distinct sources, including soils, plants, algae, bacteria, and other organisms. The complementary use of lipid biomarkers, such as tetraethers lipids, sterols, stanols, fatty acids, and *n*-alkanes, can indicate the OM contributions from allochthonous or autochthonous sources, which will be described in detail in this chapter.

### 1.3.1 Bulk geochemical parameters

#### 1.3.1.1 Elemental analyses

The atomic C to N ratios can provide basic information about the sources of POM (Lamb et al., 2006). Generally, C/N ratios in terrestrial plants are higher than those in aquatic organisms. For example, the vascular plants are composed of carbon-rich compounds (e.g., lignin), leading to high C/N ratios (>17). In contrast, microalgae are composed of protein-rich compounds, leading to low C/N ratios (5 to 7). However, caution should be taken when using C/N ratios in aquatic systems as C/N ratios could be influenced by complex processes. For example, the presence of inorganic nitrogen and the selective degradation of amino acids during diagenetic processes could affect C/N ratios, leading to uncertainties in assessing POM sources using this proxy (Lamb et al., 2006; Müller, 1977).

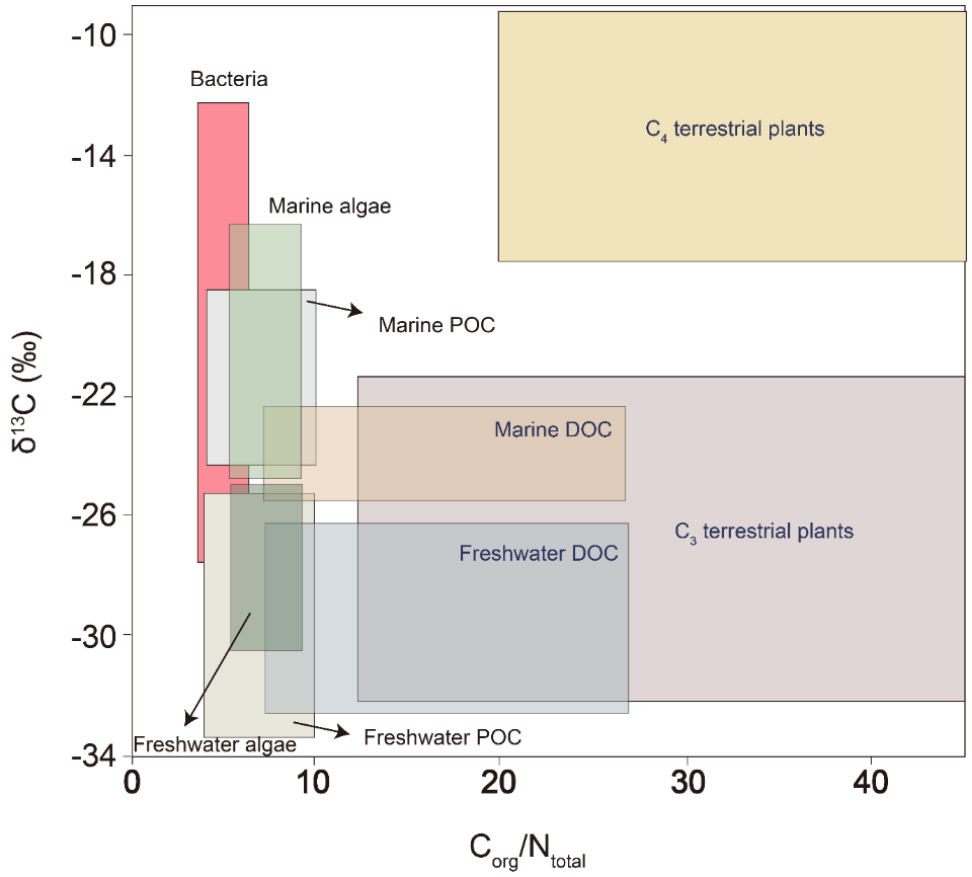



**Figure 1-4.** Typical C/N ratios and $\delta^{13}C$ in organic matter adapted after Lamb et al. (2006)

### 1.3.1.2 Stable carbon isotopic composition ($\delta^{13}C$)

As a supplement to C/N ratios, stable carbon isotopic analysis is also widely used for tracing organic matter sources. Carbon has two stable isotopes, including $^{12}C$ (the most abundant form with 6 protons and 6 neutrons in its nucleus) and $^{13}C$ (much less abundant form with 6 protons and 7 neutrons). To express the proportion of $^{13}C$ in the environment, the notation $\delta$ is used to quantify the difference between the $^{13}C/^{12}C$ ratio of a sample relative to an international standard: the Vienna Pee Dee Belemnite (VPDB; Eq. 1).

$$\delta^{13}C(‰) = \left( \frac{\left(\frac{^{13}C}{^{12}C}\right)_{\text{sample}}}{\left(\frac{^{13}C}{^{12}C}\right)_{\text{VPDB}}} - 1 \right) \times 1000 \qquad (1)$$

Generally, aquatic organisms have more positive $\delta^{13}C$ values than terrestrial plants (Lamb et al., 2006). Based upon this, the $\delta^{13}C$ has been used for determining the proportions of distinct sources in the mixture, using mixing models (Moore and Semmens, 2008). However, overlaps in $\delta^{13}C$ (and C/N ratios) of distinct source materials could complicate the applications of these bulk geochemical proxies in complex environments (Figure 1-4). Additionally, range of values in $\delta^{13}C$ values can be considerable, and terrestrial C4-plants can have higher $\delta^{13}C$ values (−12‰), which may lead to inaccuracies in calculating the relative amounts of terrestrial organic matter in marine systems using $\delta^{13}C$ values of TOC (Hedges et al., 1997). For example, C/N ratio showed an obvious decreasing trend from land to sea along the Dagu River-estuary-marine system, suggesting decreasing terrestrial input from upstream to downstream Liu et al., 2021). However, no obvious trend was observed for $\delta^{13}C$ values, which was interpreted by the significant influence of the heavier $\delta^{13}C$ of C4 plants (Liu et al., 2021). The bulk proxies should therefore be supported by lipid biomarkers, which could provide reliable source assessment of organic matter in estuaries.

Nitrogen (N) is an essential component of proteins, nucleic acids, and other biomolecules that are important for the growth and development of organisms. The availability of N can control the rate of primary production on a variety of temporal and spatial scales, which can influence the structure and functioning of ecosystems (Casciotti, 2016).

Stable isotopes of nitrogen serve as powerful tools for understanding the sources, transformations, and fates of nitrogen compounds (Denk et al., 2017). The most abundant nitrogen isotope is $^{14}N$ (99.67%), whereas the less abundant nitrogen isotope is $^{15}N$, which makes up about

0.33% of nitrogen in the natural environment. Similar to stable carbon isotope, the notation δ is used to quantify the difference between the $^{15}N/^{14}N$ ratio of a sample relative to air (Eq. 2).

$$\delta^{15}\mathrm{N}(‰) = \left( \frac{\left(\frac{^{15}N}{^{14}N}\right)_{\mathrm{sample}}}{\left(\frac{^{15}N}{^{14}N}\right)_{\mathrm{air}}} - 1 \right) \times 1000 \qquad (2)$$

Nitrogen isotopes ($\delta^{15}N$) are influenced by complex microbial processes (i.e. nitrogen fixation, nitrification, and denitrification) and anthropogenic activities (Bianchi and Canuel, 2011; Casciotti, 2016; Denk et al., 2017). While $\delta^{15}N$ analysis can provide valuable insights into the sources and cycling of nitrogen in aquatic systems, its interpretation can be complex and requires careful consideration (Bianchi and Canuel, 2011).

Despite these challenges, the nitrogen isotopes are still considered to be a useful tool to study OM sources in estuaries. For example, the $\delta^{15}N$ values of suspended matter are much higher (18-24 ‰) during the flowering periods of phytoplankton in the Scheldt estuary, which was interpreted as the nitrogenous nutrient assimilated by phytoplankton (Mariotti et al., 1984).

### 1.3.2 Lipid biomarkers

Bulk geochemical proxies can be influenced by decomposition processes, remineralization, and distinct terrigenous sources, which could inevitably complicate their applications (Lamb et al., 2006). To overcome the limitations of the bulk proxies, source-specific lipid biomarkers can be applied for tracing the sources of organic matter. The latter could provide more reliable information on POM sources compared with bulk parameters (Bianchi and Canuel, 2011).

#### 1.3.2.1 GDGTs

Glycerol dialkyl glycerol tetraethers (GDGTs) are membrane lipids of Archaea and some Bacteria. They occur ubiquitously in a wide range of terrestrial and aquatic environments, including soils (Hopmans et al., 2004), lakes (Tierney and Russell, 2009), marine settings (Schouten et al., 2002), peats (Sinninghe Damsté et al., 2000), cold seeps (Zhang et al., 2020), hydrothermal vents (Hu et al., 2012), and estuaries (Wu et al., 2014). Generally, based on structures and sources, GDGTs are divided into two groups (Figure 1-5), isoprenoid GDGTs (isoGDGTs) and branched GDGTs (brGDGTs). The isoGDGTs (cf. structures in Figure 1-5) are synthesized by Archaea, whereas the brGDGTs are produced by some Bacteria according to the distinct stereoconfiguration of their glycerol moieties and alkyl chains (Weijers et al., 2006). The exact producers of brGDGTs are not identified yet, even though some of them are attributed to the phylum *Acidobacteria* (Chen et al., 2022; Halamka et al., 2022; Sinninghe Damsté et al., 2011).

The structures of the major isoGDGTs (GDGT-0, -1, -2, -3, Crenarchaeol, and Crenarchaeol') show differences in the number of cyclopentane moieties (0-4) in their alkyl chains (Figure 1-5). It was shown that the number of cyclopentane moieties of the GDGTs produced by cultured hyperthermophillic archaea increased with growth temperature, which was suggested to be a response of the cell membrane for adapting to temperature changes (Gliozzi et al., 1983; Uda et al., 2001). This mechanism thus allowed for reconstructing the temperature at which the isoGDGTs were synthesized in marine settings, by calculating the average number of cyclopentane rings. The $TEX_{86}$ (tetraether index of tetraethers consisting of 86 carbons) proxy, which reflects the aforementioned average number of cyclopentane rings in isoGDGTs, was linearly correlated to sea surface temperatures (Schouten et al., 2002). This isoGDGT-based proxy is now used globally

for paleotemperature reconstructions in marine as well as lacustrine settings on a wide variety of timescales (Schouten et al., 2013).

The brGDGTs were shown to be produced by unknown heterotrophic bacteria (Blewett et al., 2022; Huguet et al., 2017; Weijers et al., 2010). Similar to isoGDGTs, brGDGTs also contain varying numbers of cyclopentane rings, with 0 (suffix a), 1 (suffix b), and 2 (suffix c) cyclopentane moieties. In addition to these cyclopentane moieties, there is also a varying number of methyl groups. Specifically, brGDGTs with tetramethylated (prefix I), pentamethylated (prefix II), and hexamethylated (prefix III) alkyl backbones can be distinguished (Figure 1-5). In soils, the degree of cyclisation of brGDGTs (CBT) was shown to be related with pH, whereas the degree of methylation (MBT) was found to be related to mean annual air temperature (MAAT) and pH in a global soil dataset ($n$=134) (Weijers et al., 2007). This would reflect the adaptation mechanism of brGDGT-producing bacteria to environmental changes based on the hypothesis that the temperature could impact on the membrane fluidity and permeability, whereas the pH could affect the proton gradient across the membrane (Pearson and Ingalls, 2013; Weijers et al., 2007).

The brGDGT-based proxies were largely applied for reconstructing continental temperature and soil pH (Schouten et al., 2013). Recently, the improved chromatography methods allowed for separation of brGDGTs with distinct position of alkyl-chain methylations (De Jonge et al., 2014, 2013; Ding et al., 2016; Hopmans et al., 2016). The 5-methyl (methyl groups at the fifth position), 6-methyl (methyl groups at the sixth position), and 7-methyl (methyl groups at the seventh position) brGDGTs could then be separated and quantified. The fractional abundances of 6-methyl brGDGTs were significantly correlated with soil pH and were excluded from the MBT, leading to the development of MBT'$_{5Me}$. The latter index is only influenced by MAAT and no more by pH (De Jonge et al., 2014), improving the MAAT reconstructions based on this proxy. In recent years,

global MAAT calibrations have been developed both in terrestrial and aquatic settings, with different models and increasing number of samples (Russell et al., 2018; Véquaud et al., 2022).

The $IR_{6Me}$ index represents the proportion of 6-methyl brGDGTs vs. 5-methyl brGDGTs, with high values indicating higher abundance of 6- vs. 5-methyl brGDGTs (De Jonge et al., 2015). In aquatic systems, 6-methyl brGDGTs were considered as being related to *in situ* production (De Jonge et al., 2015; Kirkels et al., 2022b). $IR_{6Me}$ (Eq. 3) was calculated according to De Jonge et al. (2015) with Roman numbers referring to the structures in Figure 1-5.

$$IR_{6Me} = \frac{II_{a_6}+II_{b_6}+II_{c_6}+III_{a_6}+III_{b_6}+III_{c_6}}{II_{a_5}+II_{b_5}+II_{c_5}+II_{a_6}+II_{b_6}+II_{c_6}+III_{a_5}+III_{b_5}+III_{c_5}+III_{a_6}+III_{b_6}+III_{c_6}} \quad (3)$$

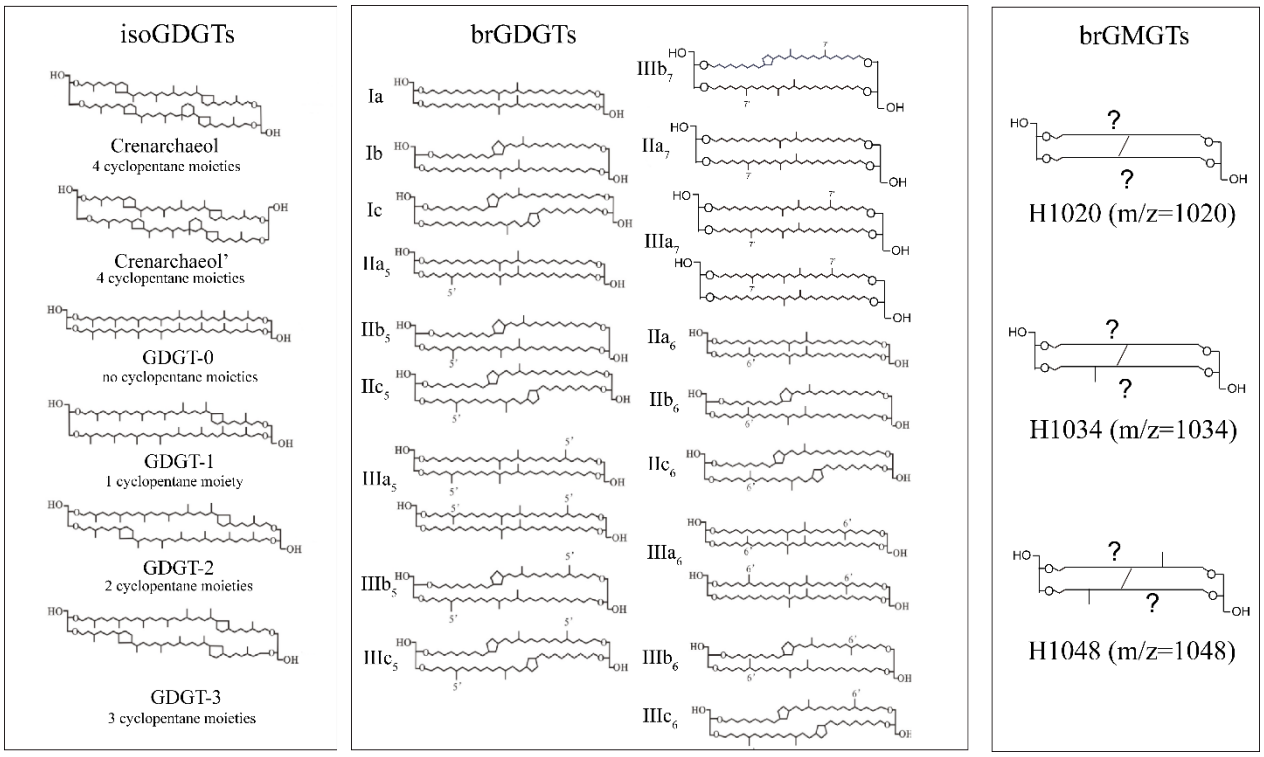



**Figure 1-5.** Structures of the main GDGTs and GMGTs studied

Initially, the brGDGTs were suggested to be mainly produced in soils and one specific isoGDGT, the crenarchaeol, by marine *Thaumarchaeota* (a phylum of Archaea) in aquatic ecosystems (Hopmans et al., 2004). Therefore, the Branched and Isoprenoid Tetraether (BIT)

index, based on the relative abundance ratio of soil-derived brGDGTs to crenarchaeol, was proposed for quantifying river-transported soil organic matter in aquatic environments (Hopmans et al., 2004).

$$\mathrm{BIT} = \frac{\mathrm{I_a}+\mathrm{II_{a_5}}+\mathrm{II_{a_6}}+\mathrm{III_{a_5}}+\mathrm{III_{a_6}}}{\mathrm{I_a}+\mathrm{II_{a_5}}+\mathrm{II_{a_6}}+\mathrm{III_{a_5}}+\mathrm{III_{a_5}}+\mathrm{crenarchaeol}} \qquad (4)$$

BIT values were suggested to be high in soil samples (close to 1) and low in open marine samples (close to 0). Consequently, high BIT values were considered as indicators of substantial soil organic matter inputs to aquatic settings. However, the application of BIT is complicated by the fact that crenarchaeol can be largely produced in soils and *in-situ* produced brGDGTs in aquatic environments (Schouten et al., 2013). Furthermore, the applicability of brGDGT-based proxies in estuaries still remains debatable because of the complex biogeochemical processes and/or long transport distance from land to sea (Cheng et al., 2021). For example, higher BIT values could be observed offshore due to more fluvial-derived terrestrial organic matter buried offshore (Wu et al., 2014). Moreover, the BIT index can be significantly influenced by elevated crenarchaeol concentration, e.g. in the Thames Estuary where it was interpreted as anthropogenic disturbance (Lopes dos Santos and Vane, 2016). Hence, additional molecular proxies for quantifying the riverine OM inputs are still needed.

#### 1.3.2.2 brGMGTs

The branched glycerol monoalkyl glycerol tetraethers (brGMGTs) are a much less studied group of lipids compared to brGDGTs. To date, brGMGTs have been identified in marine sediments (Liu et al., 2012), peats (Elling et al., 2023; Naafs et al., 2018), soils (Baxter et al., 2021), rivers (Kirkels et al., 2022a), and lake sediments (Baxter et al., 2021, 2019).

They are structurally similar to brGDGTs, but possess an additional covalent carbon–carbon bond between the alkyl chains, leading to “H-shaped” structure (Figure 1-5). The bridge of

brGMGTs was considered to be a primary adaptation to heat stress (Baxter et al., 2019; Naafs et al., 2018). Although the rigorous chemical characterization of brGMGTs is lacking and the source organisms of brGMGTs are unknown, the correlations between fractional abundances of brGMGTs and MAAT were still observed, showing their potential as temperature indicators in lakes and peats (Baxter et al., 2019; Naafs et al., 2018).

A recent study indicates that shifts in microbial community composition in response to other unknown environmental factors may also control the production of brGMGTs in peats and lignites (Elling et al., 2023). In order to use the brGMGT as environmental proxies in sedimentary records, it is still important to determine which factors influence their distributions in soils, riverine and marine environments, which are currently poorly understood (Bijl et al., 2021).

#### 1.3.2.3 *n*-alkanes

The straight-chain alkanes (*n*-alkanes) are abundant and common lipid biomarkers from terrestrial plants, aquatic plants and aquatic organisms, with $C_nH_{2n+2}$ as molecular formula (Figure 1-6). The carbon chain length for *n*-alkanes varies depending on source organisms. For example, short-chain *n*-alkanes (C<20) are mainly found in photosynthetic bacteria and algae (Cranwell et al., 1987; Pisani et al., 2013). Middle-chain *n*-alkanes ($C_{20}$-$C_{25}$) are enriched in aquatic plants (submerged and floating aquatic macrophytes) (Cranwell, 1984; Ficken et al., 2000). Long-chain *n*-alkanes (C>25) with a strong odd-to-even carbon preference are predominant in terrestrial higher plants (Ficken et al., 2000; Silva et al., 2012). Hence, sources of organic matter could be distinguished based on *n*-alkane distributions (Derrien et al., 2017). Several proxies based on *n*-alkane distributions have been proposed for identifying OM sources, including Average Chain Length ratio (ACL, also known as Mean Carbon Number), Carbon Preference Index (CPI), and aquatic proxy (Paq).

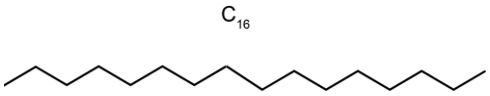



**Figure 1-6**. Structure of $C_{16}$ alkane

ACL (Eq. 5) reflects the average chain length of the *n*-alkanes and is a weighted average of the relative abundance of the different homologs. It can be used to distinguish the *n*-alkanes derived from terrestrial higher plants from those of microbial and algal sources (Derrien et al., 2017). In addition, ACL values can be utilized to predict climate-dependent vegetation change. The vegetation in drier and warmer temperatures biosynthesizes longer chain alkyl lipids than that in temperate settings, with higher ACL values indicating drier and warmer temperatures (Rommerskirchen et al., 2003; Sarkar et al., 2014). Nevertheless, this has no universal validity.

$$ACL = \left[\frac{(\sum C_i \times i)}{C_i}\right] \quad (5)$$

CPI ratio (Eq. 6) is a proxy showing odd carbon forms relative to their even carbon homologs in a certain range of carbon numbers. The *n*-alkanes from terrestrial higher plants usually have higher CPI values (>5). Compared to natural biogenic forms, petroleum-derived *n*-alkanes generally lack the odd carbon predominance. The anthropogenic hydrocarbon pollution in aquatic systems could therefore be identified by low CPI values (close to 1). In addition, natural degradation can also result in a more homogenous distribution of plant-derived (long-chain) *n*-alkanes, as reflected by low CPI values (Zhu et al., 2011).

$$\mathrm{CPI} = \frac{1}{2}\left[\frac{(C_{25}+C_{27}+C_{29}+C_{31}+C_{33})}{(C_{24}+C_{26}+C_{28}+C_{30}+C_{32})} + \frac{(C_{25}+C_{27}+C_{29}+C_{31}+C_{33})}{(C_{26}+C_{28}+C_{30}+C_{32}+C_{34})}\right] \quad (6)$$

$P_{aq}$ (Eq. 7) is calculated based on the relative proportion of two middle-chain *n*-alkanes ($C_{23}$ and $C_{25}$) to two long-chain *n*-alkanes ($C_{29}$ and $C_{31}$) (Ficken et al., 2000; Sikes et al., 2009). It can be used for differentiating the *n*-alkane inputs from terrestrial plants and aquatic plants. For example, $0.01<P_{aq}<0.25$ reflects a predominance of *n*-alkanes derived from terrestrial plants, 0.4–

0.6 from emergent aquatic plant and >0.6 from aquatic plants and marine macrophytes (Ficken et al., 2000; Sikes et al., 2009).

$$P_{aq} = \left[\frac{(C_{23} + C_{25})}{(C_{23} + C_{25} + C_{29} + C_{31})}\right] \quad (7)$$

The *n*-alkane-derived proxies could support source information of organic matter obtained from other biomarkers and bulk proxies. For example, in the Conwy Estuary, $P_{aq}$ was compared with bulk proxies ($\delta^{13}C$ and C/N) and GDGT-derived proxy (BIT) (Lopes dos Santos and Vane, 2020). These authors observed that $\delta^{13}C$ increased in the seaward direction and showed a negative correlation with C/N and BIT and a positive correlation to $P_{aq}$, which supports the seaward declines in the terrestrial contribution from the land to Conwy Bay.

#### 1.3.2.4 Fatty acids

A fatty acid consists of a straight chain with distinct numbers of carbon atoms and a carboxyl group (—COOH) at the end of the chain (Figure 1-7). Fatty acids can be classified into saturated and unsaturated fatty acids (Figure 1-7). The fatty acid is saturated if the carbon-to-carbon bonds are all single. On the other hand, the fatty acid is unsaturated and more labile if one or more double bonds exists between carbon atoms. Unsaturated fatty acids can be further divided into monounsaturated fatty acids (MUFAs) and polyunsaturated fatty acids (PUFAs) based on the number of double bonds.

Fatty acids generally exist in bound forms (esterified forms), free forms, and in combination with biochemical classes (such as glycolipids and lipoproteins). Most fatty acids are present in bound forms in neutral and polar lipids, whereas free fatty acids are less abundant in natural environments (Bianchi and Canuel, 2011).

The nomenclature for fatty acids is based on its carbon chain length, number of double bonds and position of the double bonds. For example, $C_{16:1\omega7}$ indicates that i) the number of carbon atoms is 16, ii) the number of double bonds is 1, and iii) ω7 is the position of the double bond relative to the aliphatic end (Figure 1-7).

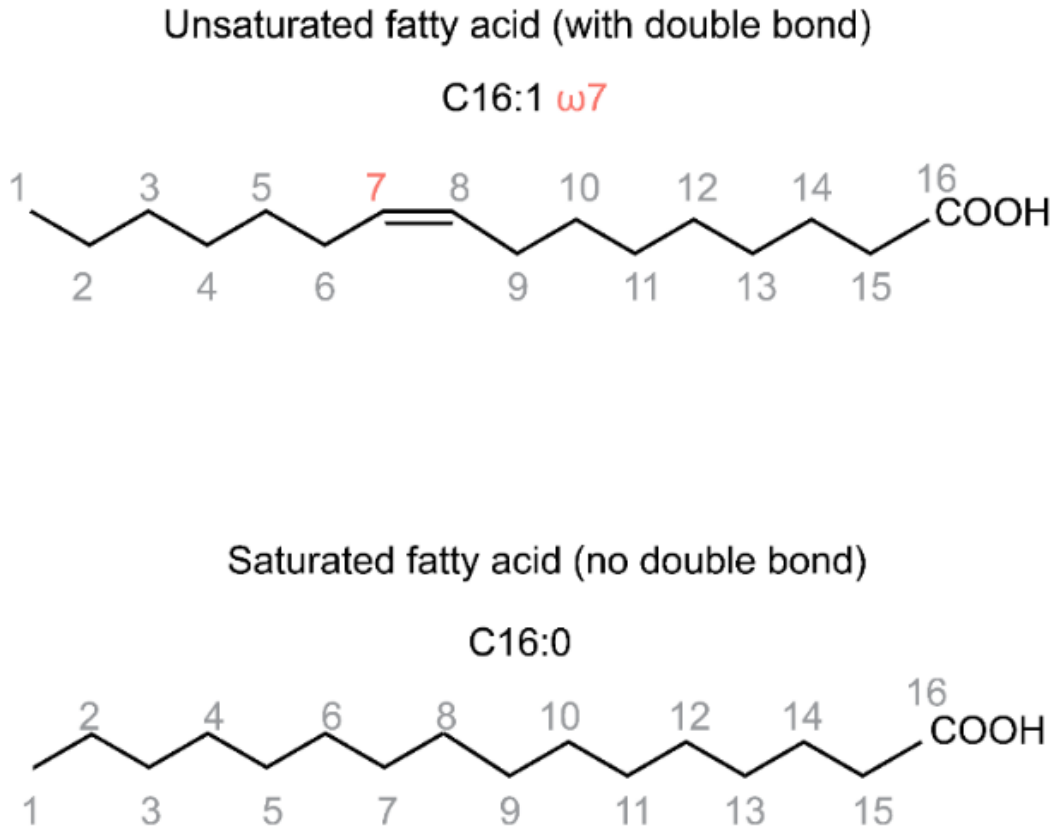



**Figure 1-7**. Structures of saturated and unsaturated fatty acids

Fatty acids are commonly used as lipid biomarkers for microbial ecology, trophic studies, and for characterizing organic matter in aquatic environments (Bianchi and Canuel, 2011). $C_{16:0}$ (palmitic acid) and $C_{18:0}$ (stearic acid) are nonspecific fatty acids, as they are abundant and widely distributed compounds in most organisms in a number of environments (terrestrial, marine, bacteria, and animals). The long chain saturated fatty acids are abundant in terrestrial higher plants ($C_{16}$-$C_{30}$), whereas the short chain ($C_{12}$-$C_{19}$) saturated fatty acids are commonly observed in autochthonous organisms (Bianchi and Canuel, 2011; Kawamura et al., 1987).

Fatty acids are thus valuable tools to trace the sources of natural organic matter in aquatic systems. For instance, as diatoms are characterized by high levels of $C_{16:1}/C_{16:0}$, this ratio can be used as a general diatom biomarker (Budge et al., 2001; Claustre et al., 1989).

Fatty acids have been widely utilized in estuaries as they provide valuable insights into sources and transformations of POM (Bianchi and Canuel, 2011). For example, in the Altamaha estuary, the contents of total fatty acids showed similar variations with the chlorophyll a, highlighting significant algal contribution into the samples (Dai and Sun, 2007).

#### 1.3.2.5 Sterols and stanols

Sterols (cf. structures in Figure 1-8), important lipids for eukaryotes, are synthesized by various types of organisms. For instance, sterol distribution varies across plants and animals, with phytosterols being more frequent in plants and cholesterol predominating in animals (Lagarda et al., 2006; Weete et al., 2010). Specifically, fungi and plants mostly synthesize ergosterols and phytosterols with 28 to 29 carbon atoms (C-28 and C-29 sterols), whereas animals primarily synthesize the cholesterol (C-27 sterol). Given their distinct distribution in various organisms/environments, sterols are widely used as biomarkers for identifying sources of natural organic matter in aquatic systems (Bianchi and Canuel, 2011).

The stigmasterol (stigmasta-5,22E-dien-3β-ol), β-Sitosterol (stigamast-5-en-3β-ol), brassicasterol (ergosta-5,22E-dien-3β-ol), and cholesterol (cholest-5-en-3β-ol) could be used for tracing the sources of natural organic matter. For instance, the stigmasterol and β-Sitosterol are biomarkers for terrestrial organic matter, whereas brassicasterol is a biomarker of aquatic organisms (Moreau et al., 2002; Volkman, 1986).

A recent study suggests that some animals can also synthesize phytosterols (Michellod et al., 2023). These authors found that sitosterol (one of the phytosterols) can be synthesized de novo

by gutless marine annelids. This indicates that, in addition to terrestrial sources, phytosterols may also be produced *in situ* in marine environments.

Stanols are saturated sterols as they have no double bonds in the ring (Figure 1-8). Some stanols are widely used biomarkers for assessing sewage contamination in aquatic environments (He et al., 2018). For example, the coprostanol (5β-cholestan-3α-ol) and epicoprostanol (5β-cholestan-3α-ol) are fecal biomarkers, derived from urban sewage inputs (Carreira et al., 2004; Grimalt et al., 1990; Leeming et al., 1996; Vane et al., 2010). Coprostanol and cholestanol are degradation products of cholesterol and are derived from different pathways. Specifically, coprostanol is derived from cholesterol when digested by omnivorous and carnivorous organisms (sewage contamination), whereas cholestanol is derived from cholesterol by microbial degradation (natural degradation product). Based upon this, the presence of sewage in aquatic environments can be identified using the ratio of coprostanol vs. coprostanol + cholestanol, with values higher than 0.7 as the criteria for sewage contamination (Grimalt et al., 1990). In addition, epicoprostanol is commonly converted from coprostanol by microbial activities and is usually found in digested sludge samples (McCalley et al., 1981). Hence, the presence of epicoprostanol could indicate that the sewage has been microbially degraded or partially digested by wastewater treatment. Based upon this, the level of wastewater treatment can be identified by the epicoprostanol/coprostanol ratio, with values lower than 0.2 for untreated sewage and higher than 0.2 for treated sewage (Mudge and Lintern, 1999).

The proxies based on sterols and stanols have been applied in estuaries for evaluating (i) the sources of natural organic matter (Mudge and Bebianno, 1997) and (ii) fecal contamination (Cordeiro et al., 2008). For example, sterols were analyzed for distinguishing sewage and marine/terrestrial organic matter inputs in the Paranaguá Estuarine System (Martins et al., 2011). These authors use principal component analysis to find distinctions between sterols from marine,

fecal, and terrestrial inputs, showing the applicability of sterols for identifying sources of natural and anthropogenic organic matter.

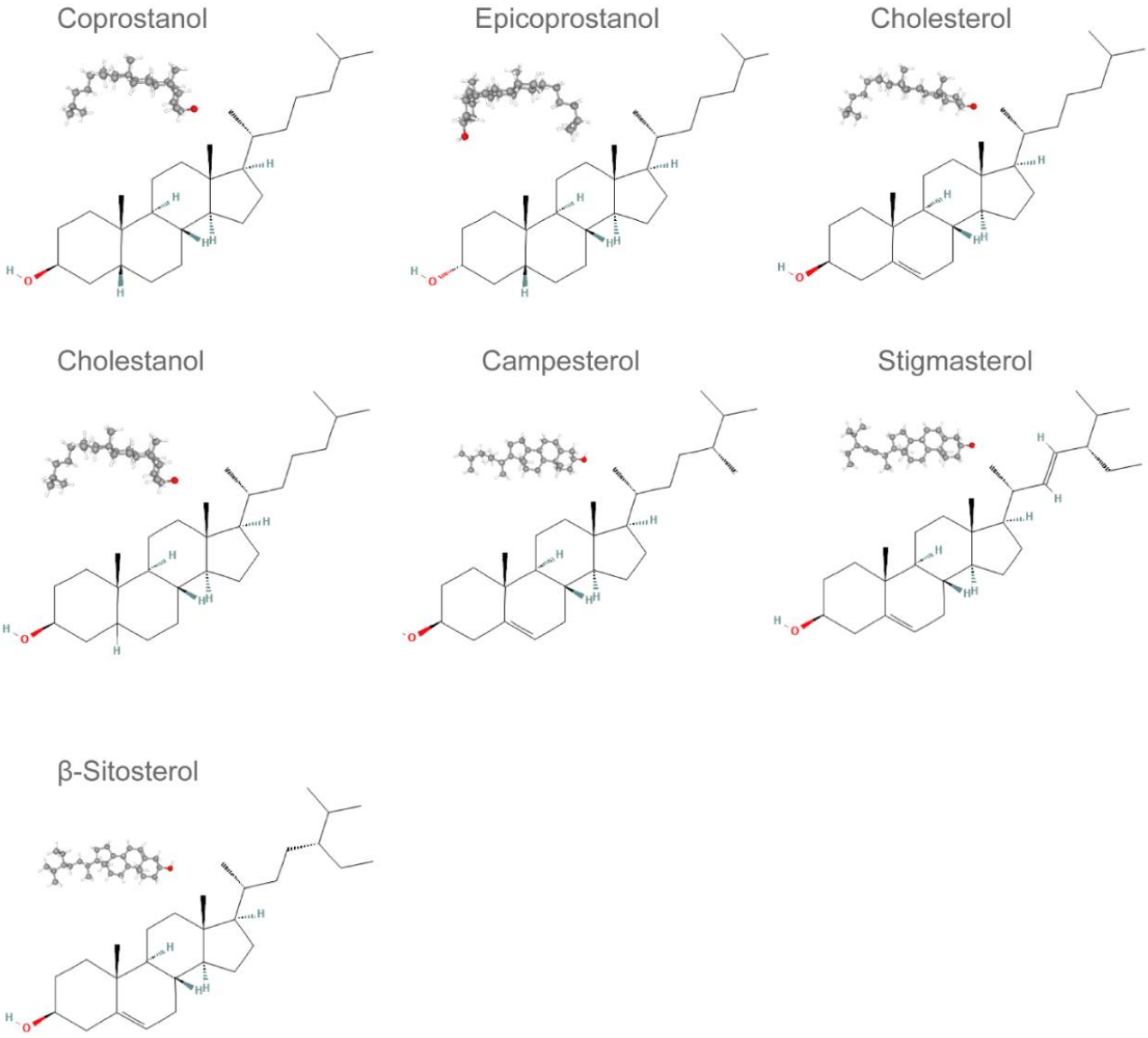



**Figure 1-8.** Structures of sterols and stanols (2D structure and 3D conformer downloaded from https://pubchem.ncbi.nlm.nih.gov)

The biomarker proxies used in this thesis and the information they provide are summarized in Table 1-1.

**Table 1-1.** Biomarker proxies used in this thesis

| Family | Proxy | Calculation | Source assignment | References |
|---|---|---|---|---|
| GDGTs | BIT | $\frac{I_a + II_{a_5} + II_{a_6} + III_{a_5} + III_{a_6}}{I_a + II_{a_5} + II_{a_6} + III_{a_5} + III_{a_5} + \text{crenarchaeol}}$ | Higher values indicate higher inputs of soil-derived OM | (Hopmans et al., 2004) |
| GDGTs | $IR_{6Me}$ | $\frac{II_{a_6} + II_{b_6} + II_{c_6} + III_{a_6} + III_{b_6} + III_{c_6}}{II_{a_5} + II_{b_5} + II_{c_5} + II_{a_6} + II_{b_6} + II_{c_6} + III_{a_5} + III_{b_5} + III_{c_5} + III_{a_6} + III_{b_6} + III_{c_6}}$ | Higher values indicate more *in situ* 6-methyl brGDGT production | (De Jonge et al., 2015) |
| Stanols | S1 | $\frac{\text{Coprostanol}}{\text{Coprostanol} + \text{Cholestanol}}$ | Presence of sewage (S1>0.7) | (Grimalt et al., 1990) |
| Sterols | S2 | $\frac{\text{Brassicasterol}}{\text{Total sterols}}$ | Higher values indicate increasing algal contribution | (Moreau et al., 2002; Volkman, 1986) |
| Fatty acids | F1 | $\frac{C16:1}{C16:0}$ | Higher values indicate diatoms contribution | (Claustre et al., 1989; Parrish et al., 2000) |
| *n*-alkanes | CPI | $\frac{1}{2}\left[\frac{(C_{25} + C_{27} + C_{29} + C_{31} + C_{33})}{(C_{24} + C_{26} + C_{28} + C_{30} + C_{32})} + \frac{(C_{25} + C_{27} + C_{29} + C_{31} + C_{33})}{(C_{26} + C_{28} + C_{30} + C_{32} + C_{34})}\right]$ | Terrestrial higher plants (CPI>5); Petrogenic/marine (CPI≈1) | (Bray and Evans, 1961) |
| *n*-alkanes | ACL | $\left[\frac{(\sum C_i \times i)}{C_i}\right]$ | Terrestrial higher plants (ACL>25); Aquatic plants (20<ACL<25); Microorganisms/plankton (ACL<25) | (Cranwell, 1984; Ficken et al., 2000) |
| *n*-alkanes | $P_{aq}$ | $\left[\frac{(C_{23} + C_{25})}{(C_{23} + C_{25} + C_{29} + C_{31})}\right]$ | Terrestrial plants ($0.01<P_{aq}<0.25$); Emergent aquatic plants ($0.4<P_{aq}<0.6$); Submerged aquatic plants ($P_{aq}>0.6$) | (Ficken et al., 2000; Sikes et al., 2009) |

### 1.3.3 POM dynamics in estuaries and challenges

Evaluating the sources and fate of estuarine POM is crucial for understanding global carbon budget, ecological impacts, and fisheries (Bianchi, 2007; Cai, 2011; Canuel et al., 2012; Carvajalino-Fernández et al., 2020). However, understanding the sources and fate of estuarine POM is especially challenging due to their multiple sources (i.e. riverine, estuarine, and marine contribution) and complex biogeochemical processes (i.e. autochthonous production and degradation) (Bianchi, 2007; Bianchi and Canuel, 2011; Bibi et al., 2020; Goñi et al., 2021). Such complexity could be associated with variations in riverine discharge. For example, higher water discharge can lead to increased riverine input of nutrients and terrestrial loadings of POM into the estuary, whereas low discharge affects the water residence time, thus controlling the settling and degradation of estuarine POM (Bibi et al., 2020; He et al., 2014; Xiong and Shen, 2022).

In addition to the natural factors, human activities (i.e. dam construction) are recently highlighted and are thought to play an important role in regulating the dynamics of estuarine OM (Wang et al., 2022). Such anthropogenic activities may lead to changes in nutrient inputs and primary productivity, thus influencing the processing of estuarine POM. However, it remains unknown how human activities (i.e. land use changes) control distinct types of estuarine POM.

Investigating the complex relationships between land use changes, water discharge, and POM dynamics is crucial for understanding estuarine functioning and sustainable management, which requires multi-proxy approaches that consider both natural processes and human activities. To date, the relationships between land use changes, hydroclimate conditions and POM dynamics are primarily studied at the bulk level (Jeong et al., 2023), hampering a comprehensive understanding of the behaviors of estuarine POM.

## 1.4 DOM characterization

DOM is one of the largest reservoir of organic carbon on Earth, playing a key role in the global carbon cycle and in all biogeochemical cycles (Hansell and Orellana, 2021; Lønborg et al., 2020). It acts as a reservoir for nutrients (i.e. nitrogen and phosphorus), preventing their immediate availability to microbes in the upper water column (Repeta, 2015). DOM also provides energy for microbes (Tranvik, 1992) and influences the transport and bioavailability of essential trace metals (Yamashita and Jaffé, 2008) and organic and inorganic pollutants (Bauer and Blodau, 2006). These biogeochemical processes are closely linked to DOM composition (Derrien et al., 2019). Characterizing DOM is thus a major environmental concern, which is important for monitoring and controlling the water quality.

In the early stages of characterizing DOM, efforts mainly focused on extracting the substances from water to obtain sufficient quantities for chemical analysis (Coble, 2007; Repeta, 2015). Several methods (i.e. resin adsorption recommended by the International Humic Substances Society (IHSS), solid-phase extraction, ultrafiltration, and reverse osmosis coupled with electrodialysis) have been utilized to selectively concentrate and extract specific fractions of DOM for further characterization (Dittmar et al., 2008; Koprivnjak et al., 2009; Leenheer and Croué, 2003; Simjouw et al., 2005; Thurman and Malcolm, 1981). However, such approaches often led to inevitable modifications of the DOM composition (Coble, 2007; Repeta, 2015). Despite this, membrane separations (such as ultrafiltration, nanofiltration, reverse osmosis/electrodialysis) are considered less invasive than other methods (Thibault et al., 2019).

In aquatic systems, a portion of DOM is colored, which is known as Colored Dissolved Organic Matter (CDOM). CDOM has been studied since the early 1900s (Kalle, 1938) when it was referred to as "yellow substance" or "Gelbstoff" in German (Kalle, 1949). Thereafter, a subset of

CDOM was observed to emit fluorescence, which was termed fluorescent DOM (FDOM) (Kalle, 1966). The relationships between DOM, CDOM, and FDOM are shown in Figure 1-9. The specific chemical structures of compounds within CDOM and FDOM can encompass various molecules such as aromatic molecules, highly unsaturated compounds, and peptides (Stubbins et al., 2014), which can undergo complex interactions and transformations in natural aquatic environments, leading to the diverse and variable nature of CDOM and FDOM.

The optical properties of DOM, particularly absorption and fluorescence properties, could provide information on amount of material and the chemical characteristics of the bulk water samples, which were associated with physical, biological, and chemical processes (Coble, 2007). Optical spectroscopic methods that do not require pre-concentration processes have emerged as sensitive, inexpensive, fast, and non-destructive approaches, which could provide valuable insights into the dynamics of DOM at a much higher temporal and spatial resolution compared to other chemical approaches (Stedmon and Nelson, 2015). Given these advantages, such approaches have become increasingly popular in aquatic science in the last decades (Derrien et al., 2019; Nelson and Siegel, 2013).

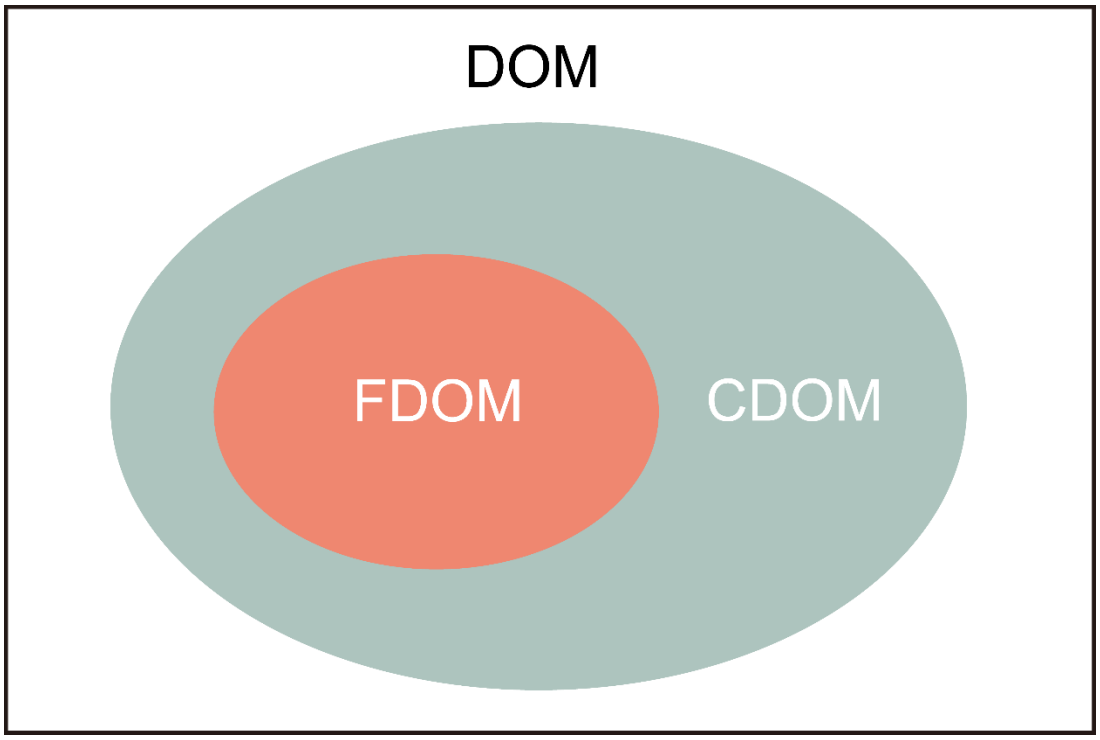



**Figure 1-9.** Schematic plot showing relationships between CDOM and FDOM adapted after Stubbins et al. (2014).

### 1.4.1 The Jablonski diagram

The Jablonski diagram is commonly used to show the processes that occur between light absorption and emission (Figure 1-10). When molecules absorb energy (light), they can pass from a lower energy state (singlet ground state, $S_0$) to a higher energy state (excited state, $S_1$ or $S_2$). Subsequently, the molecules return to the ground state ($S_0$) by emitting light of longer wavelength than the absorbed light. This emitted light (known as fluorescence), can be detected and measured to provide information about the molecules' optical properties. Furthermore, as the size of the aromatic compound increases, the energy difference between the ground and excited states decreases. Hence, the fluorescence signal of compounds with higher aromaticity typically undergoes a redshift (longer emission wavelength)(Stedmon and Nelson, 2015).

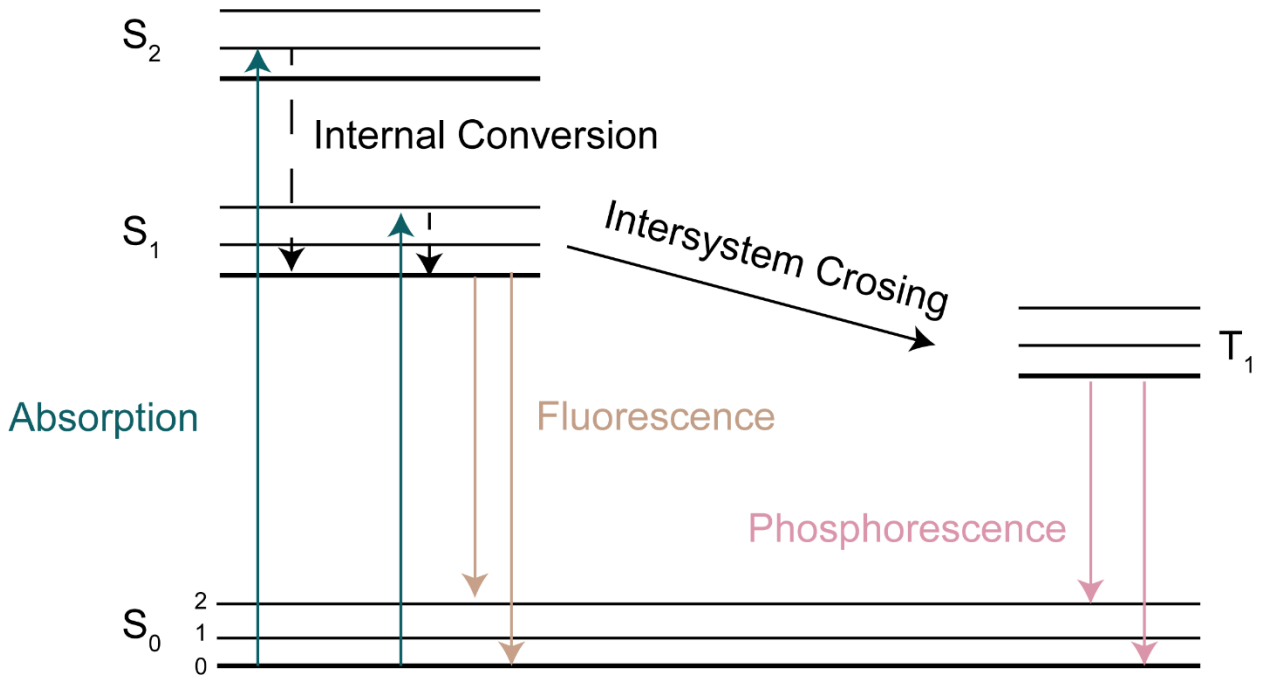



**Figure 1-10**. Jablonski diagram adapted after Lakowicz. (2006)

### 1.4.2 Absorption properties of DOM

The electromagnetic radiation is a form of energy, which can be distinguished by wavelength (Figure 1-11). The relationship between wavelength and energy is inverse. For example, as the wavelength of electromagnetic radiation increases, the energy of the radiation decreases.

Spectrometry is the study of the interaction of electromagnetic radiation with materials (e.g. OM). Spectrophotometry is a subset of spectrometry, which is the measurement of electromagnetic radiation (light) absorption as a function of wavelength. UV-Vis spectrophotometry refers to the measurement of absorption in the wavelength range between 200 nm and 800 nm, spanning ultraviolet radiation and visible parts of the spectrum (Figure 1-11). In these ranges, electromagnetic radiation-material interactions (absorption) are characterized by relatively high energy electron promotion from a lower ground energy state to an excited higher energy state

(Figure 1-10). The wavelength of ultraviolet or visible light absorbed is determined by the ease of electron promotion, which is determined by molecule structure and electron configuration.

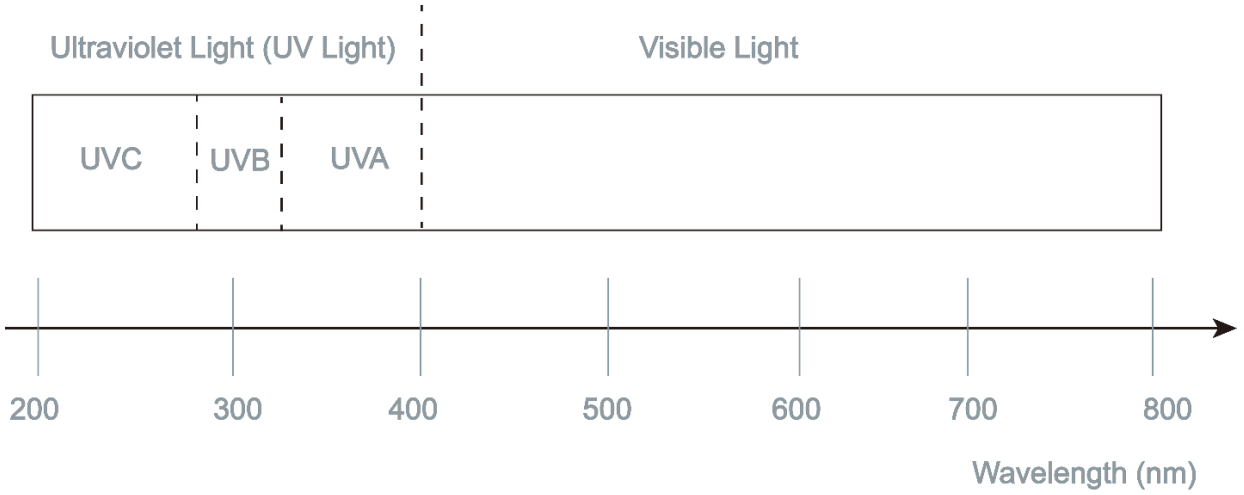



**Figure 1-11.** The spectrum of light.

In the ultraviolet (UV) region (200-400 nm) and the visible region (400-800 nm), photons have enough energy to promote bond electrons in molecules to higher energy levels, which can cause the organic molecules in CDOM to absorb light (Stedmon and Nelson, 2015). The UV-Vis spectrophotometer emits a broad spectrum of UV and visible light, which passes through the sample. As the light passes through the water sample, certain organic compounds present in the sample absorb specific wavelengths of light. The spectrophotometer measures the intensity of the transmitted light after it interacts with the sample. This is considered as a fast and nondestructive approach to analyze the CDOM properties.

Generally, the absorbance spectra of CDOM exponentially increases with shorter wavelengths (Figure 1-12). The UV-Vis spectra of CDOM are nearly featureless, with no single compound dominating (Leenheer and Croué, 2003). This featureless absorption spectrum might be due to the complex mixture of chromophores in CDOM, which overlap and interact with each other

to produce a broad absorption spectrum (Andrew et al., 2013; Boyle et al., 2009; Seritti et al., 1994; Stedmon and Nelson, 2015).

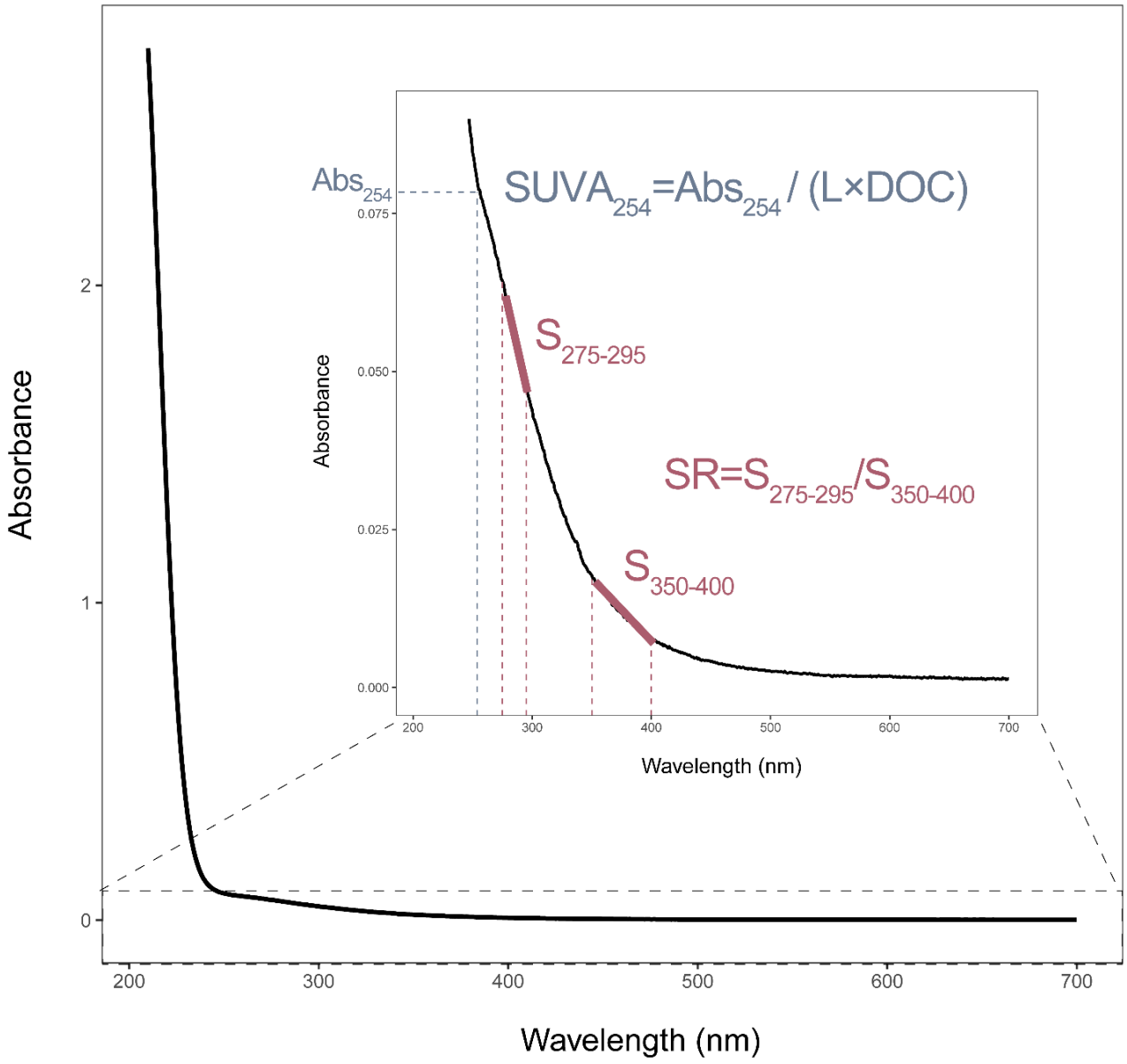



**Figure 1-12**. The UV-Vis absorbance spectra from a sub-surface water sample collected in July 2021 at Petit Couronne (Seine Estuary, France; Kilometric Point (KP) 251.3 - distance in kilometers from the city of Paris). $S_{275\text{-}295}$ represents slope for wavelengths in the 275–295 nm region, whereas $S_{350\text{-}400}$ represents slope for wavelengths in the 350–400 nm region, and SR is the ratio of these two slopes. The specific UV absorbance ($SUVA_{254}$) is measured at 254 nm.

To extract meaningful information from the UV-Vis absorbance spectra, several absorbance indices have been tested and widely applied in aquatic systems (Li and Hur, 2017). These indices are used for estimating quality and quantity of CDOM (Derrien et al., 2017). For example, the specific UV absorbance ($SUVA_{254}$) (Figure 1-12) is a parameter utilized to assess the quality of CDOM. It is calculated by dividing the absorbance at 254 nm ($Abs_{254}$) by the Dissolved Organic Carbon (DOC) concentration (mg/L), and has units of L mg-$C^{-1}$ $m^{-1}$ (Eq. 8). $SUVA_{254}$ provides information about the intensity of color of CDOM and is positively correlated with the molecular weight and aromaticity of the organic matter (Chin et al., 1994; Weishaar et al., 2003). It is a useful indicator of DOM quality in freshwater and industrial water treatment applications, and is also commonly used in estuaries (Asmala et al., 2013; Hounshell et al., 2022; Osburn et al., 2019; Zhuang et al., 2023). High $SUVA_{254}$ values (>4) indicate hydrophobic (aromatic) material. On the other hand, low $SUVA_{254}$ values (<3) indicate hydrophilic material (Edzwald and Tobiason, 1999; Matilainen et al., 2011).

$$SUVA_{254} = Abs_{254} / (L \times DOC) \qquad (8)$$

where Abs254 is the measured absorbance at 254 nm, L is the path length (m), and DOC is the dissolved organic carbon concentration (mg/L).

Moreover, the Slope Ratio (SR) is a parameter used to evaluate the variation in molecular weight of CDOM, as SR correlates negatively with molecular weight of DOM (Helms et al., 2008). It is calculated by dividing the spectral slope (S) obtained for a small UVB wavelength range (275-295 nm) by the S value obtained for a larger UVA wavelength range (350-400 nm) (Figure 1-12). Several studies have assessed the molecular weight of CDOM in estuaries using SR (Guo et al., 2014; Yang et al., 2013). The absorbance indices used in this thesis are summarized in Table 1-2.

These absorbance indices have been widely applied to evaluate aromaticity and molecular weight of estuarine DOM (Bergamaschi et al., 2012; Couturier et al., 2016; Dixon et al., 2014;

Osburn et al., 2019; Zhang et al., 2022; Zhuang et al., 2023). For example, in the Changjiang River Estuary, terrestrial DOM is less modified during transport downstream at high flows, as reflected by elevated $SUVA_{254}$ and lower SR (Zhang et al., 2022). In addition, aromatic content of DOM, reflected by $SUVA_{254}$, was highest in the river and decreased with salinity in the Newport River Estuary (Osburn et al., 2019).

**Table 1-2.** Absorbance indices used in this thesis

| Parameter | Calculation | Description | References |
|---|---|---|---|
| $SUVA_{254}$ | Abs254 / (L×DOC) | Proxy for aromaticity. High values (SUVA>4) indicate hydrophobic (aromatic) material. Low values (SUVA<3) indicate hydrophilic material | (Edzwald and Tobiason, 1999; Matilainen et al., 2011) |
| SR | $S_{275-295}/S_{350-400}$ | Proxy for average molecular weight (AMW). Higher values (1<SR<2) indicate lower AMW. Lower values (SR<1) indicate higher AMW | (Helms et al., 2008) |

### 1.4.3 Fluorescence properties of DOM

FDOM refers to a subset of CDOM that displays fluorescence when exposed to light (Figure 1-9). Since Kalle. (1949), different methods of fluorescence spectroscopy have been applied to characterize DOM in terrestrial and marine settings. These methods include (i) emission scans at specific excitation wavelengths (Miano et al., 1988), (ii) synchronous scans with a consistent wavelength offset between excitation and emission wavelengths (Lloyd, 1971), and (iii) excitation-emission matrix (EEM) that has become increasingly popular over the last decades (Coble et al., 1990; Derrien et al., 2019; Huguet et al., 2009; Kowalczuk et al., 2009, 2003; Parlanti et al., 2000).

Several fluorescence bands (α', α, β, γ) can be identified in the EEM spectra (Figure 1-13), which allow to indicate distinct sources and transformation processes of DOM (Parlanti et al.,

2000). The fluorescence bands, source information, and their Excitation/Emission wavelength at maximum fluorescence intensity are summarized in Table 1-3.

The α band (Excitation/Emission (Ex/Em) = 340-370/420-480nm) represents aromatic substances that can be derived from allochthonous or autochthonous sources (Coble, 2007; Haywood et al., 2018; Parlanti et al., 2000). Compared with the α band, the β band (Ex/Em = 310-320/360-410) is slightly blue shifted (shorter wavelength), which is associated with recently produced autochthonous material (Coble, 2007; Haywood et al., 2018; Parlanti et al., 2000). This band can thus indicate biological activity in aquatic systems (Parlanti et al., 2000). In addition, the γ band (Ex/Em = 270-280/300-350) is linked to protein-like compounds, which is an indicator of compounds from autochthonous and anthropogenic sources (*e.g.* wastewater treatment plant discharge) (Coble, 2007; Fellman et al., 2009, 2009; Haywood et al., 2018; Parlanti et al., 2000; Riopel et al., 2014).

**Table 1-3.** Spectral characteristics of the fluorescence bands

| Fluorophore | Wavelength range (nm) | Potential origin and characteristic | References |
|---|---|---|---|
| α | Ex 340-370 / Em 420-480 | Aromatic, mature, hydrophobic substances. Terrestrial or aquatic origin (difference in the position of the fluorescence emission maximum) | (Parlanti et al., 2000; Coble, 2007; Haywood et al., 2018) |
| α’ | Ex 230-260 / Em 380-500 | Mixture of aromatic, mature, Hydrophobic substances of terrestrial or aquatic origin and material of aquatic/biological origin freshly produced in the environment | (Parlanti et al., 2000; Coble, 2007; Haywood et al., 2018) |
| β | Ex 310-320 / Em 360-410 | Recent autochthonous production | (Parlanti et al., 2000; Coble, 2007; Haywood et al., 2018) |
| γ | Ex 270-280 / Em 300-350 | Protein-like compounds; biodegradable fraction; biological activity | (Parlanti et al., 2000; Coble, 2007; Fellman et al., 2009; Haywood et al., 2018) |

In addition to these fluorescence bands, several fluorescence indices (Table 1-4) have been proposed and applied for tracing sources of DOM (Derrien et al., 2017; Huguet et al., 2009; McKnight et al., 2001; Zsolnay et al., 1999). Specifically, γ/α represents the proportion of protein-like DOM compared to more aromatic and/or mature DOM, with high values indicating higher biological activity and higher biodegradability of DOM (Parlanti et al., 2000; Huguet et al., 2009). Humification Index (HIX) is a proxy for the degree of humification of DOM, which increases with humification transformation processes and aromaticity of organic material (Zsolnay et al., 1999). Biological Index (BIX) is a proxy for the degree of freshly produced DOM with microbial or biological origin (Huguet et al., 2009). Higher BIX values indicate higher contribution of fluorophore β, which is related to recent production of autochthonous DOM in aquatic systems (Huguet et al., 2009; Parlanti et al., 2000). Biological/aquatic bacterial DOM has a low HIX and a high BIX while allochthonous DOM has a high HIX and a low BIX (Huguet et al., 2009). Fluorescence Index (FI) is a proxy for distinguishing between terrigenous and microbial DOM. High FI (FI=1.9) indicates a microbial origin while low FI values (FI=1.3) reflect DOM derived from higher plants (McKnight et al., 2001).

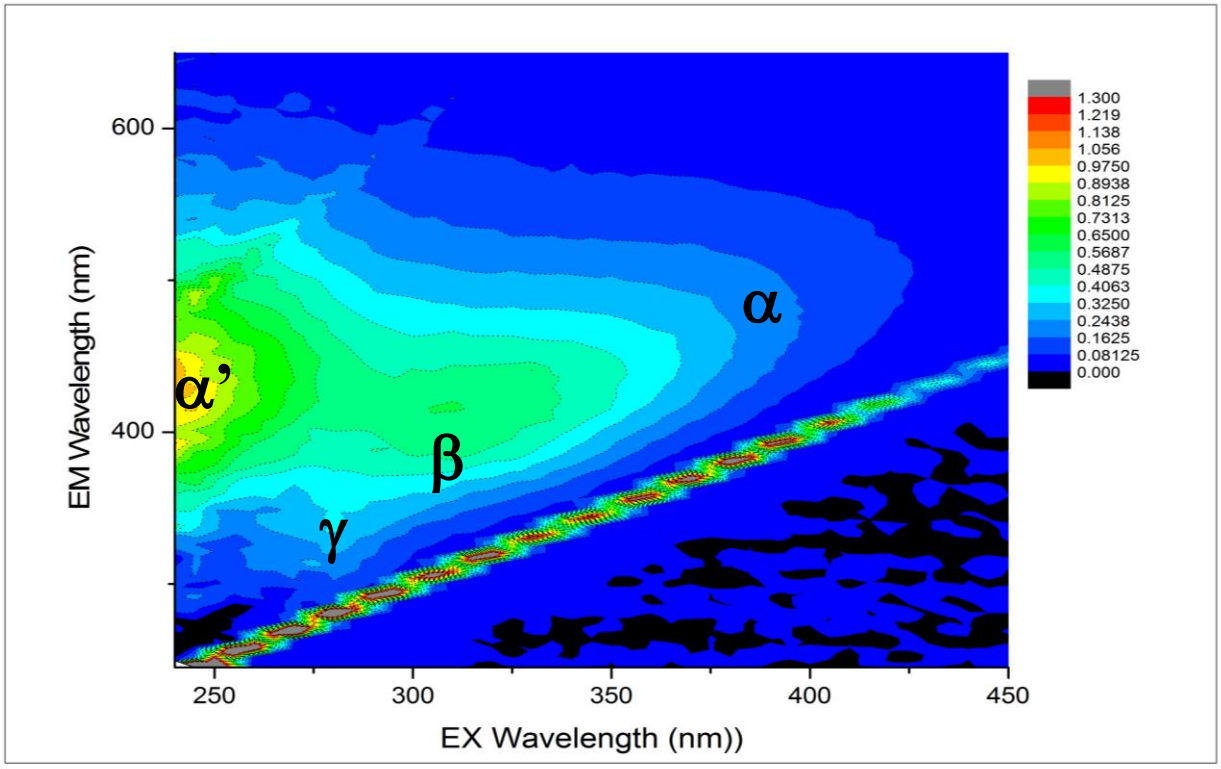



**Figure 1-13.** EEM spectrum of a subsurface water sample collected in July 2021 at Petit Couronne (Seine Estuary, France). Position of the main fluorescence bands α', α, β and γ observed by the peak picking technique.

These fluorescence indices have been used to trace FDOM sources and transformations in estuaries worldwide, such as the Gironde estuary (Huguet et al., 2009), Seine Estuary (Huguet et al., 2010), Pearl River Estuary (Liu et al., 2020), Minjiang estuary (Xie et al., 2023), and Changjiang River Estuary (Zhang et al., 2022).

For example, seasonal variability of fluorescence proxies are observed in the Pearl River Estuary, with higher HIX in the wet season (Liu et al., 2020). DOM chemistry in this estuary was thus considered to be significantly influenced by river discharge. In the Pearl River Estuary, DOM was observed to be mainly from autochthonous origin as indicated by high values of BIX (>0.8) (Xie et al., 2018).

**Table 1-4**. Fluorescence spectroscopic indices used in this thesis

| Parameter | Calculation | Description | References |
|---|---|---|---|
| γ/α | The ratio of the intensities of the bands γ ($Ex_{280}/Em_{330}$) and α ($Ex_{370}$/maximum emission) | Indicator of the proportion of protein-like DOM compared to more aromatic and/or mature DOM (terrestrial or aquatic). Higher values indicate higher biological activity and higher biodegradability of DOM. | (Parlanti et al., 2000; Huguet et al., 2009) |
| HIX | The ratio of emission spectrum areas between 435-480nm and 300-345nm, with excitation 255nm | Estimation of the degree of aromaticity of DOM. High values indicate the presence of processed DOM (polymerization, polycondensation) while low values indicate more recent and less aromatic DOM. HIX>16: more aromatic material (significant terrigenous contribution); HIX<4: aquatic biological or bacterial origin | (Huguet et al., 2009) |
| BIX | The ratio of fluorescence intensities at emission 380nm to that at 430nm, with excitation 310nm | Estimation of autochthonous DOM production and the presence of freshly produced DOM. Increases with biological activity. BIX>1: aquatic biological or bacterial origin; BIX<0.6: low biological activity. Linked to the biodegradability of DOM. | (Huguet et al., 2009) |
| FI | The ratio of fluorescence intensities at emission 450nm to | Estimate of DOM sources, high values (>1.9) indicate microbial DOM, while low | (McKnight et al., 2001) |

| | | | |
|---|---|---|---|
| | that at 500nm, with excitation 370nm | values (<1.3) indicate DOM from higher plants | |

The EEM spectra could be further decomposed by parallel factor analysis (PARAFAC) into distinct fluorescent components that represent groups of similar fluorophores (Stedmon et al., 2003a). This statistical technique has the capability to overcome the limitations of the conventional peak picking technique (Cory et al., 2011; Yamashita et al., 2008), as DOM in the natural environment is composed of different types of overlapping fluorophores. The decomposed fluorescent groups can provide valuable insights into distinct DOM sources (i.e. allochthonous or autochthonous) and transformation processes (i.e. photodegradation) (Ishii and Boyer, 2012; Jaffé et al., 2014; Kowalczuk et al., 2013). Furthermore, PARAFAC components were associated with molecular families determined by ultrahigh resolution mass spectrometry (FTICR-MS), which indicates that fluorescence measurements can provide insight into the biogeochemical cycling of a large proportion of the DOM pool including non-fluorescent molecules (Stubbins et al., 2014).

Over the last years, spectroscopic techniques coupled with PARAFAC have been widely used for characterizing DOM in a variety of natural environments (Chai et al., 2019; Hu et al., 2021; Kowalczuk et al., 2009; Y. Liu et al., 2021; Luo et al., 2021; Mielnik and Kowalczuk, 2018; Qin et al., 2020) as well as in engineered systems (Sanchez et al., 2014; Sciscenko et al., 2022a; Yang et al., 2015). For example, EEM-PARAFAC was shown to have important implications to study photodegradation (Mangalgiri et al., 2017; Murphy et al., 2018), microbial processes (Parr et al., 2015), priming effects (Zhuang et al., 2021), pollution sources (Wang et al., 2022), binding properties of Cu (II) (Liu et al., 2022), fluoroquinolones oxidative transformation processes (Sciscenko et al., 2022b, 2021), potable water reuse monitoring (Wells et al., 2022), as well as understanding aging mechanism of microplastics (Priyanka and Saravanakumar, 2022).

EEM-PARAFAC has notably been used to identify distinct sources and transformation of DOM in estuaries (Hounshell et al., 2017; Xie et al., 2018; Zhang et al., 2022; Zhu et al., 2017; Zhuang et al., 2023). For example, seaward decrease of the terrestrial and protein-like PARAFAC components was observed in the Pearl River Estuary, which was attributed to estuarine mixing processes (Xie et al., 2018). In the Qiantang Estuary, terrestrial and protein-like PARAFAC components decreased with increasing salinity, which was explained by physical mixing of freshwater and saltwater (Zhou et al., 2019).

### 1.4.4 DOM dynamics in estuaries and challenges

Estuaries represent highly dynamic zones where complex chemical, physical, and biological processes interact (Bianchi, 2007; Fairbridge, 1980). The DOM composition in such systems varies spatially and temporally due to mixing of water masses (Osburn et al., 2015; Santos et al., 2014; Xie et al., 2018), seasonality (Vidal et al., 2023) and different (biotic and abiotic) transformation processes, including microbial degradation (Asmala et al., 2013; Q. Chen et al., 2021), photochemical degradation (Santos et al., 2014), aggregation (Søndergaard et al., 2003), and adsorption/desorption from suspended particles (Wang et al., 2016).

Enhanced anthropogenic activities such as urbanization and wastewater runoff also contribute to anthropogenic DOM into estuaries (García-Martín et al., 2021; Hounshell et al., 2017). Such human activities can also lead to high concentrations of inorganic nutrients in estuaries, triggering autochthonous production of DOM (Wilson and Xenopoulos, 2009). Autochthonous DOM can be, at least in part, quickly degraded by microbial activity, leading to non-conservative DOM behavior and contributing to the release of $CO_2$, $CH_4$, and $N_2O$ to the atmosphere (Amaral et al., 2021; García-Martín et al., 2021).

Different land uses types, such as urban, agricultural, and forested areas have distinct influence on the DOM characteristics in nearby water bodies, by erosion, oxidation and shallow flow path (Asmala et al., 2013; Bhattacharya and Osburn, 2020; Boukra et al., 2023; García-Martín et al., 2021; Shang et al., 2018; Williams et al., 2010; Zhang et al., 2021). For example, wetland and agricultural coastal streams have high levels of aromatic and complex DOM, whereas forested and urban streams have low levels of aromatic DOM (Bhattacharya and Osburn, 2020).

In addition, many studies have shown that water discharge plays a significant role in shaping estuarine DOM compositions, with higher discharge leading to increased terrestrial DOM inputs and low-water periods allowing for intensive in-estuary DOM processing (Bittar et al., 2016; Peer et al., 2022; Regier and Jaffé, 2016; Singh et al., 2019; Xie et al., 2018).

Considering these factors, it is challenging to identify an estuarine zonation in terms of generalized DOM characteristics, despite its importance for sustainable estuarine management. A comprehensive dataset across a wide range of land use types over distinct hydroclimate conditions is thus necessary to investigate the variabilities of estuarine DOM and their controlling factors. However, to date, such estuarine dataset is still relatively limited, as it requires extensive monitoring sampling. More importantly, the analysis of such a complex dataset requires the use of advanced statistical techniques to extract meaningful insights and to identify hidden patterns (Bieroza et al., 2012; Cuss and Guéguen, 2016; He and Fan, 2016; Wheeler et al., 2017; Zhang et al., 2015).

## 1.5 Machine learning in environmental science

Machine Learning (ML) is a powerful technique, which is defined as follows: “ML algorithms build a model based on sample data, known as ‘training data’, to make predictions or decisions without being explicitly programmed to do so” (Koza et al., 1996). ML algorithms extract

patterns and relationships from the (provided) training data to build a model that can make predictions or generate clusters. In the past decade, ML has been utilized in a variety of scientific fields, including medicine (Swanson et al., 2023), material science (Hart et al., 2021; H. Tao et al., 2021), and chemistry (Fedik et al., 2022; Jorner et al., 2021).

In environmental sciences, the amount and complexity of data has significantly expanded (Fleming et al., 2021; Zhong et al., 2021a). Traditional statistical tools (e.g. Analysis of Variance and linear regression) may not be sufficient to handle such large and complex datasets. Instead, machine learning can efficiently handle and evaluate these complex datasets, uncovering hidden patterns and relationships (Peters et al., 2014; Tahmasebi et al., 2020). For example, ML has been used to predict reactivity of contaminants (Zhong et al., 2021b), arsenic concentrations in groundwater (Podgorski and Berg, 2020), and plant uptake efficiency (Bagheri et al., 2020). In addition, ML can identify the important factors that contribute to a particular environmental outcome, including pollutant concentrations (Hu et al., 2017), and uptake of contaminants (Bagheri et al., 2019). ML can also detect anomalies or abnormal patterns in environmental dataset, such as the assessment of contamination (Housh and Ostfeld, 2015).

While ML has recently gained significant popularity and success in environmental sciences, its applicability in estuarine OM research still needs to be explored. This might be due to the challenges and complexities associated with studying estuarine OM dynamics, which are influenced by both natural processes (Li et al., 2023) and anthropogenic activities (Bhattacharya and Osburn, 2020). Additionally, obtaining high-resolution and high-quality data on estuarine OM could be time-consuming and challenging. ML algorithms often require substantial amounts of training data to build robust models, and the scarcity of data can hinder the application of ML approaches in DOM research. Despite these challenges, there is growing interest in exploring the potential of ML in studying DOM dynamics. For example, a recent study has applied ML and

advanced mass spectrometry techniques to observe the complex relationships between the molecular features and $\delta^{13}C$ values in water samples collected in the China Coastal Environments (Yi et al., 2023). In addition, Ju et al. (2023) applied ML for prediction of ultraviolet absorption spectra of CDOM. Photochemical properties of DOM can also be predicted by ML models (Liao et al., 2023).

### 1.6. Research gaps

Tracing the input of terrestrial organic matter to marine environments is crucial for understanding global carbon cycling (Dai et al., 2022). Over the last years, a number of terrestrial proxies have been proposed and widely used to trace POM dynamics in estuaries (Bianchi, 2007; Bianchi and Canuel, 2011; Canuel and Hardison, 2016; Savoye et al., 2012). However, a number of research gaps remain to be filled. For example, different proxies have their own limitations. Specifically, the bulk geochemical proxies are influenced by diagenetic processes (Lamb et al., 2006), whereas the widely used molecular proxy (BIT) is controlled by various factors, including the selective degradation of branched vs. isoprenoid GDGTs (Smith et al., 2012). The development of additional proxies to trace riverine runoff processes is thus needed.

Furthermore, DOM sources, compositions, transformation processes have been extensively investigated in estuaries (Jaffé et al., 2014). Considering its multiple sources, various controlling factors, and complex transformation processes in estuaries, it remains challenging to properly assess the main DOM characteristics within specific estuarine zone. As machine learning techniques can effectively handle complex dataset and capture hidden data patterns, its applicability in disentangling DOM heterogeneity should be further explored.

Last, most of the previous studies investigated DOM and POM separately (Derrien et al., 2019). As each OM pool has its own properties and dynamics (Thibault et al., 2019), simultaneous

investigation of DOM and POM should be prioritized. In addition, the relationship between land use characteristics, water discharge, and distinct types of POM and DOM is currently not clear.

### 1.7. Research objective and questions

Within the framework of the SARTRE (GIP Seine-Aval) and RUNTIME (EC2CO CNRS/INSU/OFB) projects, the objective of this thesis is to evaluate the ecological functioning of an urbanized estuary in France (Seine Estuary) by simultaneously investigating the DOM and POM dynamics. To fill the mentioned research gaps, this thesis aims to answer the following scientific questions:

- How to trace riverine runoff processes in the Seine Estuary? What are sources of DOM and POM in the Seine Estuary? What are spatiotemporal variations of DOM and POM in the Seine Estuary?
- What are relationships between natural processes (i.e. water discharge), anthropogenic processes (i.e. land use changes), and distinct types of DOM and POM?
- What is the ecological functioning of the Seine Estuary in terms of POM and DOM cycling? Does the application of the machine learning and explainable artificial intelligence make it possible to identify estuarine zonation? Is there a generalized estuarine zonation for OM cycling in the Seine Estuary?

These scientific questions are achieved in the following chapters (Chapter 3 to 5), which represent 3 manuscripts for peer–reviewed publication. **Chapter 2** presents the detailed material and methods in this thesis. **Chapter 3** investigates the POM dynamics in the Seine River basin and explores novel indicators for riverine runoff processes. **Chapter 4** assesses the relationships between water discharge, land use characteristics, and different types of POM in the Seine Estuary.

**Chapter 5** attempts to study DOM dynamics in the Seine Estuary and tests the potential of using machine learning and explainable artificial intelligence for disentangling DOM composition and identifying estuarine zonation. **Chapter 6** synthesizes the results and provides future perspectives.

# Chapter 2 :

## Material and methods

## 2.1. Study area

The Seine River basin (Seine River and its estuary; Figure 2-1) has a surface area over 76000 km$^2$ and is characterized by high population density, draining through the greater Paris region (over 17 million inhabitants) to the English Channel (Flipo et al., 2021). The Seine estuary, approximately 160 km in length, extends from Poses (the upper limit of tidal influence; Figure 2-1) to the English Channel and is characterized as a macrotidal estuary based on its small depth, high tidal range, and morphology (Avoine et al., 1981; Guézennec et al., 1999).

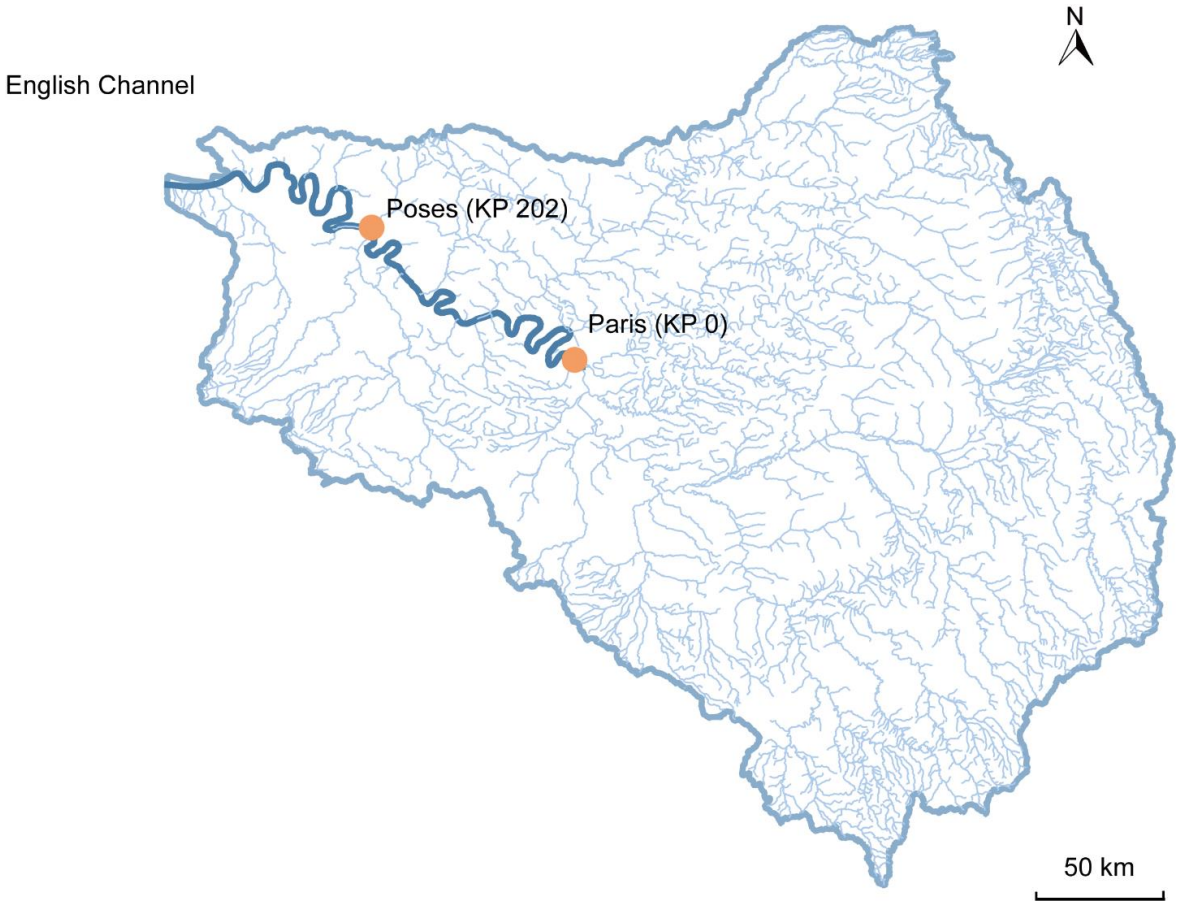



**Figure 2-1**. Hydrological network of the Seine River basin. Kilometric Point (KP) represents the distance in kilometers from the city of Paris (KP 0). A dam at Poses (KP 202) constitutes the boundary between the Seine River and the Seine Estuary.

The river basin is 97% contained inside the sedimentary Paris basin, Europe's largest groundwater reservoir (Triassic to Tertiary, Figure 2-2). The basin lithology includes carbonates (69.6%) and sandy formations (13.6%), which are interbedded with poorly permeable clayey and marl units (9.1%), and are covered by alluvial deposits (5.4%) (Guillocheau et al., 2000).

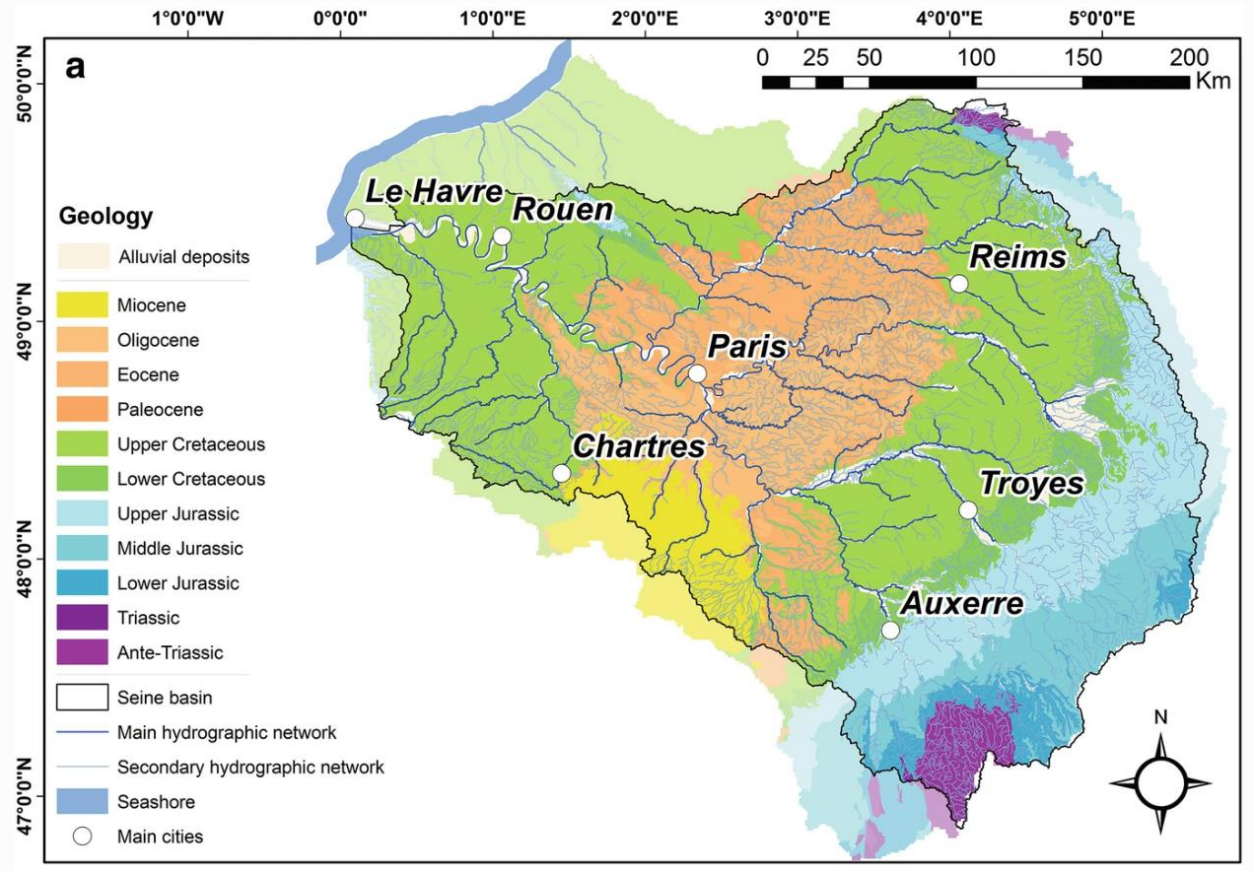



**Figure 2-2**. Geological structure of the Seine basin (Flipo et al., 2021).

The Seine River basin has a pluvial/oceanic hydrological regime (Flipo et al., 2021). The average annual rainfall over the basin is 800 mm, and it varies spatially. Near the coast and in the Morvan mountain range, the maximum rainfall reaches approximately 1200 mm per year, while in the center of the basin, it is only 650 mm per year (Flipo et al., 2020; Quintana-Segui et al., 2008; Vidal et al., 2010). The river flow regime is influenced by seasonal variations in real

evapotranspiration, resulting in high flows during winter and low flows during summer (Flipo et al., 2021).

Over the past two centuries, the urbanized area of the basin has expanded significantly. Population density near urban tributaries in the Paris area ranges from 1000 to 5000 inhabitants per square kilometer, while it is considerably lower, averaging less than 20 inhabitants per square kilometer, in the upstream regions of the basin (Flipo et al., 2021). For over a century, the city of Paris, and later the expansion of the Parisian conurbation, have caused urban pollution, resulting in a significant impact on the water quality of the lower Seine River and its estuary, with depleted oxygen levels, elevated concentrations of ammonia and nitrite, and the presence of fecal bacteria (Flipo et al., 2021; Mouchel et al., 2021; Servais et al., 2007). Due to modernization of wastewater treatment plants (WWTPs), the water quality of the Seine River has shown significant improvement over the past two decades (Romero et al., 2016). However, there are still occasional crises, such as summer low-flow conditions, that pose a threat to maintaining the good ecological status of riverine and estuarine waters (Garnier et al., 2021).

## 2.2 Sampling

During 5 low-frequency campaigns from June 2019 to June 2021, 130 samples were collected at Les Andelys (KP 175; KP represents kilometric point and is defined as the distance in kilometers from the city of Paris), Oissel (KP 229.4), Val-des-Leux (KP 265.55), Caudebec (KP 310.5), and Tancarville (KP 337) (Figure 2-3 and Table 2-1). Pictures taken during these sampling campaigns are shown in Figure 2-4. Both sub-surface and bottom water (2.2-16 m depth) samples were retrieved using a pump into precleaned 20L FLPE Nalgene carboys. Estuarine water samples (Oissel, Val-des-Leux, Caudebec, and Tancarville) were collected at three tide periods (high tide, low tide and mid tide). In November 2020, 4 water samples were collected at Poses (KP 202),

Triel-sur-Seine (KP 80), Bougival (KP 40), and Marnay-sur-Seine (KP-200) (Figure 2-3, 2-5 and Table 2-1). For these sites, 0.25-43L of water were immediately filtered using pre-combusted Whatman GF/F 0.7 µm glass fiber filters. After filtration, filters were freeze-dried, scratched and stored frozen at -20°C prior to analysis.

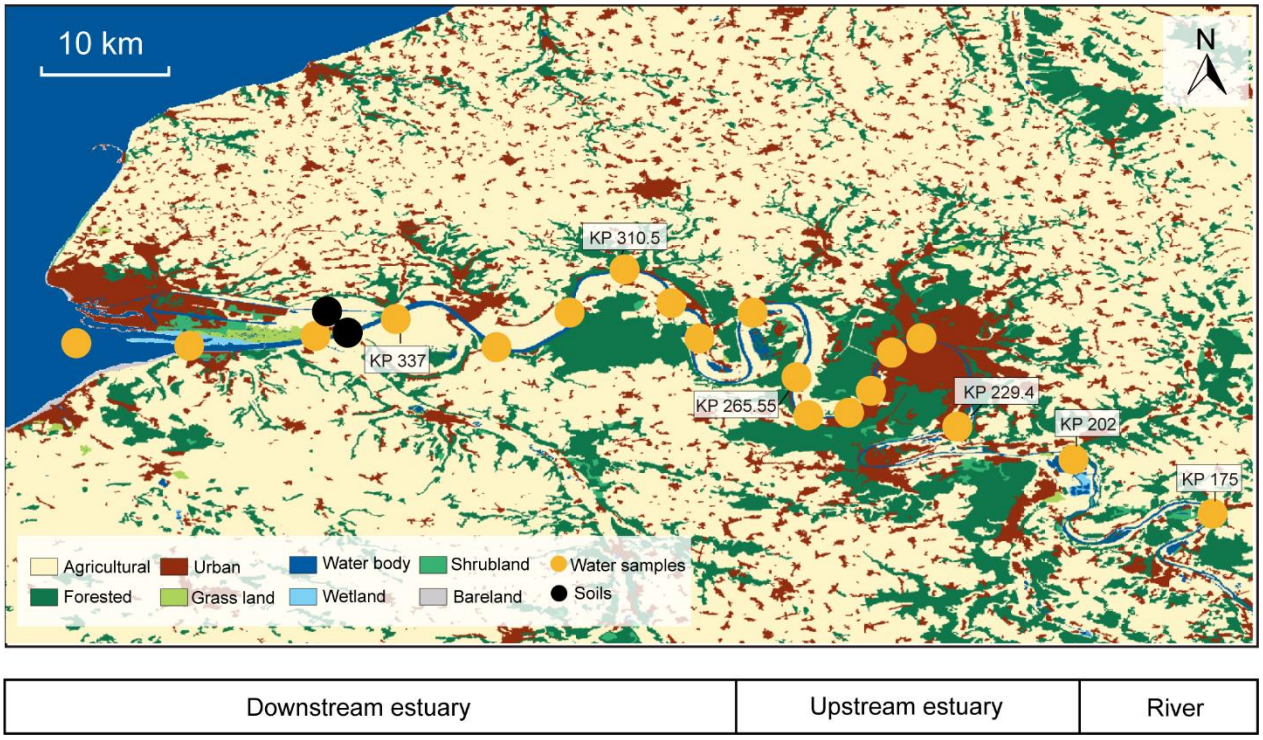



**Figure 2-3.** Map showing the sample locations and land use characteristics along the Seine Estuary. The land use data was retrieved from GLOBELAND30 (http://www.globallandcover.com/). Inland water body and seawater are combined into a single category as water body. KP (kilometric point) is defined as the distance in kilometers from the city of Paris

During 19 monitoring campaigns from June 2019 to November 2022, surface water (ca. 1m depth) samples ($n$=249) were collected at 15 locations in high-flow (over 250 $m^3/s$) and low-flow (below 250 $m^3/s$) conditions across the Seine Estuary with distinct land cover regimes (Figure 2-3 and Table 2-1). For these sites, water samples were immediately filtered on board through pre-combusted (450 °C) 0.7 µm glass fiber filters (GF/F Whatman) and stored in darkness at 4 °C until analysis. Filters were freeze-dried, scratched and stored frozen at -20°C prior to analysis.


**Figure 2-4.** Pictures taken during sampling campaigns showing some of the sampling sites at the (a) Les Andelys, (b-c) Val-des-Leux, and (d) Tancarville. Photo by Zhe-Xuan Zhang.

In 2021, surficial soils (*n*=9) were collected in the lateral area of the upstream part of the Seine River (Figure 2-5). In 2018, 2020, and 2021, additional wetland soils and mudflat sediments (*n*=42) were collected in the downstream estuary (Figure 2-3), representing allochthonous material that can be flushed into the estuary by tidal effects. These samples were collected at low tide with a plexiglass® core (4.5 cm depth), then homogenized, freeze-dried, and ground in a ball mill (model MM400, Retsch®).

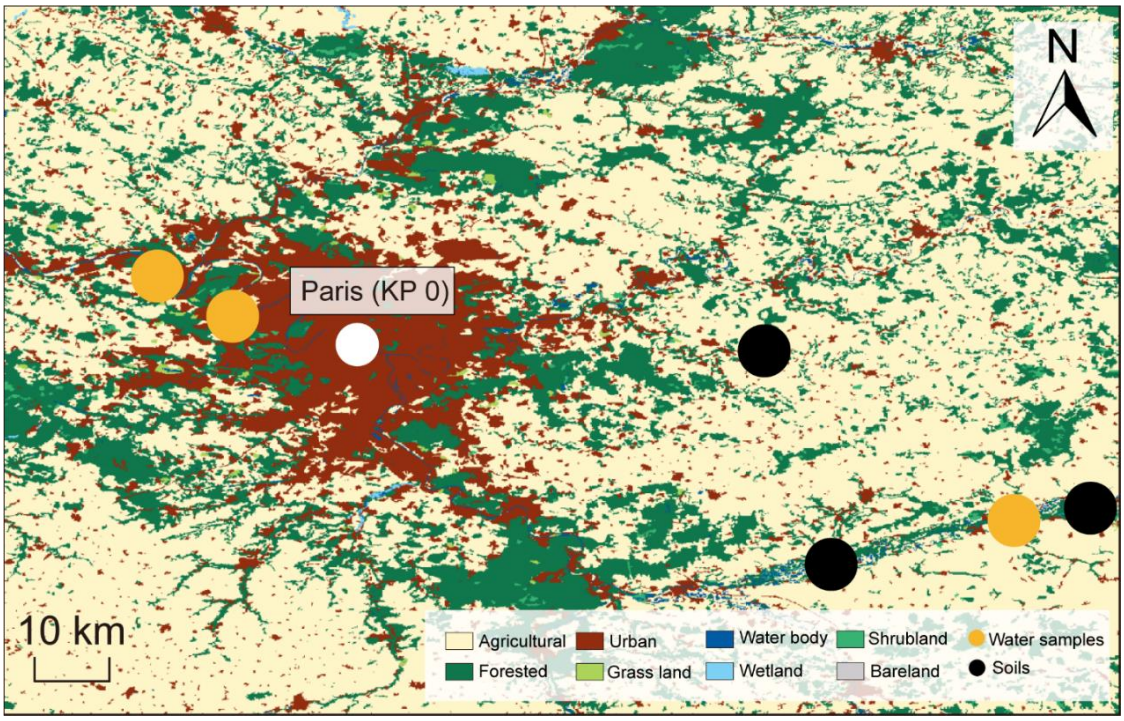



**Figure 2-5.** Map showing sample locations around the Seine River and land use characteristics. The land use data was retrieved from GLOBELAND30 (http://www.globallandcover.com/). KP (kilometric point) is defined as the distance in kilometers from the city of Paris (KP 0).

The land use data across the Seine River basin was retrieved from the worldwide surface coverage product GLOBELAND30 (http://www.globallandcover.com/) with a resolution of 30 meters in 2020. Seawater and inland water body are combined into a single category as water body. Eight land use types, including urban (industrial land use included), agricultural, forested, water body, shrubland, bareland, grassland, and wetland, can be identified across the Seine River basin (Figure 2-3 and 2-5). To calculate the land use type proportion for sampling sites, a 1 km (radius) buffer zone around each site was created using ArcGIS (10.7) software. A 1 km buffer is chosen as it can capture the effects of land use patterns in the nearby environment on the target variables (i.e. organic matter characteristics) within the water column (Hu et al., 2016; Zhang et al., 2023).

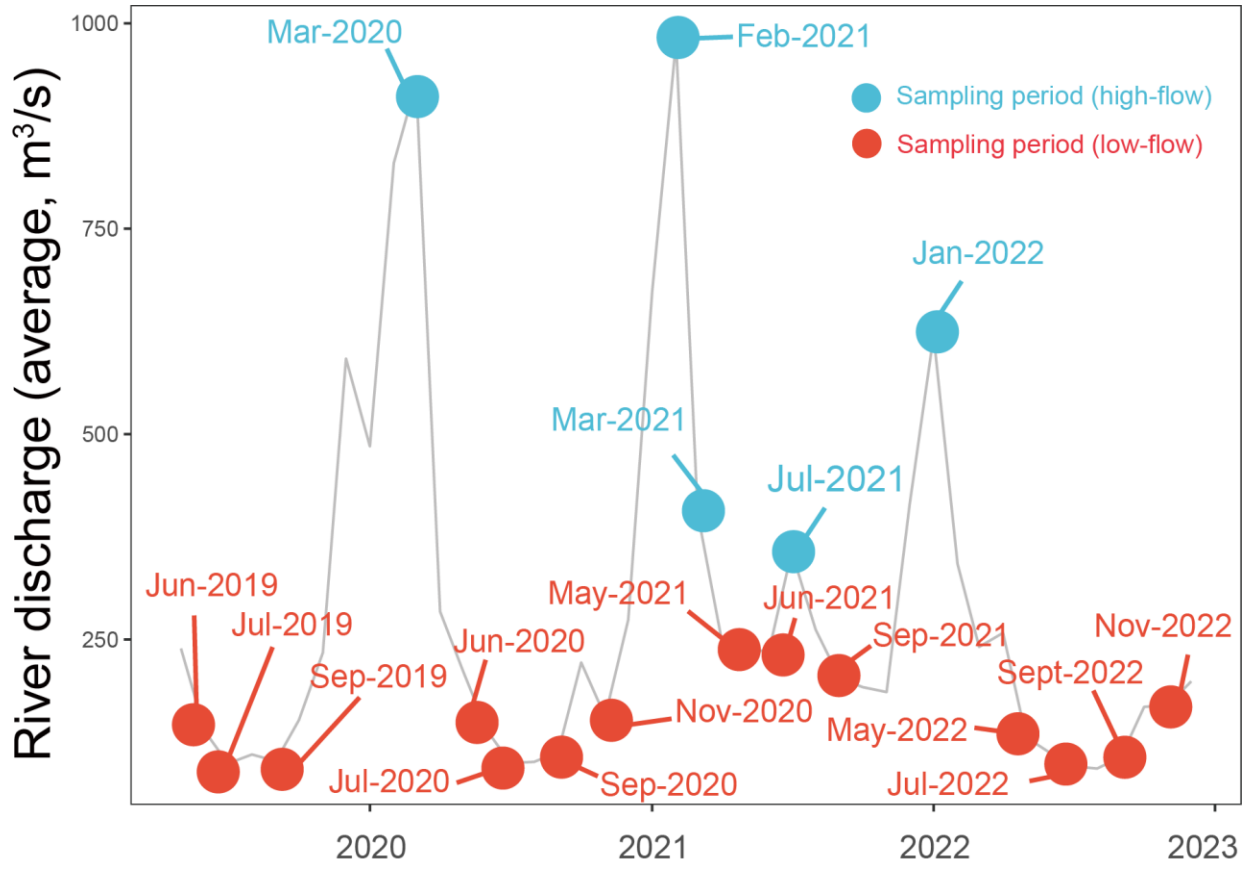



**Figure 2-6**. Mean monthly water discharge for the Seine River measured at the Paris Austerlitz station from 2019 to 2022 (data from https://www.hydro.eaufrance.fr/). Bullets represent the sampling period in high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) periods.

Generally, maximum flows occur during winter (>250 $m^3/s$), while minimum flows are observed in summer (<250 $m^3/s$) (Figure 2-6). Given the seasonally variable water discharge (Figure 2-6) and diverse land use characteristics across the estuary (Figure 2-3), this estuary is an ideal region for studying DOM and POM cycling and their relationships between human activities (i.e., land use characteristics) and natural processes (i.e., water discharge). Additionally, studying the biogeochemical functioning of the Seine Estuary in terms of DOM and POM dynamics can provide valuable insights that may also contribute to a better understanding of other urbanized estuaries.

**Table 2-1.** DOM and POM sample information

| Location | KP (km) | Longitude (°) | Latitude (°) | Zone | Type |
|---|---|---|---|---|---|
| Balise A | 360.8 | 0.110671 | 49.431828 | Estuary | DOM |
| Honfleur | 355.8 | 0.232682 | 49.432638 | Estuary | DOM and POM |
| Berville-Sur-Mer | 346 | 0.3682 | 49.441587 | Estuary | DOM and POM |
| Tancarville | 337 | 0.463442 | 49.472351 | Estuary | DOM and POM |
| Petitville | 326.6 | 0.577669 | 49.435988 | Estuary | DOM |
| Vatteville-La-Rue | 318 | 0.66614 | 49.472695 | Estuary | DOM and POM |
| Caudebec | 310.5 | 0.72753 | 49.522585 | Estuary | DOM and POM |
| Le Trait | 303 | 0.776177 | 49.483864 | Estuary | DOM |
| Heurtauville | 297.65 | 0.816867 | 49.447614 | Estuary | DOM and POM |
| Duclair | 278 | 0.873297 | 49.478666 | Estuary | DOM and POM |
| Val-des-Leux | 265.55 | 0.92 | 49.4 | Estuary | DOM and POM |
| La Bouille | 259.7 | 0.934366 | 49.35228 | Estuary | DOM |
| Haulot Sur Seine | 255.6 | 0.98475 | 49.356683 | Estuary | DOM and POM |
| Petit Couronne | 251.3 | 1.008118 | 49.379279 | Estuary | DOM and POM |
| Le Grand Quevilly | 246.6 | 1.030269 | 49.432815 | Estuary | DOM |
| Rouen | 243 | 1.06979 | 49.4428698 | Estuary | DOM |
| Oissel | 229.4 | 1.1 | 49.34 | Estuary | POM |
| Poses | 202 | 1.24 | 49.31 | Estuary | POM |
| Les Andelys | 175 | 1.4 | 49.24 | River | POM |
| Triel-sur-Seine | 80 | 2 | 48.98 | River | POM |
| Bougival | 40 | 2.13 | 48.87 | River | POM |
| Marnay-sur-Seine | -200 | 3.56 | 48.51 | River | POM |

## 2.3. Elemental and isotopic analyses

Elemental and isotopic analyses were performed on soils (surficial soils and mudflat sediments) and SPM using the approach described by Thibault et al. (2019). In brief, 1 g of soils/sediments and 40 mg of SPM were decarbonated for 2 hours with magnetic stirring at room temperature after adding 10 mL of 3 M HCl. Following that, samples were rinsed with ultrapure water and centrifuged until they reached neutral pH. The decarbonated samples were kept at −20

°C and freeze dried. Both decarbonated and non-decarbonated samples (20 mg for soils and 6 mg for SPM) were enclosed in a tin capsule. Total Organic Carbon content (TOC) and stable carbon isotopic composition ($\delta^{13}C$) were measured in decarbonated samples at the ALYSES platform (Sorbonne University / IRD, Bondy, France) using an elemental analyzer coupled with an isotope ratio mass spectrometer (Thermo Fisher Scientific Delta V Advantage). Total nitrogen (TN) and nitrogen isotope ($\delta^{15}N$) were analyzed using non-decarbonated samples, because acidification could affect N contents (Ryba and Burgess, 2002). The isotopic composition ($\delta^{15}N$ or $\delta^{13}C$) was expressed as the ratio of isotope ratios in samples and standards (atmospheric $N_2$ for nitrogen or Vienna Pee Dee Belemnite for carbon).

## 2.4. Lipid extraction and analyses

The lipids were extracted ultrasonically (3×), using 20 to 40 mL of dichloromethane (DCM): methanol (MeOH) (5/1, v/v) per extraction, from surficial soils and mudflat sediments (4-20 g) as well as SPM samples (150 mg). The total lipid extracts were separated on an activated silica gel column into fractions of increasing polarity using (i) 30 mL of heptane, (ii) 30 mL of heptane:DCM (1/4, v/v), and (iii) 30 mL of DCM/MeOH (1/1, v/v) as eluents.

An aliquot (30%) of the polar fraction containing GDGTs and GMGTs was dried, re-dissolved in heptane, and then passed through a 0.2µm polytetrafluoroethylene (PTFE) filter (Ultrafree-MC; Merck). 5 µl of the internal standard ($C_{46}$ Glycerol Trialkyl Glycerol Tetraether; 0.01025 mg/mL) was typically added to 45 µl of sample. GDGTs and GMGTs were analyzed using a Shimadzu LCMS 2020 high pressure liquid chromatography coupled with mass spectrometry with an atmospheric pressure chemical ionization source (HPLC-APCI-MS) using Selected Ion Monitoring (SIM) mode, modified from Hopmans et al. (2016) and Huguet et al. (2019). Tetraether

lipids were separated with two silica columns in tandem (BEH HILIC columns, 2.1 × 150 mm, 1.7 µm; Waters) thermostated at 30°C. Injection volume was 30 µL, with the flow rate set at 0.2 mL/min. GDGTs and GMGTs were eluted isocratically for 25 min with 82% A/18% B (A= hexane, B=hexane/isopropanol 9/1, v/v), followed by a linear gradient to 65% A/35% B in 25 min, then a linear gradient to 100% B in 30 min, and back to 82% A/18% B in 4 min, maintained for 50 min. Semi-quantification of tetraether lipids was performed by comparing the integrated signal of the respective compound with the signal of a $C_{46}$ synthesized internal standard, assuming their response factors to be identical (Huguet et al., 2006). LabSolutions software (Shimadzu) was used to process the data.

Another aliquot (6%) of the polar fraction containing sterols, stanols, and fatty acids was dried, re-dissolved in DCM, and derivatized with a mixture of N,O-bis-(trimethylsilyl) trifluoroacetamide and trimethylchlorosilane (BSTFA + TMCS, 99/1, v/v) at 70 °C for 1 hour with 0.7995 or 1.5 µg 5α-cholestane added as the internal standard. These compounds were analyzed by GC-MS using a Thermo Scientific Trace 1310 gas chromatograph fitted with a Rxi® -5Sil MS column (60 m × 250 µm × 0.25 µm; RESTEK) interfaced to a ISQ 7000 single quadrupole mass spectrometer. 1 µL of the derivatized polar fraction was injected at 2 mL/min in split mode (10:1) using He as the carrier gas. The oven temperature started at 70 °C (held 1 minute), increased to 130 °C at 20 °C/min, then increased to 320 °C (held 25 minutes) at 4 °C/min. The mass spectrometer was simultaneously operated in full scan mode (m/z 35-700) and SIM mode (m/z 75 for fatty acids, 129 for sterols, 215 for stanols, and 217 for the internal standard). The transfer line temperature was set at 320 °C and the EI voltage was set at 45 eV. Chromeleon software was used to process the data. Identification of sterols, stanols, and fatty acids was based on their retention time and mass spectra.

An aliquot (40%) of the apolar fraction containing *n*-alkanes was dried and re-dissolved in heptane with 0.03 or 0.05 µg *n*-tetracosane-d50 added as an internal standard. *n*-alkanes were analyzed with the same instrument and GC capillary column as the sterols, stanols, and fatty acids. The oven temperature program started at 50 °C and increased to 320 °C (held 30 min) at 4°C/min. 1 µl of the apolar fractions were injected using splitless mode. Carrier gas (He) was at a constant flow rate (2 mL/min). The apolar fraction was analyzed simultaneously in full scan mode (m/z 35–700) and SIM mode (m/z 57 for n-alkanes and m/z 66 for the internal standard). The transfer line temperature was set at 320 °C and EI voltage at 45 eV. Data was processed using Chromeleon software. Identification of *n*-alkanes was based on their retention time and mass spectra.

## 2.5. Water quality measurements

Water temperature, salinity, dissolved oxygen, and pH were measured by an automated YSI 6000 multi-parameter probe (YSI inc., Yellow springs, OH, USA). Water turbidity was measured with a CTD Probe Sea-bird®. Chlorophyll *a* (Chl *a*) concentrations were measured on water samples after filtration on Whatman GF/F glass fiber filters (0.7 µm). These filters were stored frozen (-20° C) before analysis. Chl *a* was extracted from filters with incubation in 90% acetone (10 mL) in the dark at 4°C for 12 hours. After two centrifugations (1700 g, 5 min), Chl a concentrations were measured by a Turner Designs Fluorometer according to the method of Strickland and Parsons (1972) as described in the reference protocol of SNO SOMLIT (Service d'observation du Milieu Littoral). These measurements were performed at the Université de Toulouse and Université de Caen Normandie.

## 2.6. DOC concentration measurement

The DOC concentrations were determined by using an aliquot of water sample that was acidified and analyzed in Non-Purgeable Organic Carbon (NPOC) mode on a Total Organic Carbon Analyzer (Shimadzu, Tokyo, Japan). For each sample, three replicate analyses were performed. The average value is reported, with the relative standard deviation below 1%.

## 2.7. Spectroscopic analyses

The spectroscopic analyses (absorbance and fluorescence) were performed in a 1 cm (path length) Hellma Suprasil® quartz cell. A Jasco® V-760 spectrophotometer was used to record the UV-Visible absorbance spectra of water samples. The absorbance spectra were acquired at 200 nm/min between 210 and 700 nm. The absorbance spectra of the ultrapure water blank daily acquired was subtracted from the spectrum of each sample. When the highest absorbance was above 0.1, samples were diluted with ultrapure water to avoid an inner-filtering effect in subsequent fluorescence analyses.

The excitation-emission matrix (EEM) fluorescence spectra were obtained between the wavelengths 240 and 800 nm at excitation (2 s integration time, 5 nm intervals) and 245-830 nm at emission (high CCD detector gain, 1 pixel (ca. 0.58 nm intervals)), by using an Aqualog spectrofluorometer (Horiba Scientific, France) equipped with a xenon lamp (150W), a double monochromator at excitation, and a CCD detector. To eliminate Raman and Rayleigh scatter peaks, each EEM was subtracted from the ultrapure water blank EEM spectrum daily acquired. The area of the Raman scattering peak of ultrapure water was calculated daily at the excitation of 350 nm, allowing the spectra to be normalized. The fluorescence intensities are expressed in Raman Units (RU). The EEM spectra were then processed to record fluorescence intensities and calculate

distinct indices, including γ/α (Huguet et al., 2009; Parlanti et al., 2000), HIX (Zsolnay et al., 1999), BIX (Huguet et al., 2009), and FI (McKnight et al., 2001), with the TreatEEM software (Omanović et al., 2023). These Fluorescence indices are summarized in Chapter 1 (Table 1-3).

## 2.8. Parallel factor analysis (PARAFAC)

A multi-way PARAllel FACtor analysis (PARAFAC) can decompose the EEM fluorescence spectra into distinct underlying fluorescent components (Stedmon et al., 2003b). This statistical approach helps to identify the fluorophores that contribute to the overall spectrum dataset and estimate their relative contribution to total DOM fluorescence. The PARAFAC model was performed for 4 to 8 components with non-negativity constraints using the DOM Fluor toolbox (version 1.7) in Matlab R2021b (Stedmon and Bro, 2008). An optimal model can be validated after split-half validation analysis and residual assessment (Murphy et al., 2013; Stedmon and Bro, 2008). The spectral characteristics of the PARAFAC components were then compared to those identified in other studies through an online spectral library (Openfluor) (Murphy et al., 2014). Tucker's congruence coefficient was used to determine the similarity between the model determined in this thesis and those in the online database, with criteria set at 95%.

## 2.9. Machine learning

Generally, Machine Learning (ML) can be categorized into unsupervised and supervised machine learning. Unsupervised learning techniques focus on discovering patterns in unlabeled data, clustering samples into distinct groups (Figure 2-7a). On the other hand, supervised machine learning models use labeled data for predictions (Figure 2-7b). Exploration and application of both

unsupervised and supervised machine learning techniques in estuarine OM research is still relatively limited, but has the potential to enhance our understanding of estuarine OM sources, transformation processes, and develop data-driven strategies for future sample collection (Tao et al., 2021; Yi et al., 2023; Zhao et al., 2023). ML algorithms can efficiently analyze the complex environmental datasets, identifying hidden patterns, trends, and correlations that may not be apparent through traditional methods (e.g. linear regression and analysis of variance).

### 2.9.1 Unsupervised machine learning

Unsupervised machine learning techniques, such as clustering algorithms, have been used in environmental sciences to divide data into different groups with similar properties and therefore reduce data complexity (Kim et al., 2021; Narvaez-Montoya et al., 2023; Rejano et al., 2023).

K-means clustering is a widely used unsupervised clustering technique given its high efficiency and concise algorithm (Li et al., 2016; MacQueen, 1967). This algorithm splits the data into K groups (clusters) and seeks to find centroids that minimize the average Euclidian distance between data points in the same cluster to the centroid (Figure 2-8) (Hartigan and Wong, 1979). The optimal number of clusters (K) can be identified by using the standard elbow method which plots the Within-Cluster Sum of Squares (WCSS) as a function of the number of clusters and selects the elbow of the curve as the number of optimal clusters (K).

$$\text{WCSS}=\sum_{C_k}^{C_n}\left(\sum_{d_i\ in\ C_i}^{d_m} distance(d_i, C_k)^2\right)$$

Where d is the data point in each Cluster and C is the cluster centroids.

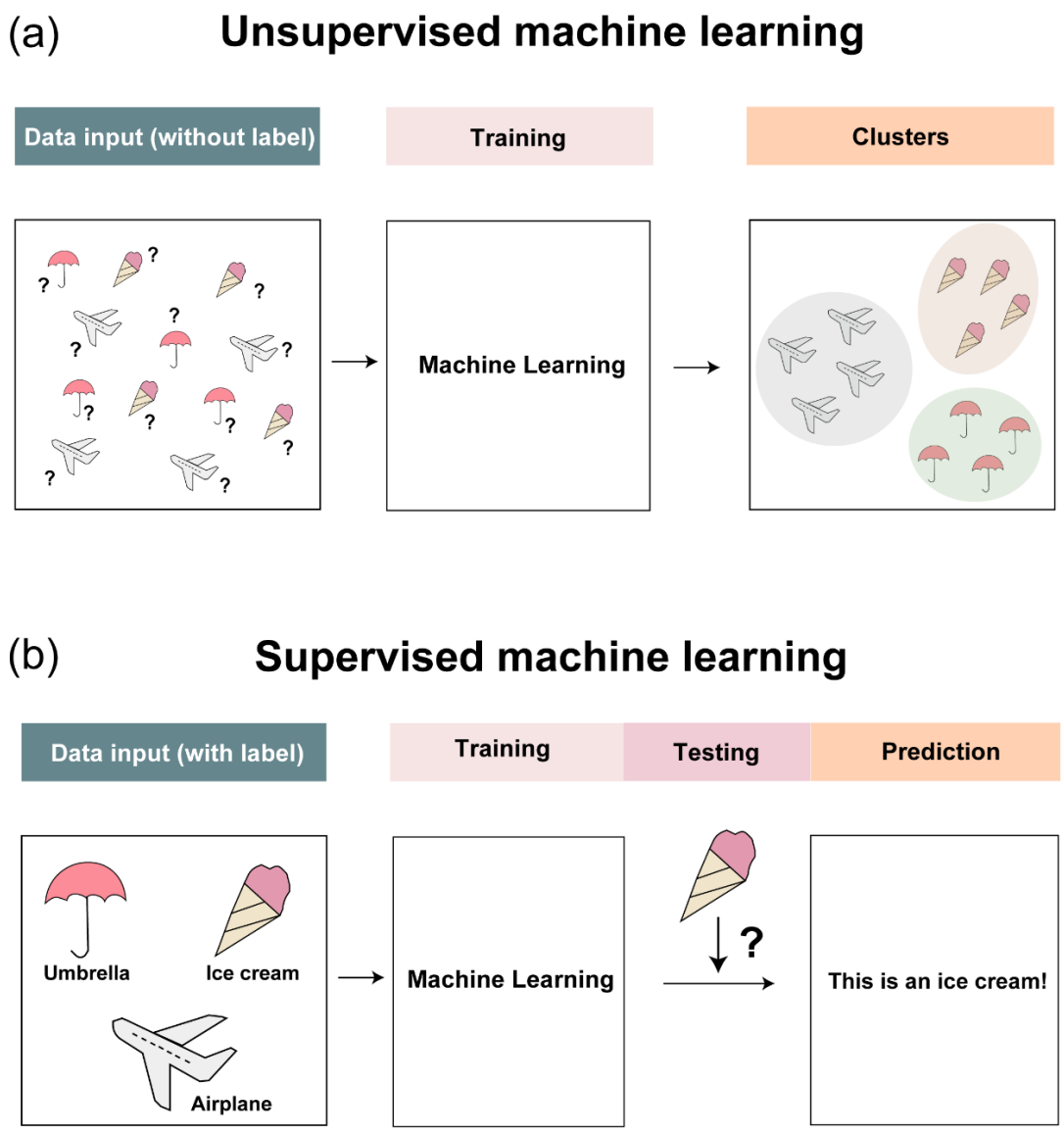



**Figure 2-7**. Schematic plot showing differences between unsupervised and supervised machine learning.

K-means clustering was used in this thesis to find clusters for the DOM optical parameters in an unlabeled dataset, which was performed using the KMeans function from the cluster module of the scikit-learn library (https://github.com/scikitlearn/) (Pedregosa et al., 2011) in Python 3.9.16. The optimal number of clusters (K) was chosen using the elbow method, after plotting WCSS (Within-Cluster Sum of Square) with varying K values from 1 to 10.

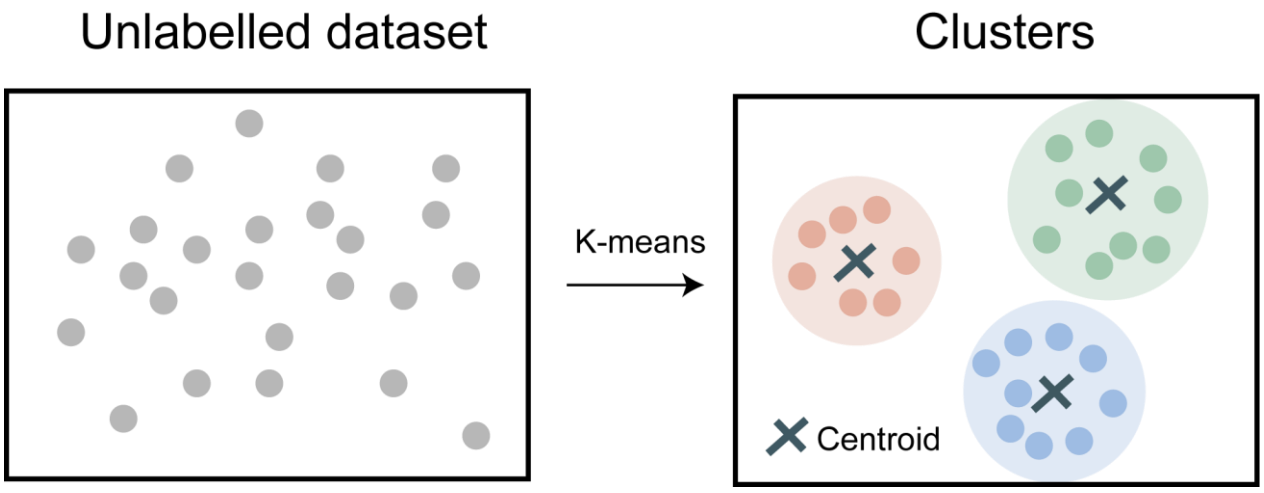



**Figure 2-8**. Schematic plot showing the mechanism of k-means clustering

### 2.9.2 Supervised machine learning

Compared with unsupervised machine learning, supervised machine learning techniques use labeled data to train models and make predictions. The Gradient Boosting Machines (GBM) is a popular supervised machine learning technique, which uses an ensemble of decision trees (weak learners) and makes more accurate predictions on tabular data compared to other machine learning algorithms. GBM algorithms are based on decision trees and are thus robust to multicollinearity. In addition, they can handle missing values, without the need for deletion/imputation of data.

Generally, GBM implementations use level-wise tree growth or leaf-wise tree growth. The level-wise tree growth is a strategy of building decision trees by expanding the tree one level at a time before moving to the next level (Figure 2-9), which can be computationally expensive. Instead of developing all nodes at each level, the leaf-wise tree growth strategy only expands the leaf nodes that contribute the most to minimizing the loss function, which is computationally efficient.

Light Gradient Boosting Machine (LightGBM) is a widely-used GBM framework that takes a leaf-wise tree growth approach, which reduces memory usage and increases the model efficiency

(Ke et al., 2017). LightGBM iteratively trains an ensemble of decision trees to minimize a loss function. This iterative process is repeated until a stopping criterion is reached. LightGBM introduces two techniques that increase the efficiency and scalability: the Gradient-based One Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) (Ke et al., 2017). GOSS retains instances with higher gradients while performing random sampling on instances with smaller gradients, which is a useful technique for obtaining accurate estimation. Meanwhile, EFB decreases the number of features by bundling the exclusive features in the sparse feature space, which increases the training speed without sacrificing accuracy.

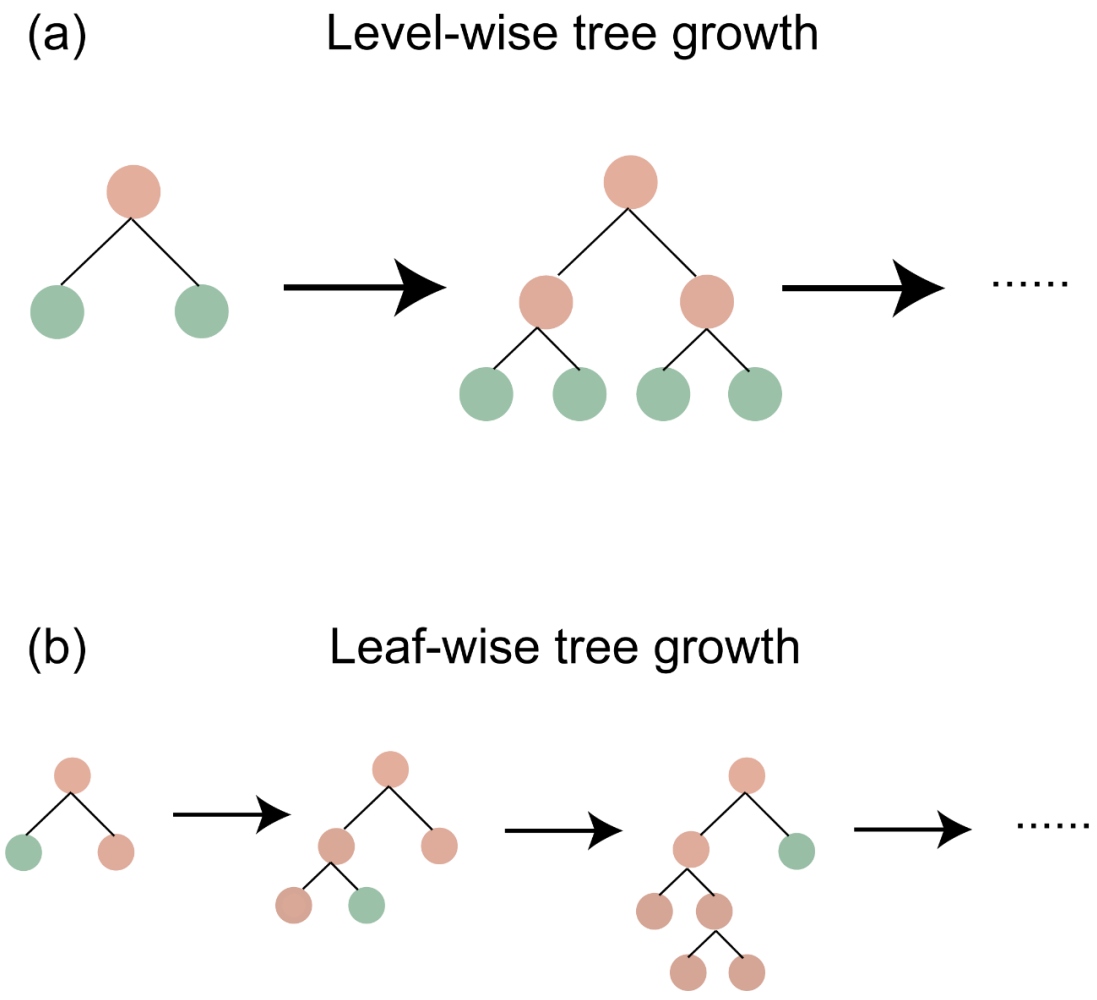



**Figure 2-9**. Schematic plot showing the mechanism of level-wise and leaf-wise tree growth model adapted after LightGBM documentation (https://lightgbm.readthedocs.io/en/latest/Features.html).

In this thesis, LightGBM was used to classify the estuarine zones based on DOM optical properties, implemented using the LightGBM package (https://lightgbm.readthedocs.io) in Pyhton 3.9.16. Before classification, the dataset was firstly divided into training dataset (75%) and test set (25%). The training set is used to fit the machine learning model, whereas the test set (an independent set of new data that was never used in training) is used to assess model performance. The class imbalance problem can be solved using a common framework (Synthetic Minority Oversampling Technique, SMOTE), which occurs when one class includes significantly fewer samples than the other classes (Chawla et al., 2002). This technique is used to oversample an imbalanced training set and is implemented in Python (version 3.9.16) with the imblearn library (Lemaître et al., 2017) (https://github.com/scikit-learn-contrib/imbalanced-learn).

### 2.9.3 Evaluation of the supervised machine learning model

To avoid overfitting and assess the model performance, 10-fold cross-validation experiments were performed. With the 10-fold cross-validation, the training set was divided into 10 parts randomly. The model was trained with nine of these parts and tested using the remaining one, which was repeated for ten times.

The overall accuracy in the independent test set, recall (sensitivity of model prediction), precision (hitting ratio of positive predictions), AU-ROC (area under the receiver operating characteristic curve), and AU-PRC (area under the precision-recall curve) were also used to evaluate the performance of the machine learning model. The ROC curve shows how well the classification model differentiates between classes, with a larger AU-ROC suggesting better model performance. Furthermore, the Precision-Recall Curve (PRC) was used to demonstrate the tradeoff between precision and recall for various thresholds. The PRC is a graph that shows recall on the x-

axis and precision on the y-axis. It is often utilized when classes are imbalanced, with higher AU-PRC indicating better classifier performance.

### 2.9.4 Explainable artificial intelligence

Machine learning models are often seen as black boxes, which makes it difficult to get explanations for the predictions they make. As interpretability and transparency are crucial in understanding and explaining the machine learning model, model explainability has recently become a basic part of the machine learning pipeline. Explainable artificial intelligence is a set of tools (frameworks) to interpret predictions made by the black box machine learning models.

SHapley Additive exPlanations (SHAP) is an explainable artificial intelligence technique based on cooperative game theory. It provides a unified framework for interpreting the output of complex machine learning models, including GBM models (Lundberg and Lee, 2017; Lundberg et al., 2020). It offers a way to evaluate the importance/contribution of each feature to the prediction by calculating the SHAP values. This makes it possible to assess how much each feature influences the prediction and if it has a positive or negative impact on the model output (Lundberg et al., 2020). The SHAP method was used in this study to evaluate the weight/importance of distinct features in the trained machine learning model, with higher SHAP values indicating a stronger positive influence of that feature on the prediction, implemented with the SHAP package (https://github.com/slundberg/shap) in Python (3.9.16). The main DOM characteristics in each class were further identified.

## 2.10. Other Statistical analyses

All statistical analyses were performed using the R software (version 4.2.1). The non-parametric statistical tests were used due to the non-normal distribution of the dataset (tested by Shapiro–Wilk normality test; p-values < 0.05). Specifically, the Spearman's correlation was used to investigate potential correlations among different features, and the unpaired two-samples Wilcoxon test (also known as Mann-Whitney test or Wilcoxon rank sum test) was used for two independent group comparisons. Significance level is indicated by asterisks: *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001; ****p-value < 0.0001; ns (not significant), p-value > 0.05.

Principal Component Analysis (PCA) is often used for visualizing high-dimensional data in lower-dimensional spaces. This visualization helps in exploring patterns, clusters, and relationships that might not be apparent in the original high-dimensional space. In this thesis, PCA was performed to statistically investigate the relationships between samples and variables, implemented with the R packages factoextra and FactoMineR. The different groups of samples were highlighted by adding 95% concentration ellipses.

Redundancy analysis (RDA) was performed using the R package vegan to investigate the relationship between variables. Angles between distinct variables were used to identify the potential correlations. Right angles (90°) reflect a lack of linear correlations, whereas small or straight angles (close to 0° or 180°, respectively) imply positive or negative linear correlations. The variables that are close to each other were assumed to be strongly linked, representing similar distribution patterns. To evaluate the relative importance of each explanatory variable (environmental parameters) on dependent variables, a hierarchical partitioning method implemented in the R package rdacca.hp was utilized. This method calculated the individual

importance (sum of the unique and total average shared effects) from all subset models, generating an unordered assessment of variable importance (Lai et al., 2022).

Spatio-temporal variations in environmental factors and variables were assessed after applying a locally estimated scatterplot smoothing (LOESS) method. This method allows the identification of nonlinear data patterns and buffers the effect of aberrant data and outliers. LOESS was implemented by the geom_smooth function of the R package ggplot2.

## 2.11. Summary of the analysis

Soil samples and mudflat sediments were all analyzed for their elemental and isotopic composition, as well as for brGDGTs and brGMGTs. The analyses performed on the water samples (POM and DOM characterization) are summarized in the Table 2-2.

**Table 2-2.** List of water samples with corresponding analysis

| | | POM characterization | | | | | DOM characterization | |
|---|---|---|---|---|---|---|---|---|
| Date | Location | elemental and isotopic analysis | *n*-alkanes | sterols and stanols | fatty acids | brGDGTs and brGMGTs | UV-visible | EEMs |
| Jun-19 | Honfleur | √ | √ | √ | √ | √ | √ | √ |
| | Berville-Sur-Mer | √ | √ | √ | √ | √ | √ | √ |
| | Tancarville | √ | √ | √ | √ | √ | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | √ | √ | √ | √ | √ | √ | √ |
| | Caudebec | √ | √ | √ | √ | √ | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | √ | √ | √ | √ | √ | √ | √ |
| | Duclair | | | | | | √ | √ |
| | Val des Leux | √ | √ | √ | √ | √ | | |
| | La Bouille | | | | | | √ | √ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Haulot Sur Seine | √ | √ | √ | √ | √ | √ | √ |
| | Petit Couronne | | | | | | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| | Rouen | | | | | | √ | √ |
| | Oissel | √ | √ | √ | √ | √ | | |
| | Les Andelys | √ | √ | √ | √ | √ | | |
| Jul-19 | Honfleur | | | | | | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | √ | √ | √ | √ | √ | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | √ | √ | √ | √ | √ | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | | | | | | √ | √ |
| | Val des Leux | √ | √ | √ | √ | √ | | |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | | | | | | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| | Oissel | √ | √ | √ | √ | √ | | |
| | Les Andelys | √ | √ | √ | √ | √ | | |
| Sep-19 | Honfleur | | | | | | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | | | | | | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | | | | | | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | | | | | | √ | √ |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | | | | | | √ | √ |

| | | | |
|---|---|---|---|
| | Le Grand Quevilly | √ | √ |
| Mar-20 | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| Jun-20 | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Le Trait | √ | √ |
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |
| Jul-20 | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Le Trait | √ | √ |
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sep-20 | Honfleur | √ | √ | √ | √ | √ | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | √ | √ | √ | √ | √ | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | √ | √ | √ | √ | √ | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | √ | √ | √ | √ | √ | √ | √ |
| | Val des Leux | √ | √ | √ | √ | √ | √ | √ |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | √ | √ | √ | √ | √ | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| | Oissel | √ | √ | √ | √ | √ | | |
| | Les Andelys | √ | √ | √ | √ | √ | | |
| Nov-20 | Honfleur | | | | | | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | | | | | | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | | | | | | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | | | | | | √ | √ |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | | | | | | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| | Poses | √ | √ | √ | √ | √ | | |
| | Triel sur Seine | √ | √ | √ | √ | √ | | |
| | Bougival | √ | √ | √ | √ | √ | | |
| | Marnay sur Seine | √ | √ | √ | √ | √ | | |
| Feb-21 | Balise A | | | | | | √ | √ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Honfleur | √ | √ | √ | √ | √ | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | √ | √ | √ | √ | √ | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | √ | √ | √ | √ | √ | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | √ | √ | √ | √ | √ | √ | √ |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | √ | √ | √ | √ | √ | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| Mar-21 | Balise A | | | | | | √ | √ |
| | Honfleur | √ | √ | √ | √ | √ | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | √ | √ | √ | √ | √ | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | √ | √ | √ | √ | √ | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | √ | √ | √ | √ | √ | √ | √ |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | √ | √ | √ | √ | √ | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| May-21 | Balise A | | | | | | √ | √ |
| | Honfleur | | | | | | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | √ | √ | √ | √ | | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Caudebec | √ | √ | √ | √ | | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | | | | | | √ | √ |
| | Val des Leux | √ | √ | √ | √ | | | |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | | | | | | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| | Oissel | √ | √ | √ | √ | | | |
| | Les Andelys | √ | √ | √ | √ | | | |
| Jun-21 | Honfleur | | | | | | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | √ | √ | √ | √ | | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | √ | √ | √ | √ | | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |
| | Duclair | | | | | | √ | √ |
| | Val des Leux | √ | √ | √ | √ | | √ | √ |
| | La Bouille | | | | | | √ | √ |
| | Haulot Sur Seine | | | | | | √ | √ |
| | Petit Couronne | | | | | | √ | √ |
| | Le Grand Quevilly | | | | | | √ | √ |
| | Oissel | √ | √ | √ | √ | | √ | √ |
| | Les Andelys | √ | √ | √ | √ | | √ | √ |
| Jul-21 | Balise A | | | | | | √ | √ |
| | Honfleur | | | | | | √ | √ |
| | Berville-Sur-Mer | | | | | | √ | √ |
| | Tancarville | | | | | | √ | √ |
| | Petitville | | | | | | √ | √ |
| | Vatteville-La-Rue | | | | | | √ | √ |
| | Caudebec | | | | | | √ | √ |
| | Le Trait | | | | | | √ | √ |
| | Heurtauville | | | | | | √ | √ |

| | | | |
|---|---|---|---|
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |
| Sep-21 | Balise A | √ | √ |
| | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Le Trait | √ | √ |
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |
| Jan-22 | Balise A | √ | √ |
| | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Le Trait | √ | √ |
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |
| May-22 | Balise A | √ | √ |

| | | | |
|---|---|---|---|
| | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |
| Jul-22 | Balise A | √ | √ |
| | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Le Trait | √ | √ |
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |
| Sep-22 | Balise A | √ | √ |
| | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Le Trait | √ | √ |

| | | | |
|---|---|---|---|
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |
| Nov-22 | Balise A | √ | √ |
| | Honfleur | √ | √ |
| | Berville-Sur-Mer | √ | √ |
| | Tancarville | √ | √ |
| | Petitville | √ | √ |
| | Vatteville-La-Rue | √ | √ |
| | Caudebec | √ | √ |
| | Le Trait | √ | √ |
| | Heurtauville | √ | √ |
| | Duclair | √ | √ |
| | La Bouille | √ | √ |
| | Haulot Sur Seine | √ | √ |
| | Petit Couronne | √ | √ |
| | Le Grand Quevilly | √ | √ |

# Chapter 3:

## Environmental controls on the brGDGT and brGMGT distributions across the Seine River basin (NW France): Implications for bacterial tetraethers as a proxy for riverine runoff

This chapter is in preparation for submission to *Biogeosciences*

## Abstract

Branched glycerol dialkyl glycerol tetraethers (brGDGTs) are bacterial lipids that have been largely used as environmental proxies in continental paleorecords. Another group of related lipids, branched glycerol monoalkyl glycerol tetraethers (brGMGTs), has recently been proposed as a potential paleotemperature proxy. Nevertheless, the sources and environmental dependencies of both brGDGTs and brGMGTs along the river-sea continuum are still poorly understood, complicating their application as paleoenvironmental proxies in aquatic settings. In this study, the sources of brGDGTs and brGMGTs and the potential factors controlling their distributions are explored across the Seine River basin (NW France), which encompasses the freshwater to seawater continuum. To this aim, brGDGTs and brGMGTs were analyzed in soils, Suspended Particulate Matter (SPM) and sediments ($n$=237) collected all along this basin, from land to sea. Both types of compounds are shown to be produced *in situ*, in freshwater as well as saltwater. Redundancy analysis further shows that both salinity and nitrogen loadings dominantly control the brGDGT distributions. Furthermore, the relative abundance of 6-methyl *vs*. 5-methyl brGDGTs ($IR_{6Me}$ ratio), Total Nitrogen (TN), $\delta^{15}N$ and chlorophyll *a* concentration co-vary in a specific zone with low salinity, suggesting that 6-methyl brGDGTs are preferentially produced under low-salinity and high-productivity conditions. In contrast with brGDGTs, brGMGT distribution appears to be primarily regulated by salinity, with a distinct influence on the individual homologues. Salinity is positively correlated with homologues H1020a and H1020b, and negatively correlated with compounds H1020c and H1034b. This suggests that bacteria thriving in freshwater preferentially produce compounds H1020c and H1034b, whereas bacteria primarily growing in saltwater appear to be predominantly responsible for the production of homologues H1020a and H1020b. Based on

the abundance ratio of the freshwater-derived compounds (H1020c and H1034b) *vs.* saltwater-derived homologues (H1020a and H1020b), a novel proxy, Riverine IndeX (RIX) is proposed to trace riverine organic matter inputs, with high values (>0.5) indicating higher riverine contribution. As RIX relies on compounds that are specifically produced in certain settings (freshwater or saltwater), this index has potential to serve as a powerful proxy for riverine runoff in modern samples as well as in paleorecords.

*Keywords:* branched GDGTs; branched GMGTs; environmental proxies; land-ocean continuum; riverine runoff

## 3.1. Introduction

Branched glycerol dialkyl glycerol tetraethers (brGDGTs) are membrane lipids produced by unknown bacteria, although some of them were attributed to the phylum *Acidobacteria* (Sinninghe Damsté et al., 2011; Chen et al., 2022; Halamka et al., 2022). These compounds were observed to occur ubiquitously in a wide range of terrestrial and aquatic environments (Schouten et al., 2013; Raberg et al., 2022). The distribution of brGDGTs (number of cyclopentane moieties and methyl groups; cf. structures in Figure 1-5 in Chapter 1) was empirically linked with pH and Mean Annual Air Temperature (MAAT) in soils (Weijers et al., 2007; De Jonge et al., 2014; Véquaud et al., 2022), peats (Naafs et al., 2017; Véquaud et al., 2022) and lake sediments (Martínez-Sosa et al., 2021). The brGDGT-based proxies (i.e. MBT'$_{5ME}$ and CBT') have been largely applied to reconstruct MAAT and pH from sedimentary archives (Coffinet et al., 2018; Harning et al., 2020; Wang et al., 2020).

In aquatic settings, brGDGTs were initially suggested to be predominantly derived from watershed soils and transported by erosion in the sediments (Hopmans et al., 2004). Based on this assumption, the Branched and Isoprenoid Tetraethers (BIT) index was defined as the abundance ratio of the major brGDGTs to crenarchaeol (isoprenoid GDGT mainly produced by marine *Thaumarchaeota*). It is comprised between 0 and 1, with high BIT values (around 1) reflecting higher contribution of terrestrial organic matter compared to marine organic matter (Hopmans et al., 2004). Over the last years, the BIT index has been broadly used for quantifying the relative contribution of terrestrial organic matter in aquatic systems (Xu et al., 2020; Yedema et al., 2023) and evaluating the reliability of $TEX_{86}$ palaeothermometer (Cramwinckel et al., 2018). However, several studies have shown that brGDGTs can also be produced *in situ* in aquatic settings, either in rivers (e.g. De Jonge et al., 2015; Freymond et al., 2017; Kim et al., 2015; Zell et al., 2014, 2013)

or lakes (Tierney and Russell, 2009), adding complication for the identification of brGDGT sources in aquatic ecosystems and for the application of the brGDGTs as (paleo)environmental proxies, including the BIT index. The BIT values have all the more to be carefully interpreted as they could also be influenced by the selective degradation of branched *vs*. isoprenoid GDGTs (Smith et al., 2012). Thus, complementary molecular proxies for quantifying the input of terrestrial organic matter to aquatic settings are still needed, which may cross-validate other available terrestrial proxies, such as the $\delta^{13}C$ of organic carbon (Lamb et al., 2006), heterocyst glycolipids (Kang et al., 2023), and long-chain diols (Lattaud et al., 2017).

The improvement of analytical methods allowed the separation and quantification of 5-, 6- and 7-methyl brGDGTs (methyl groups at the fifth, sixth, and seventh positions; Supplementary Figure 3-1), that in previous chromatographic protocols co-eluted (De Jonge et al., 2014, 2013; Ding et al., 2016). Compounds eluting later than 7-methyl brGDGTs are tentatively designated 1050d and 1036d, as their exact chemical structures are currently unknown (Wang et al., 2021). The fractional abundance of the individual brGDGT isomers was shown to be influenced by distinct environmental factors. For example, the relative abundance of 5-methyl brGDGTs was correlated with temperature, whereas that of 6-methyl brGDGTs was correlated with pH (De Jonge et al., 2014). In addition to temperature and pH, other environmental factors may influence brGDGT distributions in terrestrial and aquatic settings and hence the application and interpretation of brGDGT-derived proxies. For example, recent studies in lakes observed an influence of salinity on the relative abundance of 6-methyl, 7-methyl brGDGTs and their late-eluting compounds (Wang et al., 2021; Kou et al., 2022). This suggests that salinity could also control the distribution of these compounds in other systems like river-sea continuums but this assumption has not yet been studied.

Compared with brGDGTs, the branched glycerol monoalkyl glycerol tetraethers (brGMGTs) are a much less studied group of lipids. Recent studies have revealed their presence in

diverse environments, including peatlands (Naafs et al., 2018; Tang et al., 2021), marine settings (Liu et al., 2012; Xie et al., 2014), rivers (Kirkels et al., 2022a), soils (Baxter et al., 2021; Kirkels et al., 2022a) and lakes (Baxter et al., 2021, 2019). BrGMGTs are labelled as H1020, H1034, and H1048 respectively (cf. in Figure 1-6 in Chapter 1), with isomers suggested by a suffix letter (a-c) following the order in which they elute according to Baxter et al. (2019). These compounds are structurally similar to brGDGTs, but possess an additional covalent carbon–carbon bond between the alkyl chains, leading to "H-shaped" structure. The bridge of brGMGTs was considered to be a primary adaptation to heat stress (Naafs et al., 2018; Baxter et al., 2019). Their presumed membrane stability under high temperature conditions was inferred from the behaviour of isoprenoid glycerol monoalkyl glycerol tetraethers (isoGMGTs), which were identified in a hyperthermophilic methanogen (Morii et al., 1998) and deep-sea hydrothermal vents (Schouten et al., 2008). Although a rigorous chemical characterization of brGMGTs is lacking and the source organisms of brGMGTs are unknown, correlations between the relative abundances of brGMGTs and MAAT were observed in peat soils (Naafs et al., 2018) and lakes (Baxter et al., 2019), showing their potential as temperature proxies. In addition to temperature, anoxic conditions may also trigger brGMGT production in the anoxic zone of peats (Naafs et al., 2018; Tang et al., 2021), anoxic part of the water column and/or sediments in lakes (Baxter et al., 2021), in regularly inundated soils (Kirkels et al., 2022a), as well as in the oxygen minimum zone in the marine environments (Xie et al., 2014). Furthermore, shifts in microbial community composition in response to other unknown environmental factors seem to control the relative abundances of brGMGTs in peats and lignites (Elling et al., 2023). Henceforth, in order to use the brGMGT as environmental proxies in sedimentary records, it is still necessary to understand which factors control their distributions in riverine and marine water columns and sediments, which remain to date poorly understood (Bijl et al., 2021; Sluijs et al., 2020).

Based on previous studies of brGDGTs and brGMGTs in terrestrial and marine settings (Dearing Crampton-Flood et al., 2019; Wang et al., 2021; Kirkels et al., 2022a, 2022b; Kou et al., 2022), we hypothesize (1) that both brGDGTs and brGMGTs can be produced *in situ* in aquatic systems and (2) that brGDGT and brGMGT distribution are influenced by surrounding environmental factors and vary spatially along the land-sea continuum. These compounds have a potential to be used as proxies of riverine organic matter inputs along estuaries. These hypotheses were tested by examining and comparing the distribution of brGDGTs and brGMGTs in soils, suspended particulate matter (SPM) and sediments ($n$ = 237) collected all along the Seine River basin (NW France), covering its riverine and estuarine parts. The aim of the present study was (1) to investigate the sources of brGDGTs and brGMGTs along the Seine land-sea continuum, (2) to determine the predominant environmental controls affecting the distribution of these molecules and (3) to assess the potential of brGMGTs as a riverine runoff proxy.

## 3.2. Material and methods

### 3.2.1. Study area

The Seine River basin (Seine River and its estuary; Figure 3-1a) is more than 760km long and is characterized by high population density, draining through the greater Paris region (over 12 million inhabitants) to the English Channel (Flipo et al., 2021). The Seine Estuary is a macrotidal estuary according to its high tidal range, small depth and morphology. The maximum flows are generally observed in winter (over 700 $m^3/s$; Figure 3-1b), whereas the minimum flows are observed in summer (below 250 $m^3/s$; Figure 3-1b). The tide influences the estuary up to the city of Poses (site 5, KP 202 in Figure 3-1a; KP represents kilometric point and is defined as the distance

in kilometers from the city of Paris), where a dam constitutes the boundary between the river and the estuary. The estuary can be divided into two major parts: the upper section mainly influenced by freshwater (KP 202 to KP 298, from site 5 to site 12; Figure 3-1a and Table 3-1) and the lower section (starting at KP 298, from site 12 to the coastal area; Figure 3-1a and Table 3-1).

**Table 3-1.** Location of the sampling sites along the Seine Basin, with the type of samples collected

| Site | Name | Longitude (°) | Latitude (°) | KP | Zone | Date | Type |
|---|---|---|---|---|---|---|---|
| 1 | Marnay-sur-Seine | 3.56 | 48.51 | -200 | River | 2020-11 | SPM (*n*=1) |
| 2 | Bougival | 2.13 | 48.87 | 40 | River | 2020-11 | SPM (*n*=1) |
| 3 | Triel sur Seine | 2.00 | 48.98 | 80 | River | 2020-11 | SPM (*n*=1) |
| 4 | Les Andelys | 1.40 | 49.24 | 175 | River | 2019-6; 2019-7; 2020-9 | SPM (*n*=6) |
| 5 | Poses | 1.24 | 49.31 | 202 | Upstream estuary | 2016-4; 2020-11 | SPM (*n*=2) |
| 6 | Oissel | 1.10 | 49.34 | 229.4 | Upstream estuary | 2019-6; 2019-7; 2020-9 | SPM (*n*=18) |
| 7 | Rouen | 1.03 | 49.43 | 243 | Upstream estuary | 2016-4 | SPM (*n*=1); Sediments (*n*=10) |
| 8 | Petit Couronne | 1.01 | 49.38 | 251.3 | Upstream estuary | 2020-9; 2021-2; 2021-3 | SPM (*n*=3) |
| 9 | Haulot Sur Seine | 0.98 | 49.36 | 255.6 | Upstream estuary | 2019-6 | SPM (*n*=1) |
| 10 | Val-des-Leux | 0.92 | 49.40 | 265.55 | Upstream estuary | 2019-6; 2019-7; 2020-9 | SPM (*n*=18) |
| 11 | Duclair | 0.87 | 49.48 | 278 | Upstream estuary | 2020-9; 2021-2; 2021-3 | SPM (*n*=3) |
| 12 | Heurtauville | 0.82 | 49.45 | 297.65 | Downstream estuary | 2019-6 | SPM (*n*=1) |
| 13 | Caudebec | 0.75 | 49.52 | 310.5 | Downstream estuary | 2015-4; 2015-9; 2016-4; 2019-6; 2019-7; 2020-9; | SPM (*n*=24) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | 2021-2; 2021-3 | |
| 14 | Vatteville-La-Rue | 0.67 | 49.47 | 318 | Downstream estuary | 2019-6 | SPM (*n*=1) |
| 15 | Tancarville | 0.47 | 49.47 | 337 | Downstream estuary | 2015-1; 2015-4; 2015-9; 2019-6; 2019-7; 2020-9; 2021-2; 2021-3 | SPM (*n*=24); Sediments (*n*=20) |
| 16 | Berville-Sur-Mer | 0.37 | 49.44 | 346 | Downstream estuary | 2019-6 | SPM (*n*=1) |
| 17 | Fatouville | 0.32 | 49.44 | 350 | Downstream estuary | 2015-4; 2015-7; 2015-9; 2016-4 | SPM (*n*=4); Sediments (*n*=28) |
| 18 | Honfleur | 0.23 | 49.43 | 355.8 | Downstream estuary | 2015-4; 2015-9; 2019-6; 2020-9; 2021-2; 2021-3 | SPM (*n*=6) |
| 19 | La Carosse | 0.03 | 49.48 | 370 | Downstream estuary | 2015-7; 2016-4; 2016-4 | SPM (*n*=2); Sediments (*n*=10) |
| A | n.a. | 3.72 | 48.56 | n.a. | Soil (around the river) | 2021-9 | Soil (*n*=1) |
| B | n.a. | 3.23<br>3.26 | 48.43<br>48.42 | n.a. | Soil (around the river) | 2021-9 | Soil (*n*=5) |
| C | n.a. | 3.11<br>3.13 | 48.83<br>48.85 | n.a. | Soil (around the river) | 2021-10 | Soil (*n*=3) |
| D | n.a. | 0.38<br>0.38<br>0.38 | 49.47<br>49.46<br>49.45 | n.a. | Soil (around the downstream estuary) | 2021-3; 2021-9 | Soil (*n*=8) |
| E | n.a. | 0.41<br>0.41 | 49.44<br>49.45 | n.a. | Soil (around the downstream estuary) | 2018-2; 2018-6; 2018-8; 2018-10; 2020-9; 2021-3 | Soil (*n*=34) |

### 3.2.2. Sampling

From June 2019 to March 2021, water samples (*n*=102) were collected across the Seine River (Figure 3-1a). Sub-surface water (ca. 1m depth) samples were collected in high-flow (over 250 $m^3/s$) and low-flow (below 250 $m^3/s$) periods from the three zones (river, upstream estuary and downstream estuary) of the Seine River basin (Table 3-1). At 5 sites (sites 4, 6, 10, 13, and 15, Figure 3-1a and Table 3-1), both sub-surface and bottom water (2.2-16 m depth) samples were retrieved using a pump into precleaned 20L FLPE Nalgene carboys. Estuarine water samples (sites 6, 10, 13, and 15; Figure 3-1a and Table 3-1) were collected at three tide periods (high tide, low tide and mid tide). For these sites, 0.25-43 L of water were immediately filtered using pre-combusted Whatman GF/F 0.7 µm glass fiber filters. After filtration, filters were freeze-dried, scratched and stored frozen at -20°C prior to analysis.

Additional SPM samples (*n*=16; Table 3-1) used in this study for brGDGT and brGMGT analysis were collected from the upstream and downstream estuary (site 5, 7, 13, 15, 17, 18, and 19; Figure 3-1a and Table 3-1) in 2015 and 2016, as detailed by Thibault et al. (2019). Sediments (*n*=68) from 8 cores (10cm depth) were collected in the river channel at the same sites as these SPM samples in 2015 and 2016 using a UWITEC corer as described by Thibault et al. (2019) (Table 3-1). These sediments were further sliced (1-cm thickness) and freeze-dried. Surficial soils (*n*=9) were collected in the lateral area of the upstream section of the Seine river in 2021 (site A, B, and C, Figure 3-1a and Table 3-1) and freeze-dried. Additional wetland soils and mudflat sediments (*n*=42) were collected in the downstream estuary in 2018, 2020, and 2021 (site D and E, Figure 3-1a and Table 3-1), representing allochthonous material transported into the estuary by tidal effect. These samples were collected at low tide using a plexiglass® core (4.5 cm depth), and

back to the laboratory, homogenized, freeze-dried, and ground using a ball mill (model MM400, Retsch®).

### 3.2.3. Elemental and isotopic analyses

Elemental and isotopic analyses of the soils (surficial soils and mudflat sediments, *n*=51) and SPM (*n*=102) collected from 2018 to 2021 were performed following the method described in Thibault et al. (2019). Briefly, 40 mg of SPM and 1 g of soils/sediments samples were firstly decarbonated by adding 10 mL of 3 M HCl for 2 h with magnetic stirring at room temperature. Subsequently, these samples were rinsed using ultrapure water and centrifuged until reaching neutral pH. The obtained decarbonated samples were stored at −20 °C and freeze dried. Both decarbonated and non-decarbonated samples (~6 mg for SPM and ~20 mg for soils) were enclosed in a tin capsule. Total Organic Carbon content (TOC) and stable carbon isotopic composition ($\delta^{13}C$) were measured in decarbonated samples using an elemental analyzer coupled with an isotope ratio mass spectrometer (Thermo Fisher Scientific Delta V Advantage) at the ALYSES platform (Sorbonne University / IRD, Bondy, France). Total Nitrogen (TN) and nitrogen isotope ($\delta^{15}N$) were measured in non-decarbonated samples as acidification could impact the N contents (Ryba and Burgess, 2002). The isotopic composition ($\delta^{13}C$ or $\delta^{15}N$) was expressed as relative difference between isotopic ratios in samples and in standards (Vienna Pee Dee Belemnite for carbon or atmospheric $N_2$ for nitrogen). Additional elemental and isotopic analyses of SPM and sediments collected in 2015 and 2016 (*n*=84) were carried out as described in Thibault et al. (2019).

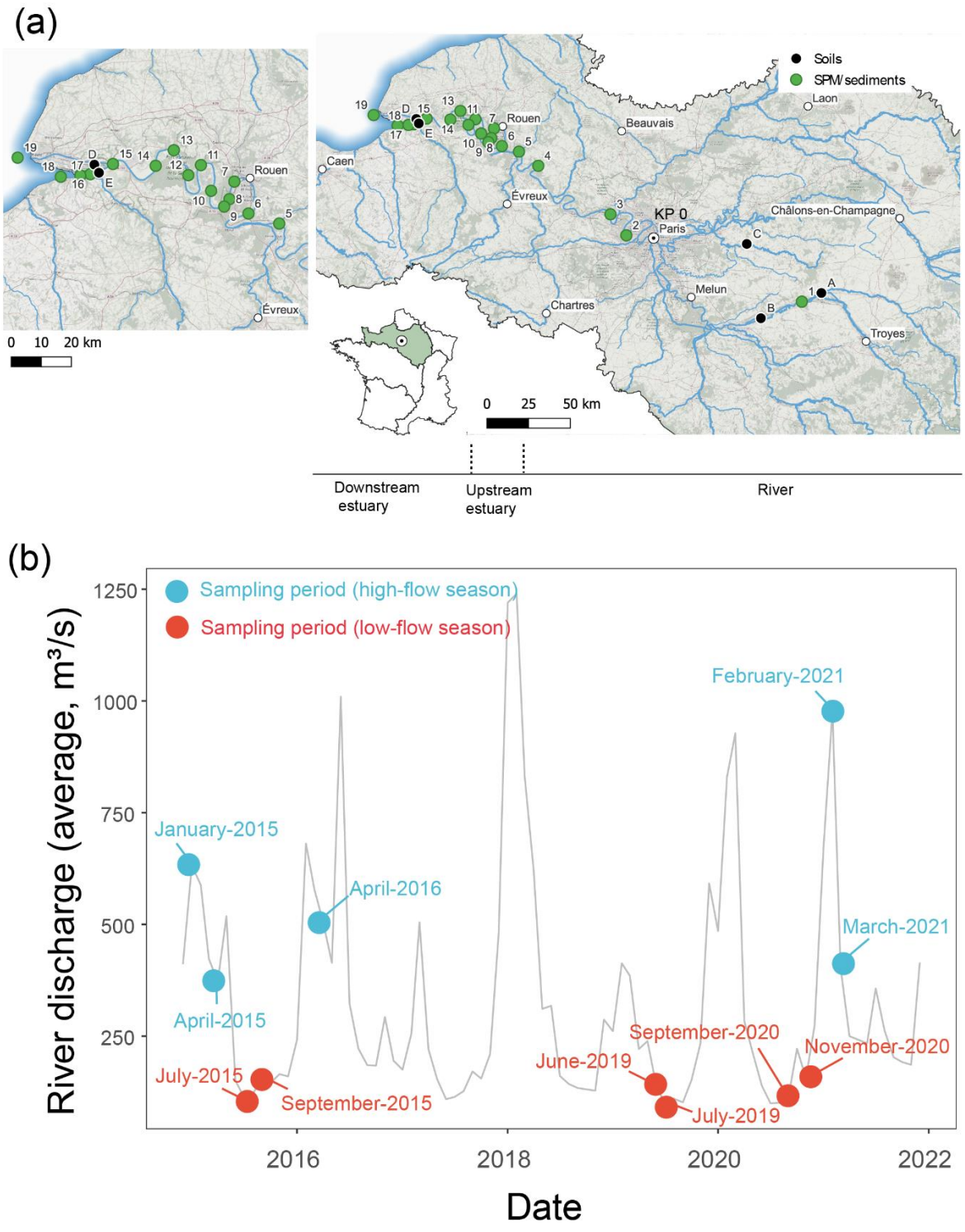



**Figure 3-1.** (a) Geographical locations of sampling sites in the Seine River Basin (KP: kilometric point, the distance in kilometers from the city of Paris (KP 0)). (b) Mean monthly water discharge for the Seine River at the Paris Austerlitz station from 2015 to 2021 (data from https://www.hydro.eaufrance.fr/). Bullets represent the sampling period in high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) conditions.

### 3.2.4. Lipid extraction and analyses

The lipids from surficial soils and mudflat sediments (4-20g, *n*=51) and from SPM samples (~150mg, *n*=102) were extracted ultrasonically (3×) with 20 to 40 mL of dichloromethane (DCM): methanol (MeOH) (5/1, v/v) per extraction. Lipids from the SPM and sediments samples (*n*=84) collected in 2015 and 2016 were previously extracted by Thibault (2018) following the same method. The total lipid extracts were then separated into fractions of increasing polarity on an activated silica gel column, using (i) 30 mL of heptane, (ii) 30 mL of heptane:DCM (1/4, v/v), and (iii) 30 mL of DCM/MeOH (1/1, v/v) as eluents. An aliquot (30%) of the third (polar) fraction containing GDGTs and GMGTs was dried, re-dissolved in heptane, and passed through a 0.2µm polytetrafluoroethylene (PTFE) filter (Ultrafree-MC; Merck). $C_{46}$ Glycerol Trialkyl Glycerol Tetraether (GTGT) was used as an internal standard (Huguet et al., 2006). 5 µl of this standard (0.01025 mg/mL) was typically added to 45 µl of sample.

GDGTs and GMGTs were analyzed using a Shimadzu LCMS 2020 high pressure liquid chromatography coupled with mass spectrometry with an atmospheric pressure chemical ionization source (HPLC-APCI-MS) in selected ion monitoring mode, modified from Hopmans et al. (2016) and Huguet et al. (2019). Tetraether lipids were separated with two silica columns in tandem (BEH HILIC columns, 2.1 × 150 mm, 1.7 µm; Waters) thermostated at 30°C. Injection volume was 30 µL. The flow rate was set at 0.2 mL/min. GDGTs and GMGTs were eluted isocratically for 25 min with 82% A/18% B (A= hexane, B=hexane/isopropanol 9/1, v/v), followed by a linear gradient to 65% A/35% B in 25 min, then a linear gradient to 100% B in 30 min, and back to 82% A/18% B in 4 min, maintained for 50 min. Identification of the different brGMGT isomers was achieved by comparison of peak retention time with that of known brGMGTs in Baxter et al. (2019) and Kirkels et al. (2022a). Semi-quantification of brGDGTs and brGMGTs was performed by comparing the

integrated signal of the respective compound with the signal of a $C_{46}$ synthesized internal standard (Huguet et al., 2006) assuming their response factors to be identical.

The detection limit was set at a signal-to-noise ratio (SNR) of 3. Peaks with lower SNR (<3) are not distinguishable from the background noise and are considered below the limit of quantification.

### 3.2.5. Calculation of GDGT proxies

The $IR_{6Me}$ index represents the proportion of 6-methyl brGDGTs *vs*. 5-methyl brGDGTs and was calculated according to De Jonge et al. (2015; Eq. 1) with Roman numbers referring to the structures in annex (Supplementary Figure 3-1):

$$IR_{6Me} = \frac{II_{a_6}+II_{b_6}+II_{c_6}+III_{a_6}+III_{b_6}+III_{c_6}}{II_{a_5}+II_{b_5}+II_{c_5}+II_{a_6}+II_{b_6}+II_{c_6}+III_{a_5}+III_{b_5}+III_{c_5}+III_{a_6}+III_{b_6}+III_{c_6}} \quad (1)$$

The BIT index including the 6-methyl brGDGTs was calculated following De Jonge et al. (2015; Eq. 2):

$$BIT = \frac{I_a+II_{a_5}+II_{a_6}+III_{a_5}+III_{a_6}}{I_a+II_{a_5}+II_{a_6}+III_{a_5}+III_{a_5}+crenarchaeol} \quad (2)$$

Based on duplicate injections, the average analytical error was 0.005 for $IR_{6Me}$ and 0.06 for BIT.

### 3.2.6. Water quality measurements

Water turbidity was measured by a CTD Probe Sea-bird®. Water temperature, dissolved oxygen, salinity, and pH were measured using an automated YSI 6000 multi-parameter probe (YSI inc., Yellow springs, OH, USA). Chlorophyll *a* (Chl *a*) concentrations were measured on water samples after filtration on Whatman GF/F 0.7 µm glass fiber filters, which were stored frozen (-20° C) before analysis. Chl *a* was extracted from filters with incubation in 10 ml of 90% acetone for 12 hours in the dark at 4°C. After two centrifugations (1700 g, 5 min), Chl *a* concentrations

were measured using a Turner Designs Fluorometer according to the method of Strickland and Parsons (1972) as described in the reference protocol of SNO SOMLIT (Service d'observation du Milieu Littoral). Water quality measurements were performed at the Laboratoire Ecologie Fonctionnelle et Environnement (Université de Toulouse) as well as at UMR BOREA (Université de Caen Normandie).

### 3.2.7. Statistical analyses

All statistical analyses were performed using the R software (version 4.2.1). The non-parametric statistical tests were used due to the non-normal distribution of the dataset (tested by Shapiro–Wilk normality test; p-values < 0.05). Specifically, the Spearman's correlation was used to investigate potential correlations among different features (environmental parameters, fractional abundances of brGDGTs and brGMGTs, and proxies derived from these compounds), and the unpaired two-samples Wilcoxon test (also known as Mann-Whitney test or Wilcoxon rank sum test) was used for two independent group comparisons. Significance level is indicated by asterisks: *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001; ****p-value < 0.0001; ns (not significant), p-value > 0.05.

A Principal Component Analysis (PCA) was performed on the fractional abundances of brGDGTs and brGMGTs, using the R packages factoextra and FactoMineR. The different groups of samples were highlighted by adding 95% concentration ellipses. The proportion of variance in brGDGT and brGMGT compositions that can be explained by different groups was evaluated by permutational multivariate analysis of variance using distance matrices (adonis) in the adonis2 function of the R package Vegan, using the Bray-Curtis distances and 999 permutations.

A Redundancy analysis (RDA) was performed using the R package vegan to investigate the relationship between environmental parameters and brGDGT or brGMGT distributions in

SPM. Angles between brGDGTs or brGMGTs and environmental factors were used to identify the potential correlations. Right angles (90°) reflect a lack of linear correlations, whereas small or straight angles (close to 0° or 180°, respectively) imply positive or negative linear correlations. The compounds that are close to each other were assumed to be strongly linked, representing similar distribution patterns and comparable responses to the environmental conditions. To evaluate the relative importance of each explanatory variable (environmental parameters) on brGDGT or brGMGT distributions, a hierarchical partitioning method implemented in the R package rdacca.hp was used. This method calculated the individual importance (sum of the unique and total average shared effects) from all subset models, generating an unordered assessment of variable importance (Lai et al., 2022).

Spatio-temporal variations of environmental factors and proxies derived from brGDGTs and brGMGTs were assessed after applying a locally estimated scatterplot smoothing (LOESS) method. This method allows the identification of nonlinear data patterns and buffers the effect of aberrant data and outliers. LOESS was implemented by the geom_smooth function of the R package ggplot2.

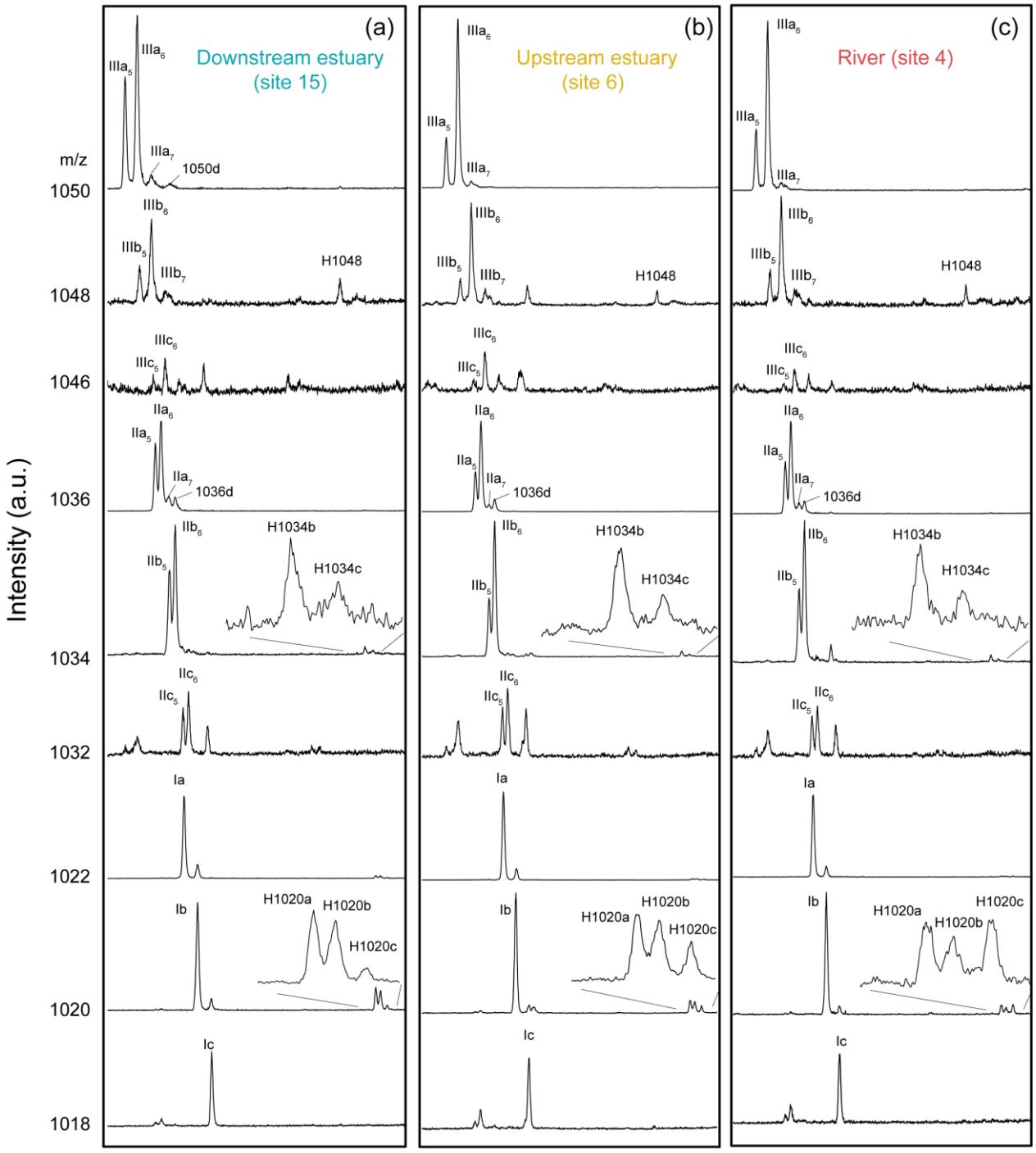



**Figure 3-2.** Extracted chromatograms of brGDGTs and brGMGTs for the SPM samples collected in (a) site 15 (Tancarville, September 2020), (b) site 6 (Oissel, July 2019) and (c) site 4 (Les Andelys, July 2019). The nomenclature for the penta- and hexamethylated brGDGTs: 5-methyl brGDGTs ($IIIa_5$, $IIIb_5$, $IIIc_5$, $IIa_5$, $IIb_5$, and $IIc_5$); 6-methyl brGDGTs ($IIIa_6$, $IIIb_6$, $IIIc_6$, $IIa_6$, $IIb_6$, and $IIc_6$); 7-methyl brGDGTs ($IIIa_7$, $IIIb_7$, and $IIa_7$).

## 3.3. Results

### 3.3.1. Distribution of bulk parameters from land to sea

The total organic carbon (TOC) content was significantly higher in the upstream estuary (4.64±1.42 %, based on SPM and river channel sediments) than in downstream estuary (3.30±1.69 %, based on SPM and sediments), soils (3.02±3.49 %, based on surficial soils and mudflat sediments) and river (2.88±1.14 %, based on SPM) (Figure 3-3). The total nitrogen (TN) content was higher in the upstream estuary (0.51±0.17 %, based on SPM and sediments) than in the river (0.37±0.15 %, based on SPM), downstream estuary (0.31±0.14 %, based on SPM and river channel sediments), and soils (0.24±0.17 %, based on surficial soils and mudflat sediments) (Figure 3-3). Much lower values of $\delta^{13}C_{org}$ were observed in river (-31.30±1.91 ‰, based on SPM) and upstream estuary (-30.62±1.66 ‰, based on SPM and sediments) than in the downstream estuary (-26.45±1.34 ‰, based on SPM and sediments) and soils (-26.55±1.13 ‰, based on surficial soils and mudflat sediments) (Figure 3-3). In addition, no significant differences in $\delta^{15}N$ were observed along the river basin (Figure 3-3).

### 3.3.2. Distribution of brGDGTs from land to sea

The different brGDGTs were detected in all studied samples. The brGDGT chromatograms from upstream samples (SPM and river channel sediments) differed markedly from downstream estuarine samples (SPM and sediments). For example, 6-methyl brGDGTs were much more abundant than 5-methyl brGDGTs in the river (SPM) and upstream estuary (SPM), whereas the strong predominance of 6-methyl *vs.* 5-methyl brGDGTs decreased in the downstream SPM samples (Figure 3-2). Furthermore, the peaks of the recently described 7-methyl brGDGTs and

their late-eluting isomers (i.e. 1050d) were more pronounced in the downstream estuary than in the rest of the Seine basin (Figure 3-2).

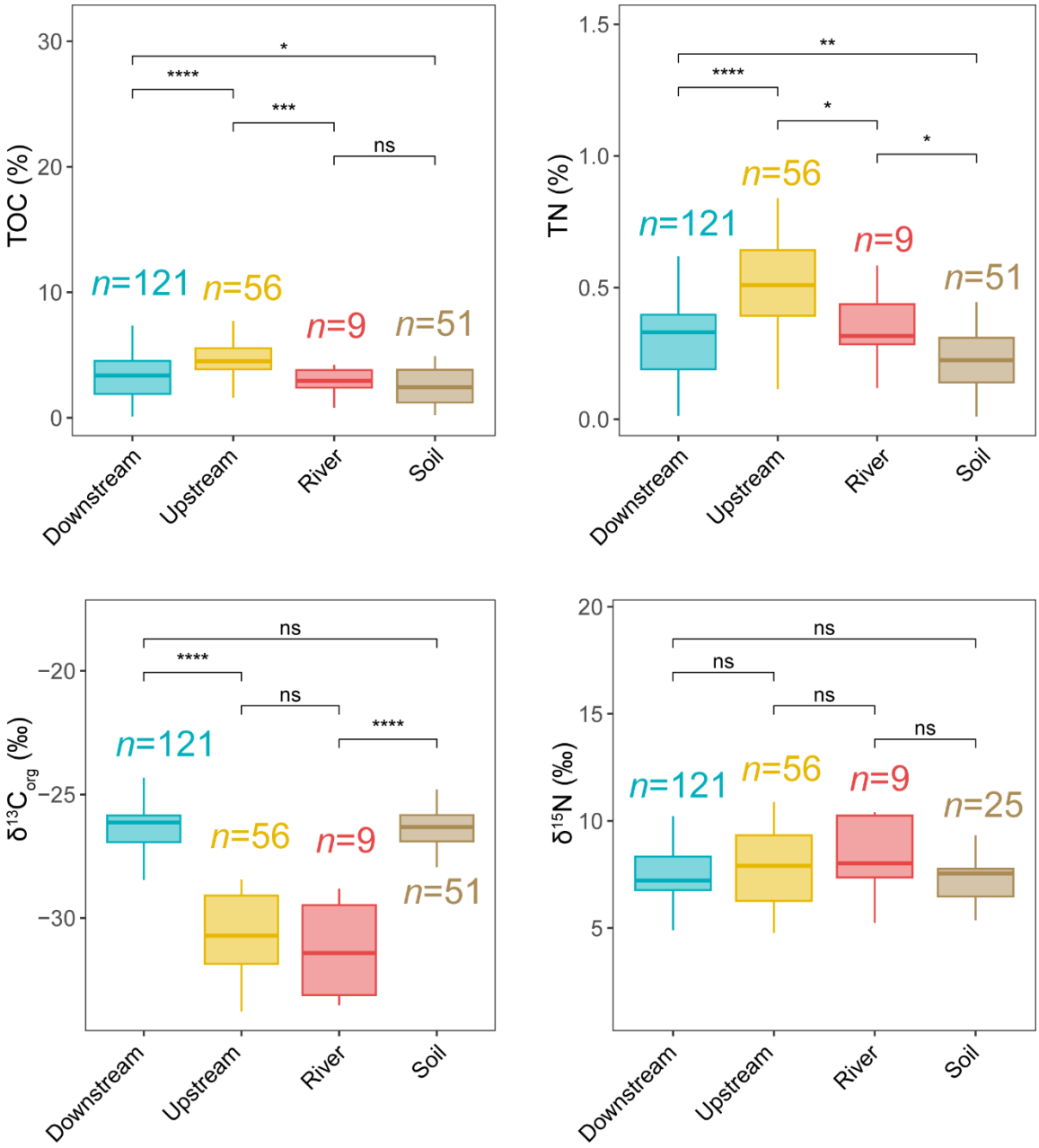



**Figure 3-3.** Distribution of bulk parameters (TOC, TN, $\delta^{13}C_{org}$ and $\delta^{15}N$) from soils (surficial soils and mudflat sediments) as well as river, upstream estuary and downstream estuary samples across the Seine River basin. Box plots of upstream and downstream estuary samples are based on SPM and sediments, whereas those of river samples are based only on SPM. Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

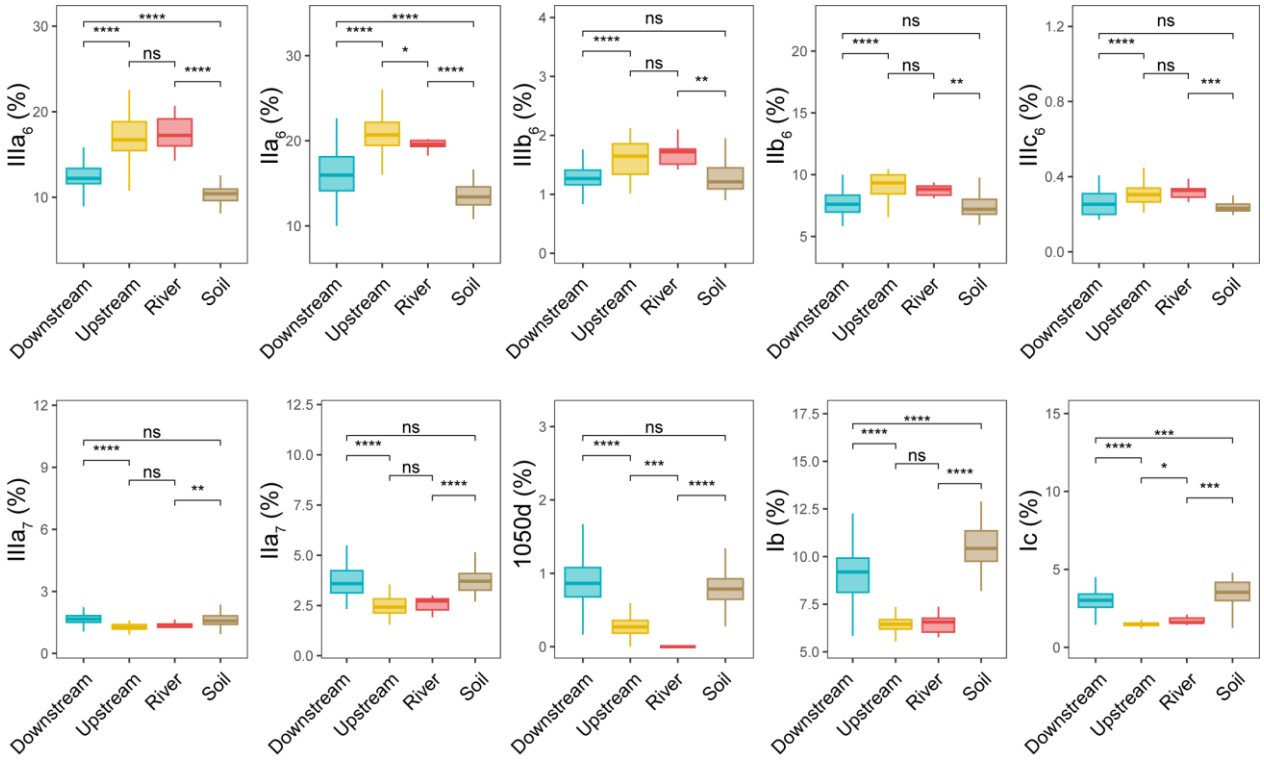



**Figure 3-4.** Relative abundances of selected individual brGDGTs from soils (surficial soils and mudflat soils/sediments, *n*=51), river (*n*=9), upstream estuary (*n*=56), and downstream estuary (*n*=121) samples across the Seine River basin: cyclopentane-containing tetramethylated brGDGTs (Ib and Ic), 6-methyl brGDGTs ($IIa_6$, $IIIa_6$, $IIb_6$, $IIIb_6$, and $IIIc_6$), 7-methyl brGDGTs ($IIa_7$ and $IIIa_7$) and brGDGTs 1050d. Box plots of upstream and downstream estuary samples are based on SPM and sediments, whereas those of river samples are based only on SPM. Boxes are color-coded based on the sample type (soil in brown, river in red, upstream estuary in yellow, and downstream estuary in blue). Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

The relative abundances of the brGDGTs were determined all along the Seine River basin (Figure 3-4 and Supplementary Figure 3-1, 3-2). The 6-methyl brGDGTs ($IIIa_6$, $IIa_6$, $IIIb_6$, $IIb_6$, and $IIIc_6$) were significantly higher in river (SPM) and upstream estuary (SPM and river channel sediments) than in soils (surficial soils and mudflat sediments) and downstream estuary (SPM and river channel sediments). In addition, the relative abundances of 7-methyl brGDGTs ($IIIa_7$ and $IIa_7$) and their late-eluting compound (1050d) in downstream estuary (SPM and river channel sediments) and soils (surficial soils and mudflat sediments) were significantly higher than those in river (SPM) and the upstream estuary (SPM and river channel sediments).

The concentration of total brGDGTs also showed differences along the land to sea continuum (Supplementary Figure 3-3a). The total brGDGTs concentration decreased from river (10.51 ± 5.91 μg/g organic carbon ($C_{org}$), based on SPM samples) to upstream estuary (7.52 ± 5.09 μg/g $C_{org}$, based on SPM and sediments) and downstream estuary (4.95 ± 4.09 μg/g $C_{org}$, based SPM and sediments). In soils from all the Seine basin, the concentration in total brGDGTs (1.55 ± 1.61 μg/g $C_{org}$, based on surficial soils and mudflat sediments) was significantly lower than that in SPM and sediments (Supplementary Figure 3-3a).

A Principal Component Analysis (PCA) was performed to statistically compare the fractional abundances of brGDGTs from different location (river, upstream and downstream estuary, based on SPM and sediments collected in the river channel), which explained 40.9% of the variance in two dimensions, with negative loadings for most of the 6-methyl brGDGTs and positive loadings for the remaining brGDGTs (Figure 3-5a). Samples from the downstream estuary clustered well apart from those from the river and upstream parts. Specifically, the brGDGT distribution was dominated by 6-methyl brGDGTs ($IIIa_6$, $IIIb_6$, $IIIc_6$, $IIa_6$, and $IIb_6$) in river and upstream estuarine samples, whereas in downstream estuary, it was driven by 5-methyl brGDGTs ($III_5$, $IIa_5$, $IIc_5$, $IIb_5$ and $IIIb_5$), tetramethylated brGDGTs (Ia, Ib, and Ic), 7-methyl brGDGTs ($IIIa_7$, $IIa_7$, and $IIb_7$), and their late-eluting compounds (1050d and 1036d). The brGDGT distributions of soils (surficial soils and mudflat sediments) were included in the PCA biplot performed on SPM and river channel sediments. This revealed that the brGDGT distribution in soils mostly overlap with the one in downstream SPM and river channel sediments (Figure 3-5a).

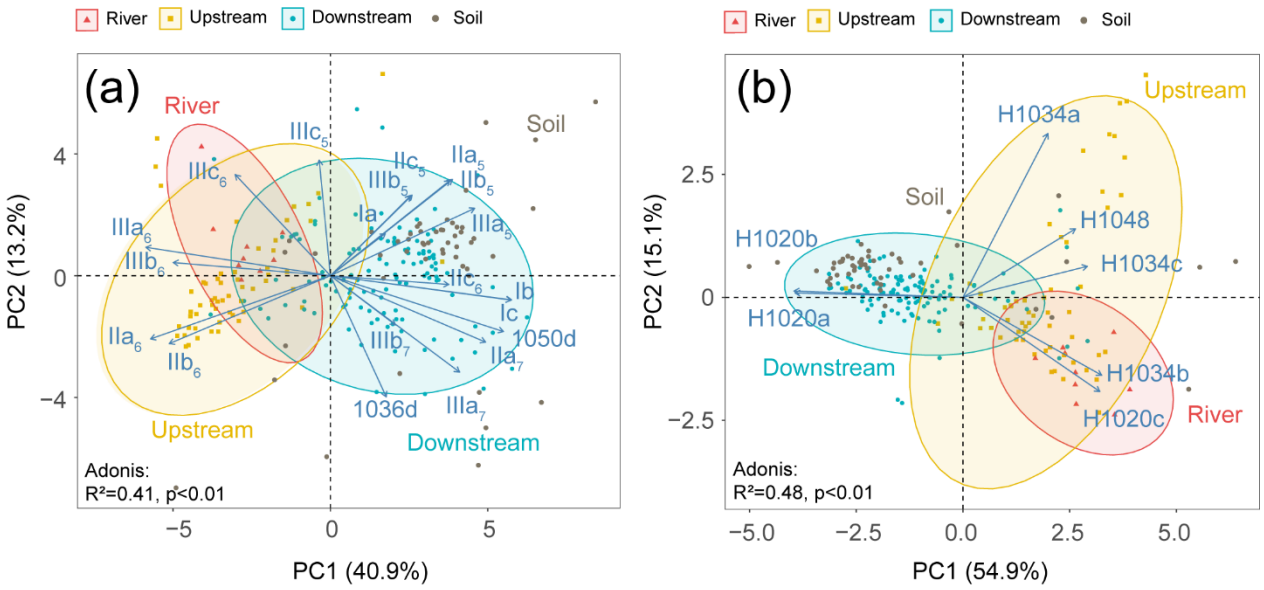



**Figure 3-5.** PCA analysis of fractional abundances of (a) brGDGTs and (b) brGMGTs. PCA scores of soils were added passively as an overlay. Their coordinates are predicted based on the information provided by the PCA performed on SPM and sediments (active individuals). Adonis analysis was used to evaluate how variation can be explained by the variables (999 permutations).

A Redundancy analysis (RDA) was performed to investigate the influence of the environmental factors (TOC, TN, temperature, water discharge and salinity) on the brGDGT distributions in SPM samples (Figure 3-6a and Supplementary Table 3-2). It allowed to explain 39.79% of the variability through two dimensions. The RDA triplot (Figure 3-6) showed how these factors correlate to the distributions of individual brGDGTs. The first axis of the RDA explained 33.16% of the variability and was primarily correlated with salinity and TN, whereas the second axis explained 6.63% of the variability and was associated with temperature, water discharge and TOC (Figure 3-6a and Supplementary Table 3-2). Based on hierarchical partitioning, salinity and TN were the two most important variables in explaining the brGDGT variations (individual importance of 14.97 % for salinity and 13.47 % for TN; Figure 3-6b and Supplementary Table 3-2). Compared with the salinity and TN, other available parameters have much lower individual importance (3.68 % for water discharge, 3.6 % for temperature and 2.12 % for TOC; Figure 3-6b and Supplementary Table 3-2).

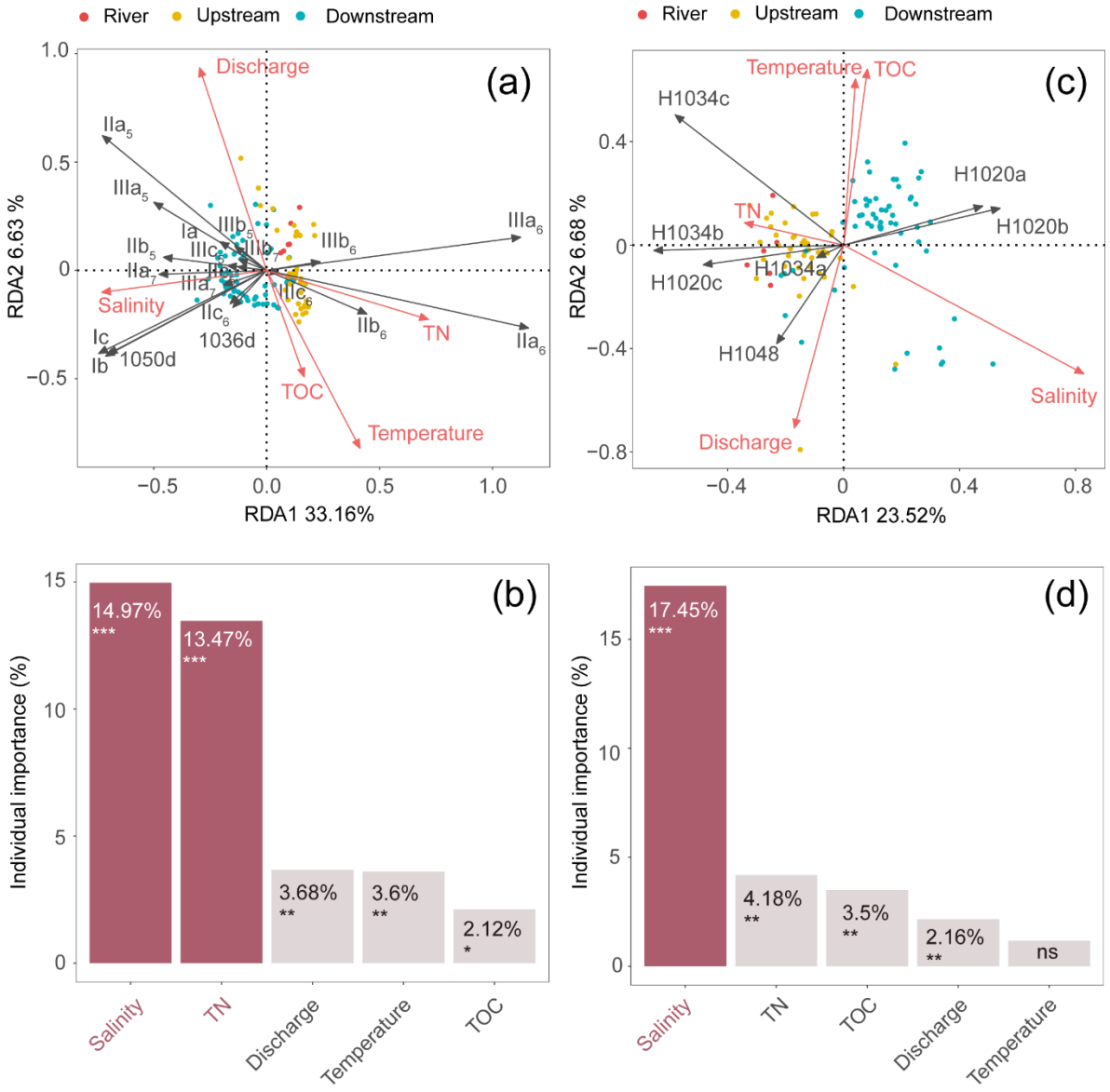



**Figure 3-6.** RDA analysis showing relationships between environmental factors (TN, TOC, salinity, temperature, discharge, red arrows) and fractional abundances of (a) brGDGTs and (c) brGMGTs. The individual importance of the environmental factors (TN, TOC, salinity, temperature, and discharge) explaining the variation in (b) brGDGT and (d) brGMGT distributions was determined by hierarchical partitioning analysis. The dataset used for RDA analysis is composed of SPM from river ($n$=6; red), upstream estuary ($n$=42; yellow) and downstream estuary ($n$=59; blue). Significance level is indicated by asterisks: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns, not significant, $p > 0.05$. $p$-values are derived from permutation tests (999 randomizations).

### 3.3.3 Distribution of brGMGTs from land to sea

The brGMGTs identified in previous studies were detected in the samples collected across the Seine River basin (Figure 3-2). H1034a is the least abundant isomer and is below detection limit for most of the samples in the Seine River basin (Supplementary Figure 3-4). The chromatograms revealed distinct distributions in brGMGTs in the different parts of the basin (SPM and sediments), with e.g. a higher intensity for the homologue H1020c in the river samples (SPM) than in those from the upstream (SPM) and downstream estuary (SPM) (Figure 3-2). These spatial variations were apparent when calculating the fractional abundances of the individual brGMGTs (Figure 3-7 and Supplementary Figure 3-4, 3-5). From upstream to downstream, the relative abundances in H1020a and H1020b increased, whereas those in 1020c and H1034b decreased (Figure 3-7). In SPM and river channel sediments, the total brGMGT concentration was observed to be slightly higher in the riverine part (0.26 ± 0.24 μg/g $C_{org}$) than in downstream (0.20 ± 0.13 μg/g $C_{org}$) and upstream estuary samples (0.17 ± 0.18 μg/g $C_{org}$; Supplementary Figure 3-3b). The total brGMGT concentrations were the lowest in soils (surficial soils and mudflat sediments) all over the basin (0.07 ± 0.09 μg/g $C_{org}$; Supplementary Figure 3-3b).

The PCA analysis based on the brGMGT relative abundances (Figure 3-5b) explained 70 % of the variance, which allows to observe that samples from the different parts of the basin clustered well apart from each other. The first axis explained 54.9 % of the variance, separating downstream samples from riverine and upstream samples, with negative loadings for two brGMGTs (H1020a and H1020b), and positive loadings for the remaining brGMGTs (H1020c, H1034a, H1034b, H1034c, and H1048). The second axis explained 15.1% of the variance and mainly separated the riverine and upstream samples, with higher relative abundances of compounds H1020c and H1034b in riverine samples (Figure 3-5b). The soil brGMGT distributions were

passively added to the PCA biplot based on SPM and sediments, revealing that the soils largely overlap with the SPM and sediments collected in the downstream estuary (Figure 3-5b).

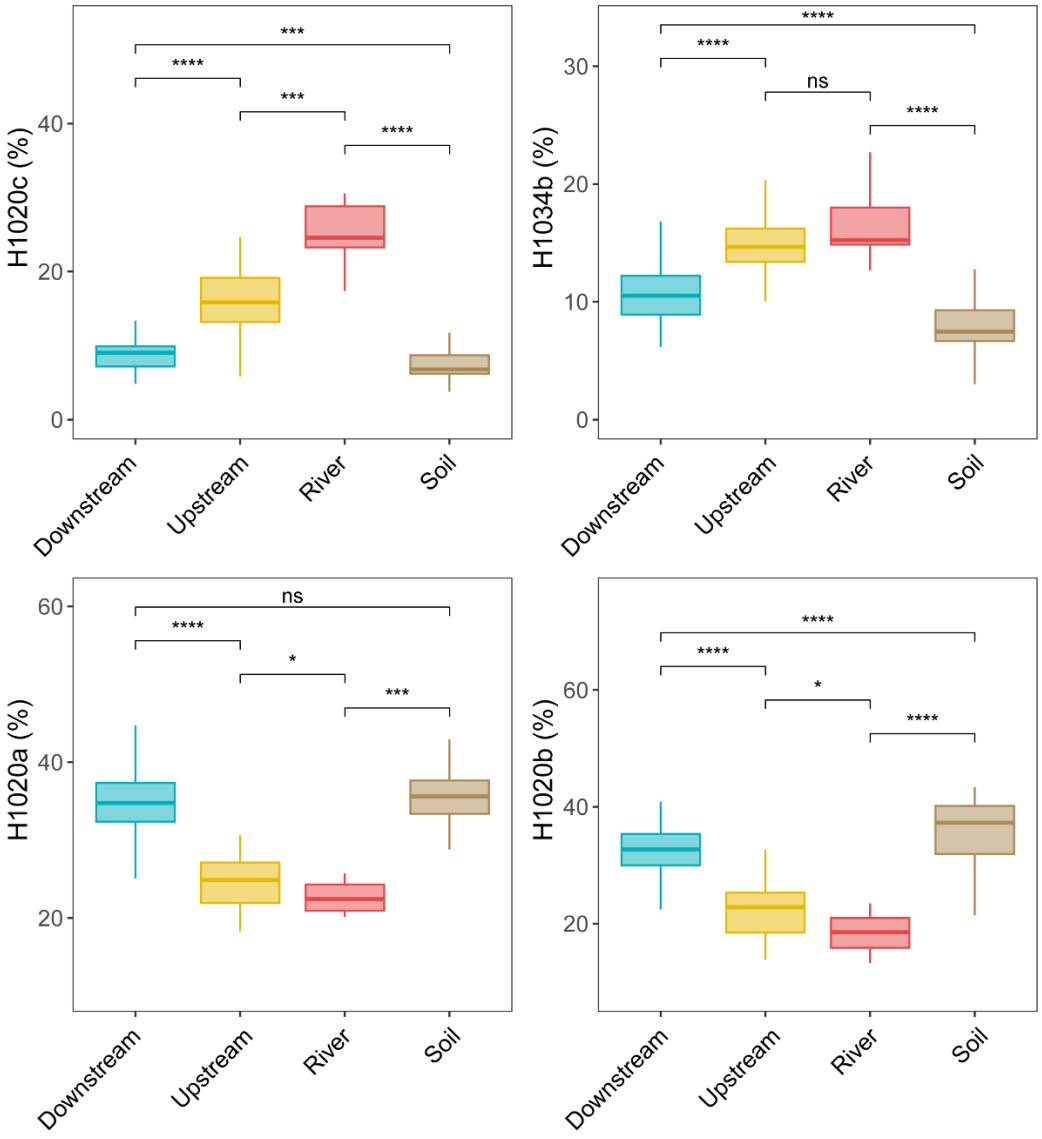



**Figure 3-7.** Relative abundance of selected individual brGMGTs from soils (surficial soils and mudflat soils/sediments, *n*=51), river (*n*=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) across the Seine River basin. Box plots of upstream and downstream estuary are composed of SPM and river channel sediments, whereas those of river are composed of SPM. Boxes are color-coded based on the sample type (soil in brown, river in red, upstream estuary in yellow, and downstream estuary in blue). Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

The RDA was performed to investigate the factors that could explain the variability of brGMGT distributions in SPM samples (Figure 3-6c and Supplementary Table 3-2) and allows to

explain 30.2 % of the variance. The RDA triplot showed that the first axis, accounting for 23.52 % of the variability, was mainly associated with salinity and to a lesser extent TN, while the second axis (6.68 %) was mainly driven by temperature, TOC and water discharge (Figure 3-5c and Supplementary Table 3-2). Based on hierarchical partitioning, salinity had the highest individual importance (17.45 %) in explaining the variability of brGMGT distribution followed by TN (4.18 %), TOC (3.5 %), and water discharge (2.16 %) (Figure 3-6d and Supplementary Table 3-2).

## 3.4. Discussion

### 3.4.1. Sources of brGDGTs and environmental controls on their distribution

#### 3.4.1.1 Sources of brGDGTs

In order to determine the predominant origin of brGDGTs in the Seine River basin, the overall brGDGT concentrations and distributions in SPM and river channel sediments ($n$=186) were compared with those in soils (surficial soils and mudflat sediments, $n$=51). The brGDGT concentrations (normalized to $C_{org}$) and relative abundances of several brGDGTs (i.e. $IIIa_6$, $IIa_6$, $IIIb_6$, $IIb_6$, and $IIIc_6$) in the SPM and sediments were significantly higher than those in soils ($p$<0.05, Wilcoxon test; Supplementary Figure 3-3a and Figure 3-4). Such differences in brGDGT concentrations and relative abundances between soils and aquatic settings (SPM and sediments) imply that at least part of the brGDGTs in the water column and sediments of the Seine River basin is produced *in situ*. This is in agreement with previous findings which suggested an *in situ* aquatic contribution to the brGDGT pool (Crampton-Flood et al., 2021; De Jonge et al., 2015; Kirkels et al., 2022b; Peterse et al., 2009).

More specifically, the fractional abundances of the two major 6-methyl brGDGTs ($IIa_6$ and $IIIa_6$) are significantly higher in the Seine River and upstream estuary than in soils (Figure 3-4). This confirms that these brGDGTs are mostly produced within the river, adding to the growing body of evidence supporting riverine 6-methyl brGDGT production in water column and/or sediment (De Jonge et al., 2015; Bertassoli et al., 2022; Kirkels et al., 2022b). A subsequent shift in the brGDGT distributions in the downstream compared to the upstream areas is observed in the Seine River basin. The PCA analysis shows a separation of downstream estuarine samples (influenced by seawater intrusion) from riverine and upstream estuary ones (without significant seawater intrusion) (Figure 3-5a). This difference is predominantly driven by the higher abundances of 6-methyl brGDGTs in riverine and upstream estuarine samples *vs.* higher abundances of 5- and 7-methyl brGDGTs as well as compounds Ib, Ic, and late eluting brGDGTs 1050d, 1036d in downstream estuarine samples (Figures 3-4, 3-5a and Supplementary Figure 3-2). It may reflect the fact that riverine 6-methyl brGDGTs are more easily degraded than soil-derived homologues and only partially transferred downstream. In addition to that, the riverine brGDGT signal may be diluted by brGDGTs from other sources during downstream transport. The first hypothesis is based on a previous study, which showed a shift in brGDGT distribution from the Yenisei River to the Kara Sea (De Jonge et al., 2015). They interpreted this to be a preferential degradation of labile (riverine) 6-methyl brGDGTs and the enrichment in less labile (soil-derived) 5-methyl brGDGTs during transport (De Jonge et al., 2015). This suggests that only limited amounts of riverine 6-methyl brGDGTs are transferred to the ocean, as also shown in other recent studies (Cao et al., 2022; Kirkels et al., 2022b). Such preferential degradation of 6-methyl brGDGTs over other brGDGTs could be attributed to variations in how these molecules are attached to soil particles (Huguet et al., 2008). In addition, a shift in brGDGT distribution during downstream transport could be explained by mixing with autochthonous (i.e. estuarine-produced)

brGDGTs (Crampton-Flood et al., 2021). The relative abundance of several brGDGTs (i.e. Ib, Ic, $IIIa_7$, $IIa_7$ and 1050d) in the downstream part of the Seine River basin is indeed significantly higher than the one in the upstream part ($p<0.05$, Wilcoxon test; Figure 3-4), suggesting *in situ* brGDGT production in saltwater. Such a saltwater contribution can be visualized by the PCA based on brGDGT distribution, showing the positive score of the aforementioned compounds with the first axis (Figure 3-5a). This axis is dominated by downstream samples influenced by seawater intrusion in the Seine Estuary (Figure 3-5a). It should be noted that brGDGT distributions in soils were roughly similar to those observed in downstream samples (SPM and river channel sediments; Figure 3-5a). Hence, it cannot be excluded that brGDGTs detected in downstream samples are at least partly derived from soils of the watershed. Nevertheless, the soil-derived brGDGT contribution to the downstream samples is expected to be much lower than the autochthonous one, as the average brGDGT concentration in soils was ca. 3 times lower than the one in downstream (i.e. SPM and river channel sediment) samples (Supplementary Figure 3-3a).

#### 3.4.1.2. Environmental controls on the brGDGT distribution

As several individual brGDGTs are suggested to be preferentially produced either in the riverine or estuarine parts of the Seine basin, their distribution might be related to ambient environmental factors. The RDA (performed on SPM samples) highlights the relationships between the available environmental variables (salinity, TN, TOC, water discharge and temperature) and the relative abundances of brGDGTs. Hierarchical partitioning indicates that salinity is the most important factor influencing the brGDGT distribution (14.97 %) in the Seine River basin (Figure 3-6b and Supplementary Table 3-2). Salinity is related to the relative abundances of compounds Ib, Ic, 7-methyl brGDGTs and the late-eluting homologs 1050d and

1036d that scored negatively on the first axis of the RDA (Figure 3-6a). This is in line with the positive significant correlations between salinity and the relative abundances of these compounds (Supplementary Figure 3-6). This trends also support the assumption made about the aquatic production of ring-containing tetramethylated brGDGTs (Ib and Ic) in Svalbard fjords which was thought to be linked to a salinity change (Dearing Crampton-Flood et al., 2019). The 7-methyl brGDGTs and their late-eluting isomers were also shown to be much more abundant in hypersaline lakes than those of lower salinity (Wang et al., 2021). Such a salinity-dependent brGDGT composition has previously been interpreted by membrane adaptation to salinity changes or by a shift in bacterial community composition (Dearing Crampton-Flood et al., 2019; Wang et al., 2021). Hence, the significant positive correlations between salinity and these compounds in the Seine River basin suggest that brGDGT-producing bacteria have similar physiological mechanisms (i.e., membrane adaptation) to those reported in other aquatic settings (lakes and fjords) and/or that the diversity of these bacteria is changing along the river-sea continuum.

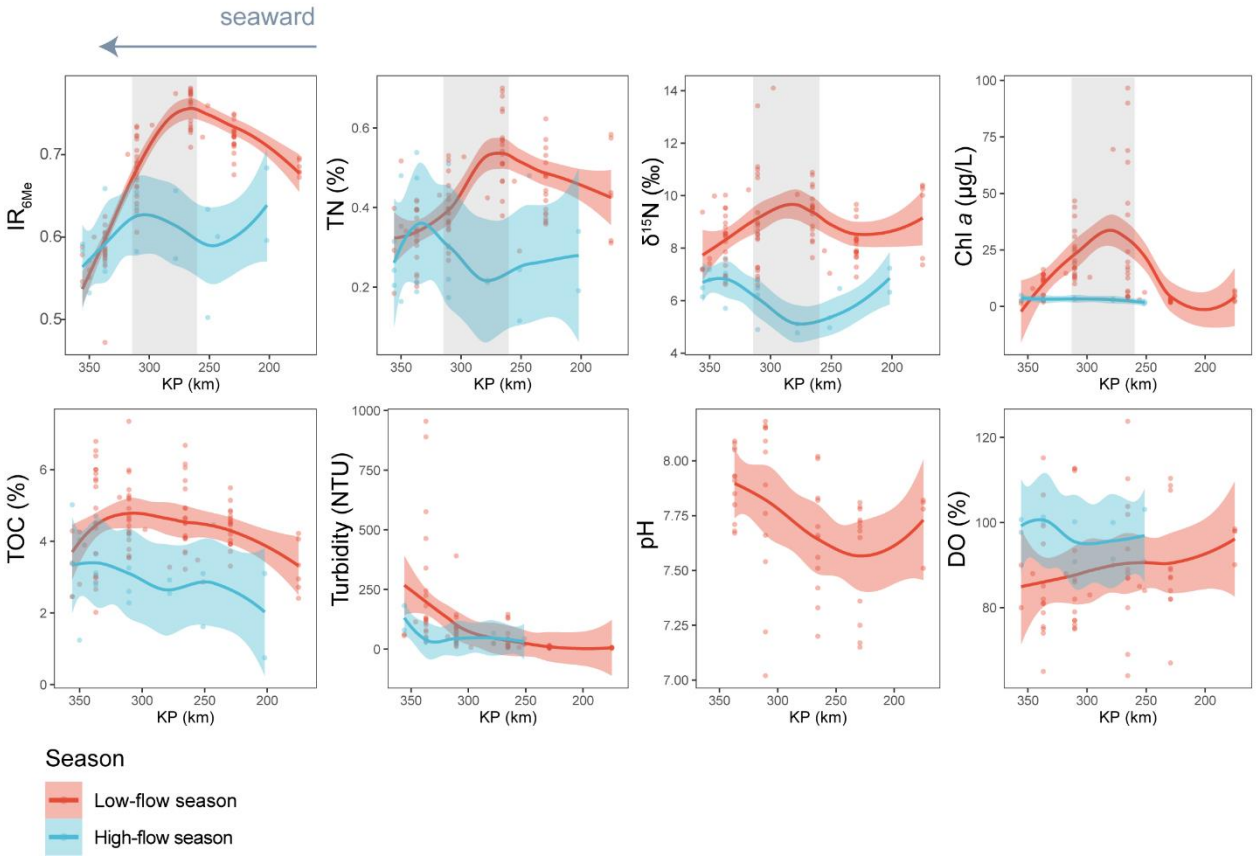



**Figure 3-8.** Spatio-temporal variations of $IR_{6Me}$ and several environmental factors, including TN (%), $\delta^{15}N$ (‰), Chla (µg/L), TOC (%), turbidity (NTU) pH, and dissolved oxygen saturation (DO, %). The trends showing variations were based on locally estimated scatterplot smoothing (LOESS) method with 95% confidence intervals. KP (kilometric point) represents the distance in kilometers from the city of Paris (KP 0). Dataset is composed of SPM. The shaded area highlights a zone (260 < KP < 340) where $IR_{6Me}$ and several environmental parameters co-vary.

The relative abundances of several 6-methyl brGDGTs (i.e. $IIa_6$, $IIIa_6$, and $IIb_6$) in the Seine River basin reveal significant negative correlations with salinity ($p<0.05$, Wilcoxon test; Supplementary Figure 3-6), which is in contrast with the positive relationships previously found in lakes (Wang et al., 2021). The distinct behavior of 6-methyl brGDGTs between lakes and the Seine river-sea continuum might be due to the lower salinity range in the Seine River basin (0-32 psu) *vs.* the lakes (0-376 psu) investigated by Wang et al. (2021). This suggests that the limited range

of salinity variation in the Seine River basin might be insufficient to trigger significant 6-methyl brGDGT production as observed in hypersaline lakes.

Alternatively, the significant negative correlations between the salinity and the relative abundance of 6-methyl brGDGTs in the Seine basin suggest that the bacteria producing 6-methyl brGDGTs are preferentially present in the low salinity area of the estuary. To explore this hypothesis, we investigate the spatio-temporal variations of the 6-methyl *vs*. 5-methyl brGDGTs ratio: $IR_{6Me}$ (Figure 3-8). High $IR_{6Me}$ values (0.69 ± 0.10) are associated with enhanced *in situ* production of 6-methyl brGDGTs within the Yenisei river (De Jonge et al., 2015). In the Seine River basin, seasonal variation in $IR_{6me}$ is observed. Specifically, much higher $IR_{6Me}$ values are observed in a specific zone of the estuary (260 < KP < 340) with a low salinity range (1.18 ± 2.71 psu) during low-flow season (Figure 3-8), suggesting that 6-methyl brGDGTs are preferentially produced in this zone when water discharge is low. Similarly, preferential production of 6-methyl brGDGT at low discharges was previously observed in other river systems, including the Amazon River basin (Kirkels et al., 2020; Crampton-Flood et al., 2021; Bertassoli et al., 2022) as well as Black and White Rivers (Dai et al., 2019). It was suggested that the enhanced 6-methyl brGDGT production at low flows was due to slow flow velocity and reduced soil mobilization. Although these hypotheses could account for the temporal variation in $IR_{6Me}$ in the Seine River basin, they are unlikely to explain the substantially high $IR_{6Me}$ values in this specific zone. Other environmental variables such as dissolved oxygen contents (Wu et al., 2021) and pH (De Jonge et al., 2014, 2015) were previously suggested to have a potential influence on 6-methyl brGDGT distributions. Nevertheless, these two environmental factors do not co-vary with $IR_{6Me}$ in the present study and can be ruled out as causes of variation in 6-methyl brGDGT distribution along the Seine river-sea continuum (Figure 3-8). Hence, the production of 6-methyl brGDGTs in this zone of the Seine Estuary has to be triggered by other factors, such as the nutrient concentration.

High nutrient levels were shown to favor the production of 6-methyl versus 5-methyl brGDGTs in the water column of mesocosm experiments (Martínez-Sosa and Tierney, 2019). As the nutrient concentration is higher in the upstream part of the Seine estuary (Wei et al., 2022), the substantial 6-methyl brGDGT production observed in the aforementioned zone (260 < KP < 340, Figure 3-8) at low flows could be due to the high amount of nutrients, especially nitrogen. This is supported by the RDA triplot showing strong correlation of TN with the brGDGT distribution in the Seine basin, with the major 6-methyl brGDGTs (i.e. $IIa_6$ and $IIIa_6$) plotting close to TN in the RDA triplot (Figure 3-6a). In addition, TN and $\delta^{15}N$ are observed to co-vary with $IR_{6Me}$ and to peak in the same zone (260 < KP < 340; Figure 3-8) during the low-flow season. Nitrate from sewage effluents and manure are generally enriched in $^{15}N$ compared to other sources, leading to much elevated $\delta^{15}N$ values (10–25‰) (Andrisoa et al., 2019; Leavitt et al., 2006). Nutrients, in the form of nitrogen, can be concentrated at low discharges, thus triggering phytoplankton blooms (Romero et al., 2019). Hence, the elevated TN and $\delta^{15}N$ signals in a specific zone of the estuary (260 < KP < 340) could be attributed to the increase of nitrogen loadings and $^{15}N$-enriched nitrate uptake by phytoplankton developing intensively during the low-flow season. The much higher chlorophyll *a* concentrations in this zone under low discharge conditions support the hypothesis of phytoplankton blooms (Figure 3-8). This high phytoplankton biomass might consequently create an environment that accelerates the growth and production of heterotrophic bacteria, which can in turn transform phytoplankton-derived organic matter (Buchan et al., 2014). As the brGDGT-producers were suggested to have a heterotrophic lifestyle (Weijers et al., 2010; Huguet et al., 2017; Blewett et al., 2022), they may transform phytoplankton-derived organic matter and thus participate in N-cycling during blooms. Hence, the co-variations of all the parameters ($IR_{6Me}$, TN, $\delta^{15}N$, and Chl *a* concentration) peaking in the low salinity area during low-flow season suggest that low salinity

range and high phytoplankton productivity represent favorable conditions for 6-methyl brGDGT production.

### 3.4.2. Sources of brGMGTs and environmental controls on their distribution

#### 3.4.2.1 Sources of brGMGTs

Similarly to the brGDGTs, the brGMGTs can also be produced *in situ* within the water column and/or sediments (Baxter et al., 2021; Kirkels et al., 2022a). In previous studies, brGMGTs were detected only in part of the soils surrounding the Godavari River basin (India; Kirkels et al., 2022a) and Lake Chala (East Africa; Baxter et al., 2021), suggesting a limited brGMGT production in soils in comparison to aquatic settings. Consistently, in the Seine River basin, concentrations of brGMGTs in SPM and sediment samples are significantly higher than those in soils ($p<0.05$, Wilcoxon test; Supplementary Figure 3-3b), pointing out their predominant aquatic source.

A notable compositional shift in brGMGT distribution is observed along the Seine River basin, as revealed by the separation of riverine, upstream and downstream estuarine samples in the PCA (Figure 3-5b). The relative abundance of 2 brGMGTs (H1020c and H1034b) gradually decreases across the basin (Figure 3-7) and is significantly correlated with those of 6-methyl brGDGTs (Supplementary Figure 3-7). As 6-methyl brGDGTs are mainly produced in freshwaters in the Seine basin, this suggests that brGMGTs H1020c and H1034b and 6-methyl brGDGTs have a common freshwater origin and that the mixture of fresh and marine waters along the estuary leads to the dilution of these compounds during downstream transport. H1020c is the dominant brGMGT homologue in SPM from the riverine zone of the Seine and one of the most abundant brGMGT in the upstream part of the estuary (Figure 3-7). Such a trend was also observed in SPM and riverbed sediments from the upper part of the Godavari River basin, which was attributed to *in situ* riverine brGMGT production of this compound (Kirkels et al., 2022a).

The fractional abundance of H1020a and H1020b homologues gradually increases along the Seine River basin. This is consistent with the higher abundances of H1020a and H1020b previously reported in marine sediments from the Bay of Bengal (Kirkels et al., 2022a). The predominance of these compounds in such samples was attributed to their *in situ* production in the marine realm. In line with this hypothesis, the relative abundances of brGMGTs H1020a and H1020b positively correlate with brGDGTs Ib, Ic, $IIIa_7$, $IIa_7$ and 1050d (Supplementary Figure 3-7) in the Seine Estuary, suggesting a similar marine origin.

#### 3.4.2.2. Environmental controls on the distribution of brGMGTs

The current knowledge on the parameters controlling the brGMGT distributions in the terrestrial and marine realm is still limited. The correlations between the brGDGT and brGMGT relative abundances in the Seine River basin (Supplementary Figure 3-7) suggest that both types of compounds might be derived from overlapping source microorganisms, with common environmental factors controlling their membrane lipid composition. In the Seine River basin, salinity is shown to be the main environmental parameter influencing the brGMGT distribution, as also observed for brGDGTs (Figure 3-6). This is reflected in the significant ($p$<0.05) increase in the relative abundances of homologues H1020a and H1020b with salinity and a concomitant significant negative correlation between this parameter and the relative abundances of homologues H1020c and H1034b ($p$<0.05, Wilcoxon test; Figures 3-9, a-d). Nevertheless, the individual effect of TN on brGMGT relative abundances is observed to be much lower compared to that observed for brGDGTs (Figure 3-6 and Supplementary Table 3-2). This implies that, while having common controlling factors such as the salinity, they are also influenced by distinct parameters (i.e. TN), likely indicating distinct sources. This is consistent with a recent study showing that brGDGTs and brGMGTs likely originate from overlapping, but not identical origins (Elling et al., 2023).

The shift in brGMGT distribution observed across the Seine River basin (Figure 3-6) could be due to a change in the diversity of brGMGT-producing bacteria and/or to an adaptation of these microorganisms to environmental changes occurring from upstream to downstream. The latter hypothesis seems unlikely, as a physiological adaptation of a given bacterial community would make it difficult to explain why the relative abundance of three isomers of compound H1020, which share a similar structure, varies differently in response to salinity changes. Hence, a shift in brGMGT-producing bacterial communities across the basin is more likely. Compounds H1020c and H1034b could predominantly be produced by bacteria preferentially growing in freshwater, and homologues H1020a and H1020b by bacteria preferentially living in brackish or saltwater.

### 3.4.3. Potential implications for brGMGTs as a proxy for riverine runoff

The distinct brGMGT distributions in freshwater and saltwater could be used to trace the Organic Matter (OM) produced upstream all along the Seine basin. To trace such a riverine runoff signal, we propose a new proxy, the Riverine IndeX (RIX), based on the fractional abundances of brGMGTs H1020c and H1034b versus H1020a and H1020b (Eq. 3):

$$\mathrm{RIX} = \frac{H1020c + H1034b}{H1020c + H1034b + H1020a + H1020b} \quad (3)$$

The RIX is calculated for the SPM and sediment samples from the Seine River basin, showing an obvious decreasing trend from upstream to downstream (Figure 3-9e). The RIX in river (0.51±0.06, SPM) and upstream estuarine (0.40±0.07, SPM and sediments) samples is significantly higher than for downstream estuarine (0.23±0.06, SPM and sediments) samples. RIX values around 0.50 could therefore be considered reflecting the riverine endmember, while those below 0.30 could represent the saltwater endmember.

A significant overlap between brGMGT distribution soils and downstream samples was observed (Figure 3-5b). This suggests that part of the brGMGT signal in the water masses of the

Seine may be partially derived from surrounding soils. Hence, RIX was also calculated for the soil samples. The RIX values of the soil samples were 0.21±0.13, close to those of the downstream estuarine samples. However, the average concentrations of brGMGTs are an order of magnitude lower in the soils than in the river channel sediments and SPM samples of the Seine basin (Supplementary Figure 3-3b). Therefore, it can be assumed that the impact of soil-derived brGMGTs on the observed RIX signal in the water column of the Seine basin is low.

In order to test the general applicability of the RIX, it was then applied to riverine and marine samples (SPM and sediments) collected in the Godavari River basin and Bay of Bengal (Kirkels et al., 2022a). This site represents the only other river-sea continuum besides the Seine basin for which brGMGT data are presently available. Significant differences in RIX between the SPM and sediment samples from the Godavari River basin are observed ($p$<0.05, Wilcoxon test; Figure 9f). In addition, 96% of the RIX values in riverine SPM and riverbed sediments from the Godavari basin exceed 0.5, whereas all of the RIX values observed in marine sediments from the Bay of Bengal are below 0.3. This suggests that the RIX cutoff values defined using the samples from the Seine basin may be broadly applicable and valid across other river-sea continuums. This deserves further studies.

Further confirmation of the RIX potential as a tracer of riverine OM comes from the significant correlations observed between this index and other commonly used proxies for tracing OM sources, i.e. the BIT and $\delta^{13}C_{org}$ ($p$<0.05, Wilcoxon test; Figure 3-9, g-h). These proxies show roughly similar spatial and temporal variations in the Seine River basin. In the low-flow season, RIX and BIT gradually decrease while $\delta^{13}C_{org}$ increase across the basin (Figure 3-9, i-k). Such trends during the low discharge periods likely reflects the continuous dilution process of riverine OM caused by the mixing of fresh and marine water masses (Thibault et al., 2019). The gradual dilution of the riverine OM signal along the Seine River basin could be due to the increase of

seawater intrusion, and thus marine-derived OM, at low discharges (Kolb et al., 2022; Ralston and Geyer, 2019). In contrast, during the high-flow season, no such gradual dilution trend is observed. Instead, at high discharges, the RIX, BIT and $\delta^{13}C_{org}$ remain roughly stable from KP 202 to 310.5, before, steeply decreasing for BIT and RIX, and increasing for $\delta^{13}C_{org}$. This trend can be explained by the fact that at high flow rates, the limit of saltwater intrusion in the estuary shifts seawards rather than landwards, allowing the riverine OM to be flushed further downstream than under low discharge conditions. After this region, the riverine OM is diluted because of the mixing with marine water masses, as observed during the low-flow season. The trends observed in the Seine Estuary are consistent with previous studies in other regions, showing that terrestrial OM was only effectively transported downstream at high flow rates (Kirkels et al., 2022b, 2020).

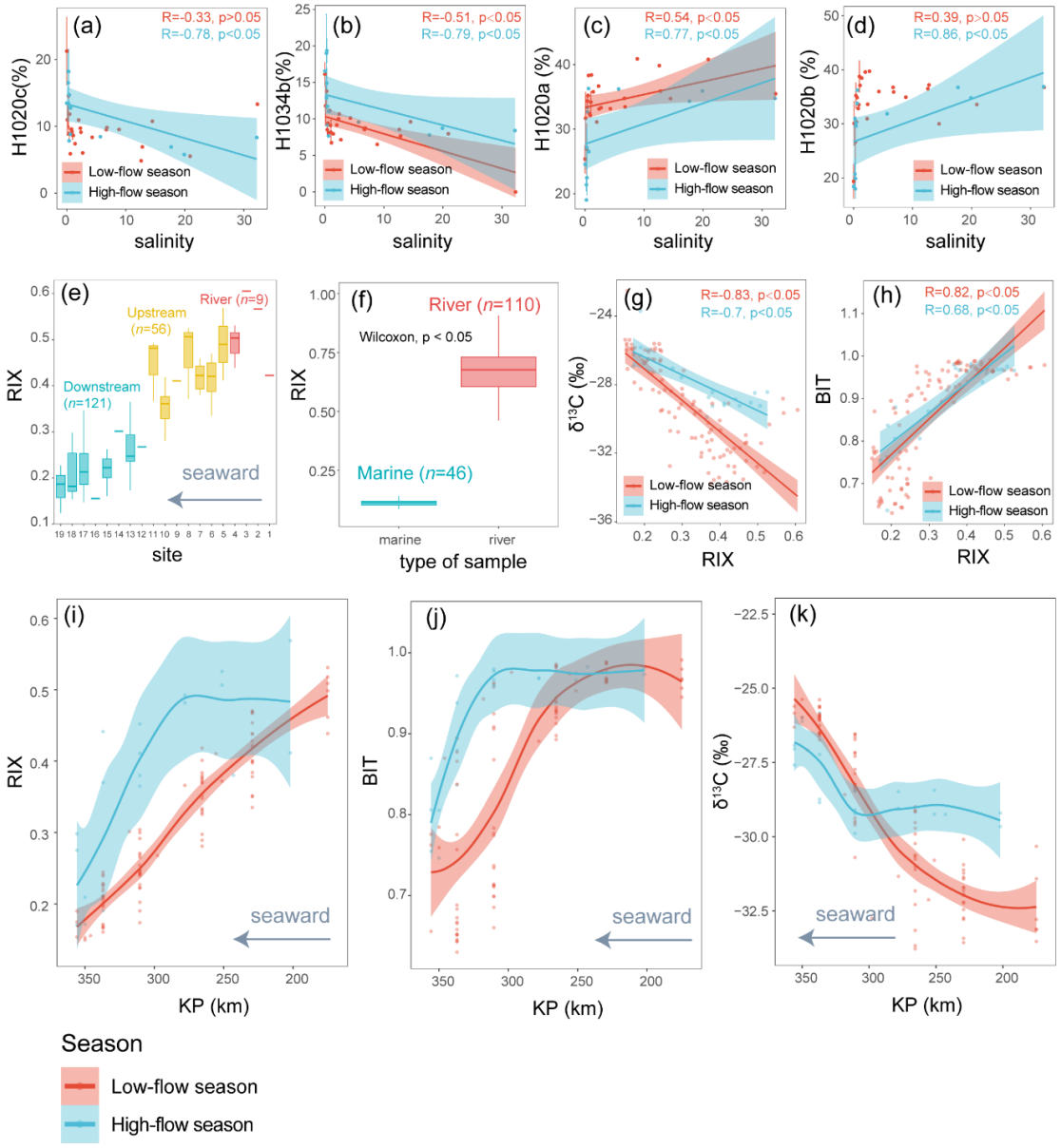



**Figure 3-9.** (a-d) Salinity plotted versus relative abundance of brGMGTs. Shaded area represents 95% confidence intervals. Vertical error bars indicate mean ± s.d for samples with the same salinity. Dataset is composed of SPM. (e) Distribution of RIX across the Seine River basin. Boxes are color-coded based on the sample type (river in red, upstream estuary in yellow, and downstream estuary in blue). Dataset is composed of SPM and river channel sediments. (f) RIX in the Godavari River basin (India) and Bay of Bengal sediments (data from Kirkels et al. (2022a)). Statistical testing was performed by a Wilcoxon test. (g-h) RIX plotted versus $\delta^{13}C$ and BIT. Shaded area represents 95% confidence intervals. (i-k) Spatio-temporal variations of RIX and several other terrestrial proxies, including BIT and $\delta^{13}C$ (‰). The trends showing spatio-temporal variations were based on locally estimated scatterplot smoothing (LOESS) method with 95% confidence intervals. KP (kilometric point) represents the distance in kilometers from the city of Paris (KP 0). Dataset is composed of SPM

Although the BIT is successfully used in the Seine River basin as well as in previous studies to trace riverine (terrestrial) OM inputs (Hopmans et al., 2004; Xu et al., 2020), this index can be biased by *in situ* production of brGDGTs in the water column and/or sediments (Dearing Crampton-Flood et al., 2019; Sinninghe Damsté, 2016) and selective degradation of crenarchaeol *vs.* brGDGTs (Smith et al., 2012). Hence, high BIT values do not necessarily indicate higher contribution of terrestrial OM in some settings (Smith et al., 2012). Unlike the BIT index, based on two different families of compounds (isoGDGTs and brGDGTs), the RIX is based on 4 compounds from the same family (brGMGTs) that likely have similar degradation rates and therefore not influenced by selective degradation. Furthermore, the RIX is based on the relative abundances of abundant brGMGTs which are all predominantly produced in aquatic settings, two of them (H1020c and H1034b) being mainly produced in freshwater and two of them (H1020a and H1020b) mainly in saltwater. Therefore, the RIX is based on compounds which are more specifically produced in the two endmembers (freshwater or saltwater), which could avoid the biases encountered with the BIT. Overall, our work shows that, in addition to the BIT and $\delta^{13}C_{org}$, the RIX successfully captures the spatio-temporal dynamics of riverine OM in the Seine River basin, making this proxy a promising and complementary one tracing riverine runoff in modern samples as well as paleorecords. A potential application to paleorecords is still necessary to further test the applicability of the RIX as a riverine proxy.

## 3.5. Conclusions

In this study, the brGDGT and brGMGT concentrations and distributions in soils, SPM and sediments (*n*=237) across the Seine River basin were investigated. Higher concentrations and distinct distributions of brGDGTs and brGMGTs in SPM and sediments compared with soils imply that both types of compounds can be produced *in situ* in aquatic settings. The distribution of both brGDGTs and brGMGTs are largely related to salinity, but only brGDGT distributions are significantly influenced by nitrogen nutrient loadings. In addition, covariations of $IR_{6Me}$, TN, $\delta^{15}N$, and Chl *a* concentration within the low salinity region suggest that riverine (6-methyl) brGDGT production is favored by low-salinity and high-productivity conditions.

In the Seine River basin, salinity correlates positively with H1020a and H1020b, and negatively with H1020c and H1034b. This indicates that compounds H1020c and H1034b could be produced by bacteria that preferentially grow in freshwater, while homologues H1020a and H1020b could be produced by bacteria that mainly live in saltwater. Based on this, a novel proxy, the Riverine IndeX (RIX) is proposed to trace riverine OM input. The average value of RIX for the riverine samples is 0.51, which is much higher than that in downstream estuarine (0.23 on average) samples. This suggests that RIX values over 0.5 imply considerable riverine contributions, whereas RIX values below 0.3 indicate higher marine contributions. This cutoff value defined in the Seine River basin also works in the Godavari River basin (India), which implies that this novel proxy based on brGMGTs may be broadly applicable and warrants further exploration in present ecosystems as well as paleorecords.

## 3.6. Annexes

**Supplementary Table 3-1** bulk geochemical data as well as proxies derived from brGDGTs and brGMGTs

| Date | Site | Type | Season | TOC (%) | TN (%) | $\delta^{13}C$ (‰) | $\delta^{15}N$ (‰) | BIT | $IR_{6Me}$ | RIX |
|---|---|---|---|---|---|---|---|---|---|---|
| 2015-1 | site 15 | SPM | High-flow | 4.70 | 0.54 | -27.23 | 7.04 | 0.88 | 0.57 | 0.21 |
| 2015-4 | site 13 | SPM | High-flow | 2.28 | 0.30 | -29.89 | 6.47 | 0.98 | 0.64 | 0.36 |
| 2015-4 | site 15 | SPM | High-flow | 3.64 | 0.47 | -25.83 | 7.62 | 0.84 | 0.56 | 0.19 |
| 2015-4 | site 17 | SPM | High-flow | 1.24 | 0.16 | -26.12 | 7.18 | 0.75 | 0.53 | 0.17 |
| 2015-4 | site 18 | SPM | High-flow | 2.45 | 0.32 | -26.93 | 7.20 | 0.75 | 0.54 | 0.17 |
| 2015-7 | site 17 | SPM | Low-flow | 4.26 | 0.52 | -25.99 | 7.60 | 0.78 | 0.55 | 0.15 |
| 2015-7 | site 19 | SPM | Low-flow | 4.54 | 0.52 | -22.52 | 5.70 | 0.77 | 0.53 | 0.16 |
| 2015-9 | site 13 | SPM | Low-flow | 4.31 | 0.46 | -28.24 | 8.85 | 0.96 | 0.73 | 0.30 |
| 2015-9 | site 15 | SPM | Low-flow | 2.66 | 0.34 | -25.39 | 7.86 | 0.78 | 0.57 | 0.16 |
| 2015-9 | site 17 | SPM | Low-flow | 2.80 | 0.35 | -25.38 | 7.41 | 0.76 | 0.56 | 0.15 |
| 2015-9 | site 18 | SPM | Low-flow | 3.37 | 0.40 | -25.61 | 7.17 | 0.78 | 0.55 | 0.15 |
| 2016-4 | site 5 | SPM | High-flow | 3.10 | 0.34 | -29.20 | 6.32 | 0.98 | 0.60 | 0.41 |
| 2016-4 | site 7 | SPM | High-flow | 4.46 | 0.48 | -29.25 | 6.46 | 0.98 | 0.60 | 0.38 |
| 2016-4 | site 13 | SPM | High-flow | 4.18 | 0.51 | -29.10 | 6.21 | 0.98 | 0.63 | 0.41 |
| 2016-4 | site 17 | SPM | High-flow | 4.04 | 0.48 | -25.98 | 6.79 | 0.81 | 0.57 | 0.21 |
| 2016-4 | site 19 | SPM | High-flow | 2.09 | 0.30 | -23.70 | 6.60 | 0.64 | 0.53 | 0.19 |
| 2015-4 | site 15 | Sediment | n.a. | 5.13 | 0.62 | -25.88 | 6.91 | 0.82 | 0.66 | 0.33 |
| 2015-4 | site 15 | Sediment | n.a. | 4.82 | 0.56 | -26.13 | 6.58 | 0.77 | 0.66 | 0.24 |
| 2015-4 | site 15 | Sediment | n.a. | 4.96 | 0.57 | -26.09 | 6.75 | 0.78 | 0.63 | 0.20 |
| 2015-4 | site 15 | Sediment | n.a. | 4.50 | 0.51 | -26.08 | 6.57 | 0.80 | 0.65 | 0.26 |
| 2015-4 | site 15 | Sediment | n.a. | 3.06 | 0.33 | -26.57 | 6.63 | 0.81 | 0.63 | 0.26 |
| 2015-4 | site 15 | Sediment | n.a. | 1.72 | 0.19 | -27.07 | 7.18 | 0.77 | 0.57 | 0.22 |
| 2015-4 | site 15 | Sediment | n.a. | 5.22 | 0.61 | -25.99 | 6.91 | 0.78 | 0.64 | 0.19 |
| 2015-4 | site 15 | Sediment | n.a. | 4.07 | 0.46 | -26.19 | 6.39 | 0.80 | 0.61 | 0.25 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2015-4 | site 15 | Sediment | n.a. | 4.36 | 0.49 | -26.36 | 8.11 | 0.80 | 0.66 | 0.24 |
| 2015-4 | site 15 | Sediment | n.a. | 4.58 | 0.54 | -25.90 | 7.07 | 0.81 | 0.60 | 0.24 |
| 2015-4 | site 17 | Sediment | n.a. | 4.33 | 0.54 | -26.33 | 7.28 | 0.80 | 0.65 | 0.22 |
| 2015-4 | site 17 | Sediment | n.a. | 4.58 | 0.55 | -25.73 | 6.89 | 0.75 | 0.63 | 0.21 |
| 2015-4 | site 17 | Sediment | n.a. | 3.51 | 0.40 | -26.25 | 6.34 | 0.74 | 0.55 | 0.35 |
| 2015-4 | site 17 | Sediment | n.a. | 3.58 | 0.40 | -26.37 | 6.63 | 0.77 | 0.61 | 0.21 |
| 2015-4 | site 17 | Sediment | n.a. | 3.03 | 0.35 | -26.28 | 7.77 | 0.76 | 0.60 | 0.23 |
| 2015-4 | site 17 | Sediment | n.a. | 1.52 | 0.18 | -26.05 | 7.10 | 0.73 | 0.56 | 0.17 |
| 2015-4 | site 17 | Sediment | n.a. | 1.90 | 0.21 | -25.97 | 7.52 | 0.77 | 0.64 | 0.26 |
| 2015-4 | site 17 | Sediment | n.a. | 1.26 | 0.13 | -27.06 | 6.48 | 0.71 | 0.56 | 0.18 |
| 2015-4 | site 17 | Sediment | n.a. | 1.83 | 0.18 | -26.16 | 6.37 | 0.72 | 0.55 | 0.24 |
| 2015-4 | site 17 | Sediment | n.a. | 1.25 | 0.14 | -26.56 | 6.90 | 0.79 | 0.64 | 0.25 |
| 2015-9 | site 15 | Sediment | n.a. | 3.79 | 0.45 | -25.78 | 7.20 | 0.79 | 0.63 | 0.23 |
| 2015-9 | site 15 | Sediment | n.a. | 1.19 | 0.12 | -25.88 | 7.78 | 0.74 | 0.56 | 0.25 |
| 2015-9 | site 15 | Sediment | n.a. | 0.36 | 0.03 | -26.14 | 7.69 | 0.80 | 0.61 | 0.32 |
| 2015-9 | site 15 | Sediment | n.a. | 0.15 | 0.02 | -25.73 | 9.16 | 0.72 | 0.58 | 0.22 |
| 2015-9 | site 15 | Sediment | n.a. | 0.44 | 0.05 | -26.58 | 7.81 | 0.78 | 0.64 | 0.23 |
| 2015-9 | site 15 | Sediment | n.a. | 0.96 | 0.11 | -25.98 | 8.23 | 0.74 | 0.56 | 0.19 |
| 2015-9 | site 15 | Sediment | n.a. | 2.05 | 0.22 | -25.97 | 7.87 | 0.78 | 0.62 | 0.24 |
| 2015-9 | site 15 | Sediment | n.a. | 2.55 | 0.28 | -26.06 | 7.84 | 0.75 | 0.64 | 0.21 |
| 2015-9 | site 15 | Sediment | n.a. | 0.63 | 0.07 | -26.00 | 7.21 | 0.71 | 0.57 | 0.21 |
| 2015-9 | site 15 | Sediment | n.a. | 2.06 | 0.23 | -25.90 | 7.25 | 0.79 | 0.68 | 0.24 |
| 2015-9 | site 17 | Sediment | n.a. | 3.53 | 0.41 | -25.40 | 6.83 | 0.76 | 0.65 | 0.23 |
| 2015-9 | site 17 | Sediment | n.a. | 2.28 | 0.26 | -25.68 | 7.05 | 0.68 | 0.56 | 0.19 |
| 2015-9 | site 17 | Sediment | n.a. | 3.24 | 0.37 | -25.85 | 7.46 | 0.75 | 0.64 | 0.18 |
| 2015-9 | site 17 | Sediment | n.a. | 2.20 | 0.25 | -25.78 | 8.01 | 0.68 | 0.56 | 0.19 |
| 2015-9 | site 17 | Sediment | n.a. | 2.91 | 0.34 | -25.66 | 7.22 | 0.69 | 0.56 | 0.17 |
| 2015-9 | site 17 | Sediment | n.a. | 3.33 | 0.37 | -25.93 | 7.86 | 0.76 | 0.65 | 0.19 |
| 2015-9 | site 17 | Sediment | n.a. | 3.15 | 0.37 | -25.85 | 6.90 | 0.76 | 0.65 | 0.22 |
| 2015-9 | site 17 | Sediment | n.a. | 2.92 | 0.33 | -25.76 | 7.55 | 0.47 | 0.54 | 0.15 |
| 2015-9 | site 17 | Sediment | n.a. | 0.38 | 0.05 | -25.80 | 7.74 | 0.76 | 0.63 | 0.34 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2015-9 | site 17 | Sediment | n.a. | 1.24 | 0.13 | -25.81 | 7.44 | 0.85 | 0.50 | 0.27 |
| 2016-4 | site 7 | Sediment | n.a. | 7.26 | 0.82 | -28.60 | 6.11 | 0.98 | 0.69 | 0.42 |
| 2016-4 | site 7 | Sediment | n.a. | 6.81 | 0.74 | -28.53 | 5.86 | 0.50 | 0.55 | 0.16 |
| 2016-4 | site 7 | Sediment | n.a. | 6.50 | 0.72 | -28.51 | 5.61 | 0.97 | 0.67 | 0.45 |
| 2016-4 | site 7 | Sediment | n.a. | 6.22 | 0.70 | -28.44 | 5.81 | 0.97 | 0.65 | 0.40 |
| 2016-4 | site 7 | Sediment | n.a. | 7.71 | 0.84 | -28.57 | 5.69 | 0.98 | 0.68 | 0.44 |
| 2016-4 | site 7 | Sediment | n.a. | 6.50 | 0.69 | -28.61 | 5.49 | 0.98 | 0.67 | 0.42 |
| 2016-4 | site 7 | Sediment | n.a. | 6.93 | 0.76 | -28.45 | 5.39 | 0.98 | 0.69 | 0.46 |
| 2016-4 | site 7 | Sediment | n.a. | 7.10 | 0.79 | -28.67 | 5.92 | 0.97 | 0.64 | 0.44 |
| 2016-4 | site 7 | Sediment | n.a. | 6.92 | 0.77 | -28.51 | 5.63 | 0.97 | 0.65 | 0.42 |
| 2016-4 | site 7 | Sediment | n.a. | 5.98 | 0.67 | -28.55 | 5.28 | 0.98 | 0.63 | 0.38 |
| 2016-4 | site 17 | Sediment | n.a. | 1.34 | 0.14 | -26.50 | 7.00 | 0.86 | 0.60 | 0.23 |
| 2016-4 | site 17 | Sediment | n.a. | 1.04 | 0.10 | -26.30 | 6.50 | 0.78 | 0.57 | 0.21 |
| 2016-4 | site 17 | Sediment | n.a. | 1.32 | 0.14 | -26.87 | 6.98 | 0.86 | 0.62 | 0.29 |
| 2016-4 | site 17 | Sediment | n.a. | 0.99 | 0.10 | -26.20 | 6.40 | 0.80 | 0.54 | 0.26 |
| 2016-4 | site 17 | Sediment | n.a. | 0.11 | 0.01 | -26.30 | 9.50 | 0.75 | 0.56 | 0.23 |
| 2016-4 | site 17 | Sediment | n.a. | 1.85 | 0.20 | -26.40 | 6.30 | 0.82 | 0.63 | 0.32 |
| 2016-4 | site 17 | Sediment | n.a. | 2.51 | 0.29 | -26.00 | 5.90 | 0.75 | 0.56 | 0.21 |
| 2016-4 | site 17 | Sediment | n.a. | 1.04 | 0.11 | -26.10 | 7.30 | 0.81 | 0.63 | 0.29 |
| 2016-4 | site 19 | Sediment | n.a. | 1.98 | 0.22 | -24.71 | 5.61 | 0.58 | 0.61 | 0.21 |
| 2016-4 | site 19 | Sediment | n.a. | 1.21 | 0.13 | -24.77 | 6.21 | 0.97 | 0.59 | 0.47 |
| 2016-4 | site 19 | Sediment | n.a. | 1.28 | 0.15 | -24.82 | 6.83 | 0.57 | 0.62 | 0.12 |
| 2016-4 | site 19 | Sediment | n.a. | 1.22 | 0.14 | -24.33 | 5.49 | 0.53 | 0.61 | 0.23 |
| 2016-4 | site 19 | Sediment | n.a. | 1.25 | 0.15 | -24.36 | 6.03 | 0.52 | 0.58 | 0.21 |
| 2016-4 | site 19 | Sediment | n.a. | 1.00 | 0.12 | -24.20 | 7.10 | 0.68 | 0.56 | 0.18 |
| 2016-4 | site 19 | Sediment | n.a. | 1.19 | 0.14 | -24.40 | 7.40 | 0.51 | 0.59 | 0.16 |
| 2016-4 | site 19 | Sediment | n.a. | 1.85 | 0.22 | -24.18 | 6.73 | 0.51 | 0.61 | 0.19 |
| 2016-4 | site 19 | Sediment | n.a. | 0.76 | 0.09 | -24.40 | 7.00 | 0.53 | 0.63 | 0.15 |
| 2016-4 | site 19 | Sediment | n.a. | 1.40 | 0.17 | -24.21 | 7.38 | 0.54 | 0.56 | 0.16 |
| 2019-6 | site 4 | SPM | Low-flow | 4.22 | 0.58 | -33.53 | 10.35 | 0.99 | 0.68 | 0.46 |
| 2019-6 | site 4 | SPM | Low-flow | 4.06 | 0.57 | -33.13 | 10.01 | 0.98 | 0.67 | 0.44 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2019-7 | site 4 | SPM | Low-flow | 2.72 | 0.44 | -32.79 | 10.25 | 0.97 | 0.69 | 0.50 |
| 2019-7 | site 4 | SPM | Low-flow | 2.41 | 0.43 | -33.12 | 10.40 | 0.97 | 0.70 | 0.52 |
| 2020-9 | site 4 | SPM | Low-flow | 3.33 | 0.31 | -31.42 | 7.60 | 0.95 | 0.68 | 0.51 |
| 2020-9 | site 4 | SPM | Low-flow | 2.95 | 0.32 | -30.33 | 7.36 | 0.96 | 0.69 | 0.53 |
| 2019-6 | site 13 | SPM | Low-flow | 4.04 | 0.49 | -29.37 | 11.10 | 0.96 | 0.72 | 0.29 |
| 2019-6 | site 13 | SPM | Low-flow | 5.28 | 0.52 | -30.70 | 13.42 | 0.96 | 0.73 | 0.27 |
| 2019-6 | site 13 | SPM | Low-flow | 4.22 | 0.39 | -28.38 | 10.68 | 0.92 | 0.70 | 0.27 |
| 2019-6 | site 13 | SPM | Low-flow | 4.58 | 0.42 | -28.32 | 9.35 | 0.91 | 0.68 | 0.25 |
| 2019-6 | site 13 | SPM | Low-flow | 3.55 | 0.39 | -29.53 | 10.83 | 0.87 | 0.69 | 0.24 |
| 2019-6 | site 13 | SPM | Low-flow | 4.24 | 0.53 | -29.47 | 11.00 | 0.91 | 0.68 | 0.25 |
| 2019-7 | site 13 | SPM | Low-flow | 3.78 | 0.27 | -26.73 | 8.66 | 0.71 | 0.64 | 0.20 |
| 2019-7 | site 13 | SPM | Low-flow | 3.60 | 0.28 | -27.18 | 8.52 | 0.70 | 0.65 | 0.24 |
| 2019-7 | site 13 | SPM | Low-flow | 5.94 | 0.30 | -26.54 | 8.39 | 0.70 | 0.66 | 0.24 |
| 2019-7 | site 13 | SPM | Low-flow | 4.72 | 0.36 | -26.55 | 8.53 | 0.69 | 0.61 | 0.23 |
| 2019-7 | site 13 | SPM | Low-flow | 4.92 | 0.35 | -27.16 | 8.33 | 0.71 | 0.67 | 0.30 |
| 2019-7 | site 13 | SPM | Low-flow | 4.43 | 0.40 | -27.87 | 8.97 | 0.66 | 0.63 | 0.17 |
| 2020-9 | site 13 | SPM | Low-flow | 7.35 | 0.40 | -28.46 | 7.27 | 0.78 | 0.72 | 0.24 |
| 2020-9 | site 13 | SPM | Low-flow | 5.18 | 0.43 | -28.30 | 7.15 | 0.77 | 0.72 | 0.22 |
| 2020-9 | site 13 | SPM | Low-flow | 5.44 | 0.40 | -28.25 | 7.25 | 0.71 | 0.67 | 0.19 |
| 2020-9 | site 13 | SPM | Low-flow | 5.22 | 0.36 | -27.73 | 6.92 | 0.73 | 0.68 | 0.25 |
| 2020-9 | site 13 | SPM | Low-flow | 5.99 | 0.34 | -26.63 | 6.59 | 0.71 | 0.66 | 0.24 |
| 2020-9 | site 13 | SPM | Low-flow | 4.64 | 0.36 | -27.29 | 6.81 | 0.74 | 0.69 | 0.22 |
| 2019-6 | site 6 | SPM | Low-flow | 4.71 | 0.57 | -31.80 | 7.65 | 0.99 | 0.70 | 0.42 |
| 2019-6 | site 6 | SPM | Low-flow | 4.20 | 0.53 | -32.74 | 7.86 | 0.99 | 0.69 | 0.42 |
| 2019-6 | site 6 | SPM | Low-flow | 3.91 | 0.42 | -31.23 | 7.85 | 0.99 | 0.68 | 0.38 |
| 2019-6 | site 6 | SPM | Low-flow | 4.64 | 0.51 | -31.98 | 7.68 | 0.99 | 0.71 | 0.35 |
| 2019-6 | site 6 | SPM | Low-flow | 3.93 | 0.55 | -32.23 | 6.90 | 0.98 | 0.72 | 0.40 |
| 2019-6 | site 6 | SPM | Low-flow | 4.86 | 0.62 | -32.78 | 8.17 | 0.99 | 0.71 | 0.44 |
| 2019-7 | site 6 | SPM | Low-flow | 3.72 | 0.48 | -33.42 | 9.67 | 0.97 | 0.73 | 0.40 |
| 2019-7 | site 6 | SPM | Low-flow | 3.93 | 0.48 | -33.57 | 9.47 | 0.97 | 0.73 | 0.38 |
| 2019-7 | site 6 | SPM | Low-flow | 3.31 | 0.44 | -31.71 | 9.20 | 0.97 | 0.75 | 0.45 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2019-7 | site 6 | SPM | Low-flow | 4.57 | 0.43 | -31.94 | 8.33 | 0.97 | 0.75 | 0.43 |
| 2019-7 | site 6 | SPM | Low-flow | 5.14 | 0.40 | -31.12 | 8.29 | 0.97 | 0.74 | 0.34 |
| 2019-7 | site 6 | SPM | Low-flow | 4.30 | 0.48 | -32.21 | 8.47 | 0.96 | 0.73 | 0.35 |
| 2020-9 | site 6 | SPM | Low-flow | 4.03 | 0.36 | -31.42 | 7.91 | 0.96 | 0.71 | 0.47 |
| 2020-9 | site 6 | SPM | Low-flow | 4.46 | 0.39 | -31.37 | 7.86 | 0.96 | 0.71 | 0.47 |
| 2020-9 | site 6 | SPM | Low-flow | 5.49 | 0.37 | -30.09 | 7.28 | 0.96 | 0.72 | 0.40 |
| 2020-9 | site 6 | SPM | Low-flow | 4.53 | 0.36 | -31.71 | 8.36 | 0.97 | 0.72 | 0.44 |
| 2020-9 | site 6 | SPM | Low-flow | 4.82 | 0.37 | -30.94 | 7.78 | 0.96 | 0.72 | 0.47 |
| 2020-9 | site 6 | SPM | Low-flow | 4.49 | 0.38 | -32.45 | 8.28 | 0.96 | 0.72 | 0.47 |
| 2019-6 | site 15 | SPM | Low-flow | 4.51 | 0.34 | -26.61 | 8.54 | 0.88 | 0.60 | 0.21 |
| 2019-6 | site 15 | SPM | Low-flow | 6.00 | 0.39 | -26.74 | 8.57 | 0.89 | 0.58 | 0.20 |
| 2019-6 | site 15 | SPM | Low-flow | 6.38 | 0.42 | -25.96 | 8.25 | 0.89 | 0.47 | 0.24 |
| 2019-6 | site 15 | SPM | Low-flow | 6.79 | 0.40 | -26.08 | 8.96 | 0.87 | 0.58 | 0.21 |
| 2019-6 | site 15 | SPM | Low-flow | 6.00 | 0.39 | -25.95 | 8.44 | 0.76 | 0.58 | 0.22 |
| 2019-6 | site 15 | SPM | Low-flow | 3.05 | 0.27 | -26.52 | 9.54 | 0.84 | 0.57 | 0.22 |
| 2019-7 | site 15 | SPM | Low-flow | 3.66 | 0.30 | -26.26 | 8.57 | 0.67 | 0.57 | 0.17 |
| 2019-7 | site 15 | SPM | Low-flow | 2.99 | 0.23 | -26.49 | 9.23 | 0.65 | 0.59 | 0.21 |
| 2019-7 | site 15 | SPM | Low-flow | 6.53 | 0.33 | -25.61 | 8.79 | 0.64 | 0.58 | 0.23 |
| 2019-7 | site 15 | SPM | Low-flow | 2.99 | 0.21 | -25.88 | 10.02 | 0.65 | 0.58 | 0.17 |
| 2019-7 | site 15 | SPM | Low-flow | 5.74 | 0.34 | -25.88 | 8.47 | 0.64 | 0.58 | 0.23 |
| 2019-7 | site 15 | SPM | Low-flow | 2.02 | 0.23 | -25.97 | 9.52 | 0.63 | 0.59 | 0.20 |
| 2020-9 | site 15 | SPM | Low-flow | 5.52 | 0.39 | -26.03 | 6.67 | 0.69 | 0.60 | 0.23 |
| 2020-9 | site 15 | SPM | Low-flow | 4.93 | 0.35 | -26.55 | 6.95 | 0.68 | 0.62 | 0.20 |
| 2020-9 | site 15 | SPM | Low-flow | 5.69 | 0.37 | -25.92 | 6.88 | 0.65 | 0.61 | 0.17 |
| 2020-9 | site 15 | SPM | Low-flow | 4.37 | 0.31 | -25.63 | 6.91 | 0.67 | 0.58 | 0.24 |
| 2020-9 | site 15 | SPM | Low-flow | 5.89 | 0.42 | -25.66 | 6.62 | 0.66 | 0.61 | 0.17 |
| 2020-9 | site 15 | SPM | Low-flow | 2.85 | 0.20 | -26.51 | 7.57 | 0.65 | 0.59 | 0.21 |
| 2019-6 | site 10 | SPM | Low-flow | 4.16 | 0.53 | -32.29 | 8.52 | 0.98 | 0.74 | 0.38 |
| 2019-6 | site 10 | SPM | Low-flow | 4.06 | 0.58 | -33.78 | 9.55 | 0.99 | 0.71 | 0.38 |
| 2019-6 | site 10 | SPM | Low-flow | 4.92 | 0.56 | -31.83 | 9.58 | 0.98 | 0.73 | 0.37 |
| 2019-6 | site 10 | SPM | Low-flow | 6.16 | 0.57 | -30.62 | 10.89 | 0.98 | 0.73 | 0.38 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2019-6 | site 10 | SPM | Low-flow | 3.71 | 0.54 | -33.67 | 9.57 | 0.98 | 0.73 | 0.42 |
| 2019-6 | site 10 | SPM | Low-flow | 4.64 | 0.70 | -33.15 | 9.30 | 0.98 | 0.73 | 0.37 |
| 2019-7 | site 10 | SPM | Low-flow | 3.22 | 0.38 | -28.99 | 9.65 | 0.93 | 0.77 | 0.33 |
| 2019-7 | site 10 | SPM | Low-flow | 4.50 | 0.57 | -29.91 | 10.43 | 0.91 | 0.76 | 0.28 |
| 2019-7 | site 10 | SPM | Low-flow | 4.51 | 0.47 | -29.15 | 8.72 | 0.93 | 0.76 | 0.30 |
| 2019-7 | site 10 | SPM | Low-flow | 4.65 | 0.65 | -29.72 | 9.41 | 0.89 | 0.74 | 0.33 |
| 2019-7 | site 10 | SPM | Low-flow | 4.10 | 0.51 | -30.23 | 10.33 | 0.93 | 0.78 | 0.37 |
| 2019-7 | site 10 | SPM | Low-flow | 4.09 | 0.69 | -30.76 | 9.49 | 0.90 | 0.77 | 0.29 |
| 2020-9 | site 10 | SPM | Low-flow | 6.68 | 0.47 | -29.16 | 8.29 | 0.94 | 0.77 | 0.33 |
| 2020-9 | site 10 | SPM | Low-flow | 3.67 | 0.51 | -31.11 | 8.23 | 0.94 | 0.77 | 0.33 |
| 2020-9 | site 10 | SPM | Low-flow | 4.93 | 0.42 | -29.10 | 8.84 | 0.93 | 0.78 | 0.38 |
| 2020-9 | site 10 | SPM | Low-flow | 5.69 | 0.51 | -29.53 | 7.64 | 0.92 | 0.78 | 0.35 |
| 2020-9 | site 10 | SPM | Low-flow | 5.46 | 0.64 | -30.67 | 9.88 | 0.94 | 0.78 | 0.40 |
| 2020-9 | site 10 | SPM | Low-flow | 6.05 | 0.68 | -31.01 | 10.79 | 0.93 | 0.78 | 0.35 |
| 2019-6 | site 18 | SPM | Low-flow | 4.29 | 0.29 | -26.31 | 9.37 | 0.76 | 0.59 | 0.19 |
| 2020-9 | site 18 | SPM | Low-flow | 2.46 | 0.18 | -25.84 | 7.21 | 0.66 | 0.58 | 0.17 |
| 2019-6 | site 9 | SPM | Low-flow | 3.48 | 0.47 | -32.12 | 7.90 | 0.97 | 0.72 | 0.41 |
| 2020-9 | site 8 | SPM | Low-flow | 2.86 | 0.29 | -30.79 | 7.04 | 0.95 | 0.76 | 0.37 |
| 2019-6 | site 16 | SPM | Low-flow | 3.89 | 0.34 | -27.08 | 9.98 | 0.87 | 0.60 | 0.16 |
| 2020-9 | site 15 | SPM | Low-flow | 4.48 | 0.32 | -26.04 | 7.34 | 0.67 | 0.60 | 0.24 |
| 2019-6 | site 14 | SPM | Low-flow | 5.23 | 0.43 | -28.58 | 10.22 | 0.94 | 0.70 | 0.30 |
| 2020-9 | site 13 | SPM | Low-flow | 3.54 | 0.37 | -29.60 | 8.93 | 0.78 | 0.71 | 0.27 |
| 2019-6 | site 12 | SPM | Low-flow | 4.33 | 0.53 | -32.75 | 14.10 | 0.98 | 0.74 | 0.27 |
| 2020-9 | site 11 | SPM | Low-flow | 3.27 | 0.42 | -31.34 | 10.04 | 0.89 | 0.77 | 0.37 |
| 2020-11 | site 3 | SPM | Low-flow | 1.61 | 0.27 | -29.48 | 2.76 | 0.98 | 0.67 | 0.61 |
| 2021-9 | site B | Soil | n.a. | 0.30 | 0.01 | -26.85 | 11.62 | 0.96 | 0.44 | 0.62 |
| 2021-9 | site B | Soil | n.a. | 0.40 | 0.04 | -26.75 | 9.33 | 0.94 | 0.53 | 0.69 |
| 2021-9 | site B | Soil | n.a. | 4.92 | 0.30 | -26.86 | 6.78 | 0.85 | 0.61 | 0.15 |
| 2021-9 | site B | Soil | n.a. | 13.75 | 0.44 | -28.73 | 1.11 | 0.75 | 0.78 | 0.12 |
| 2021-9 | site A | Soil | n.a. | 4.73 | 0.32 | -26.71 | 7.14 | 0.75 | 0.68 | 0.18 |
| 2021-10 | site C | Soil | n.a. | 3.64 | 0.34 | -27.94 | 3.20 | 0.86 | 0.71 | 0.37 |

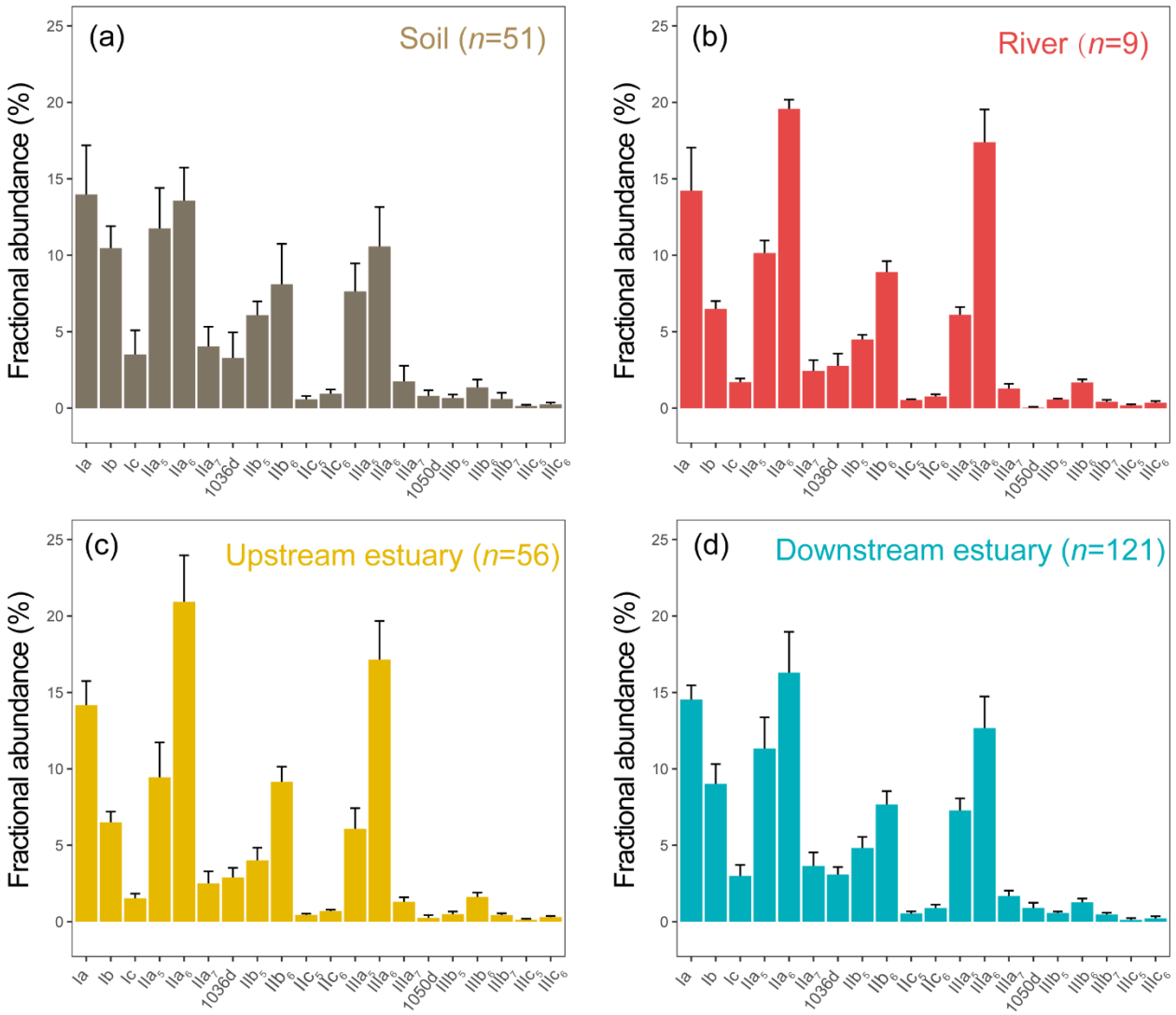
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2021-10 | site C | Soil | n.a. | 3.74 | 0.29 | -29.25 | 3.43 | 0.95 | 0.59 | 0.35 |
| 2021-9 | site B | Soil | n.a. | 22.28 | 1.07 | -27.87 | 1.58 | 0.91 | 0.71 | 0.35 |
| 2021-10 | site C | Soil | n.a. | 1.07 | 0.13 | -25.14 | 5.36 | 0.93 | 0.34 | 0.60 |
| 2021-2 | site 11 | SPM | High-flow | 2.92 | 0.22 | -28.48 | 4.77 | 0.97 | 0.57 | 0.49 |
| 2021-3 | site 11 | SPM | High-flow | 2.54 | 0.21 | -29.05 | 5.11 | 0.97 | 0.66 | 0.48 |
| 2021-2 | site 13 | SPM | High-flow | 3.26 | 0.22 | -28.45 | 4.90 | 0.97 | 0.58 | 0.38 |
| 2021-3 | site 13 | SPM | High-flow | 2.60 | 0.17 | -29.23 | 6.77 | 0.96 | 0.65 | 0.45 |
| 2021-3 | site 15 | SPM | High-flow | 2.78 | 0.21 | -28.85 | 6.75 | 0.94 | 0.66 | 0.44 |
| 2021-2 | site 15 | SPM | High-flow | 3.12 | 0.19 | -28.74 | 5.71 | 0.97 | 0.60 | 0.37 |
| 2021-3 | site 8 | SPM | High-flow | 3.10 | 0.24 | -29.24 | 4.97 | 0.98 | 0.63 | 0.51 |
| 2021-2 | site 8 | SPM | High-flow | 1.61 | 0.12 | -28.43 | 5.35 | 0.97 | 0.50 | 0.53 |
| 2021-3 | site 18 | SPM | High-flow | 5.02 | 0.24 | -27.05 | 6.49 | 0.81 | 0.59 | 0.28 |
| 2021-2 | site 18 | SPM | High-flow | 3.40 | 0.20 | -27.59 | 6.14 | 0.87 | 0.58 | 0.30 |
| 2020-11 | site 5 | SPM | Low-flow | 0.75 | 0.19 | -29.66 | 7.23 | 0.97 | 0.68 | 0.57 |
| 2020-11 | site 1 | SPM | Low-flow | 0.82 | 0.12 | -29.10 | 8.02 | 0.97 | 0.70 | 0.42 |
| 2020-11 | site 2 | SPM | Low-flow | 3.80 | 0.28 | -28.82 | 5.24 | 0.97 | 0.72 | 0.57 |
| 2021-3 | site D | Soil | n.a. | 3.23 | 0.22 | -29.33 | 5.70 | 0.88 | 0.48 | 0.07 |
| 2020-9 | site D | Soil | n.a. | 3.15 | 0.39 | -28.83 | 6.47 | 0.88 | 0.47 | 0.12 |
| 2021-3 | site D | Soil | n.a. | 1.13 | 0.14 | -26.21 | n.a. | 0.77 | 0.62 | 0.31 |
| 2020-9 | site D | Soil | n.a. | 1.78 | 0.21 | -26.24 | n.a. | 0.75 | 0.52 | 0.20 |
| 2021-3 | site D | Soil | n.a. | 4.37 | 0.30 | -26.48 | 7.76 | 0.75 | 0.56 | 0.13 |
| 2020-9 | site D | Soil | n.a. | 3.94 | 0.25 | -26.40 | 8.34 | 0.77 | 0.54 | 0.17 |
| 2021-3 | site D | Soil | n.a. | 2.74 | 0.16 | -27.12 | 7.17 | 0.81 | 0.58 | 0.22 |
| 2020-9 | site D | Soil | n.a. | 1.49 | 0.13 | -26.41 | 7.51 | 0.79 | 0.52 | 0.24 |
| 2021-3 | site E | Soil | n.a. | 4.41 | 0.37 | -28.76 | 8.72 | 0.91 | 0.63 | 0.13 |
| 2020-9 | site E | Soil | n.a. | 4.87 | 0.35 | -28.81 | 9.06 | 0.90 | 0.64 | 0.11 |
| 2021-3 | site E | Soil | n.a. | 2.17 | 0.17 | -26.95 | 8.00 | 0.84 | 0.57 | 0.13 |
| 2020-9 | site E | Soil | n.a. | 4.36 | 0.28 | -26.23 | 7.55 | 0.73 | 0.59 | 0.19 |
| 2021-3 | site E | Soil | n.a. | 3.44 | 0.28 | -25.97 | 7.64 | 0.73 | 0.61 | 0.16 |
| 2020-9 | site E | Soil | n.a. | 4.42 | 0.29 | -25.68 | 7.55 | 0.71 | 0.58 | 0.16 |
| 2021-3 | site E | Soil | n.a. | 3.90 | 0.23 | -25.99 | 7.59 | 0.73 | 0.59 | 0.16 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2020-9 | site E | Soil | n.a. | 3.72 | 0.25 | -25.74 | 7.64 | 0.73 | 0.58 | 0.19 |
| 2021-3 | site E | Soil | n.a. | 1.72 | 0.16 | -25.93 | 7.55 | 0.73 | 0.58 | 0.18 |
| 2020-9 | site E | Soil | n.a. | 1.33 | 0.14 | -25.85 | 7.67 | 0.72 | 0.56 | 0.19 |
| 2018-8 | site E | Soil | n.a. | 1.32 | 0.15 | -26.46 | n.a. | 0.93 | 0.55 | 0.14 |
| 2018-8 | site E | Soil | n.a. | 1.55 | 0.17 | -26.90 | n.a. | 0.92 | 0.54 | 0.14 |
| 2018-8 | site E | Soil | n.a. | 3.63 | 0.44 | -24.80 | n.a. | 0.91 | 0.54 | 0.14 |
| 2018-8 | site E | Soil | n.a. | 1.88 | 0.22 | -25.47 | n.a. | 0.93 | 0.54 | 0.14 |
| 2018-8 | site E | Soil | n.a. | 1.57 | 0.18 | -26.32 | n.a. | 0.92 | 0.54 | 0.17 |
| 2018-8 | site E | Soil | n.a. | 2.57 | 0.29 | -25.89 | n.a. | 0.91 | 0.55 | 0.16 |
| 2018-6 | site E | Soil | n.a. | 3.11 | 0.39 | -25.36 | n.a. | 0.92 | 0.55 | 0.16 |
| 2018-6 | site E | Soil | n.a. | 3.65 | 0.41 | -25.81 | n.a. | 0.93 | 0.55 | 0.16 |
| 2018-6 | site E | Soil | n.a. | 4.02 | 0.44 | -25.85 | n.a. | 0.93 | 0.55 | 0.16 |
| 2018-6 | site E | Soil | n.a. | 3.24 | 0.34 | -26.17 | n.a. | 0.91 | 0.56 | 0.16 |
| 2018-6 | site E | Soil | n.a. | 3.93 | 0.44 | -25.38 | n.a. | 0.90 | 0.54 | 0.17 |
| 2018-6 | site E | Soil | n.a. | 2.45 | 0.26 | -26.31 | n.a. | 0.89 | 0.55 | 0.17 |
| 2018-10 | site E | Soil | n.a. | 1.94 | 0.25 | -25.54 | n.a. | 0.70 | 0.55 | 0.18 |
| 2018-10 | site E | Soil | n.a. | 0.88 | 0.13 | -25.14 | n.a. | 0.68 | 0.54 | 0.08 |
| 2018-10 | site E | Soil | n.a. | 0.97 | 0.12 | -26.89 | n.a. | 0.70 | 0.52 | 0.20 |
| 2018-10 | site E | Soil | n.a. | 1.29 | 0.15 | -26.58 | n.a. | 0.68 | 0.54 | 0.18 |
| 2018-2 | site E | Soil | n.a. | 0.47 | 0.06 | -25.45 | n.a. | 0.77 | 0.54 | 0.15 |
| 2018-2 | site E | Soil | n.a. | 0.80 | 0.08 | -25.91 | n.a. | 0.72 | 0.56 | 0.14 |
| 2018-2 | site E | Soil | n.a. | 0.24 | 0.03 | -25.17 | n.a. | 0.71 | 0.53 | 0.14 |
| 2018-2 | site E | Soil | n.a. | 0.39 | 0.05 | -25.22 | n.a. | 0.74 | 0.55 | 0.16 |
| 2018-10 | site E | Soil | n.a. | 1.54 | 0.17 | -27.09 | n.a. | 0.77 | 0.54 | 0.17 |
| 2018-10 | site E | Soil | n.a. | 1.17 | 0.14 | -26.89 | n.a. | 0.86 | 0.54 | 0.20 |
| 2018-2 | site E | Soil | n.a. | 0.62 | 0.06 | -26.48 | n.a. | 0.79 | 0.54 | 0.30 |
| 2018-2 | site E | Soil | n.a. | 0.22 | 0.03 | -27.80 | n.a. | 0.74 | 0.53 | 0.23 |

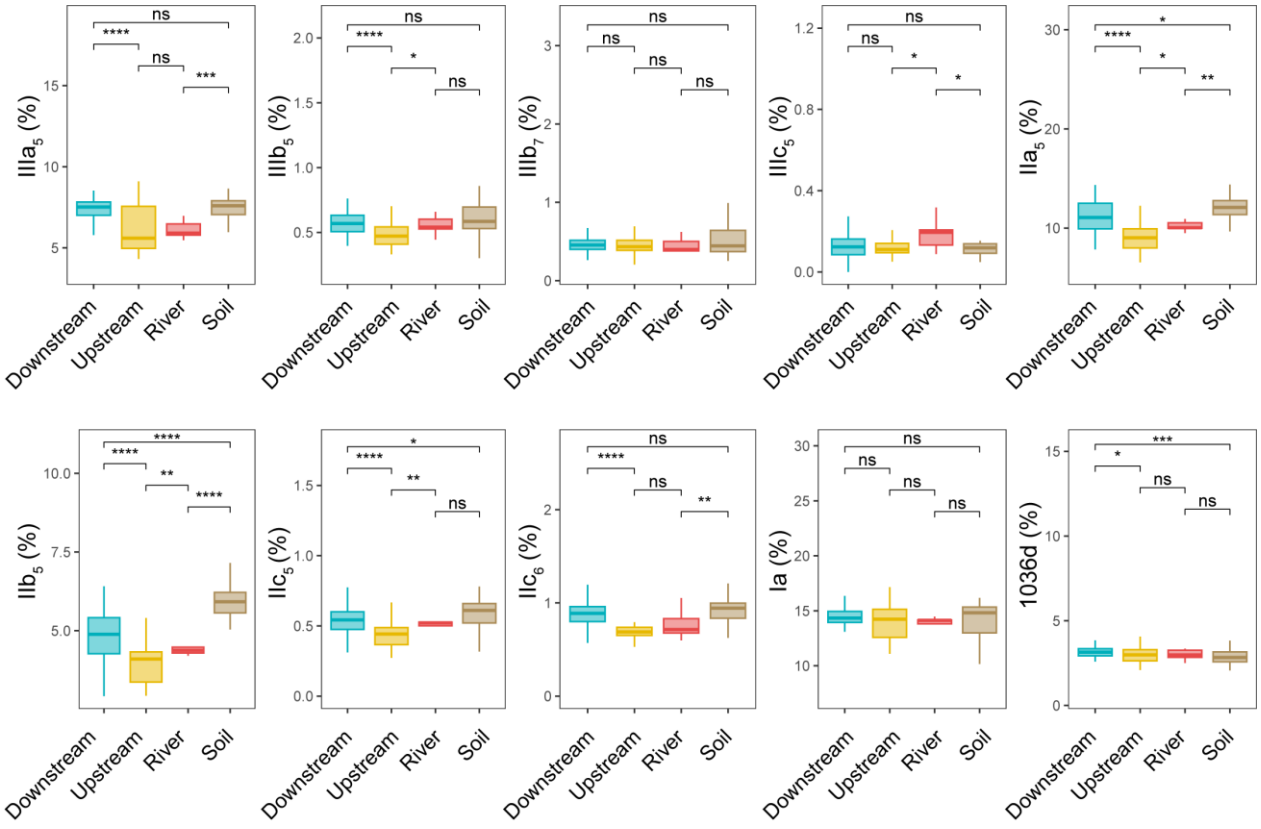
n.a.= not applicable

Supplementary Table 3-2 RDA results

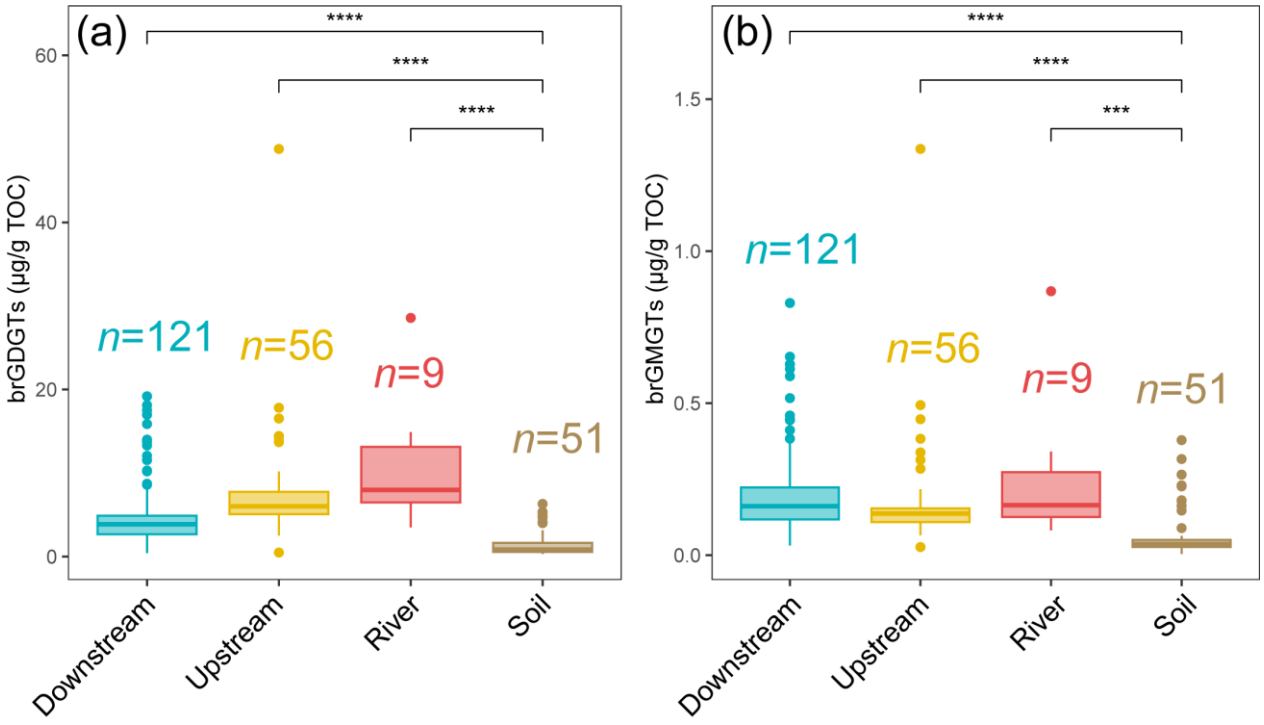
| | variables | RDA scores | | Individual Importance (%) |
|---|---|---|---|---|
| | | Axis 1 | Axis 2 | |
| brGDGTs | TOC | 0.17 | -0.49 | 2.12 * |
| | TN | 0.72 | -0.23 | 13.47 *** |
| | Temperature | 0.41 | -0.82 | 3.6 ** |
| | Salinity | -0.73 | -0.10 | 14.97 *** |
| | Water discharge | -0.30 | 0.93 | 3.68 ** |
| brGMGTs | TOC | 0.08 | 0.68 | 3.5 ** |
| | TN | -0.34 | 0.09 | 4.18 ** |
| | Temperature | 0.04 | 0.64 | 1.17 ns |
| | Salinity | 0.83 | -0.50 | 17.45 *** |
| | Water discharge | -0.17 | -0.71 | 2.16 * |



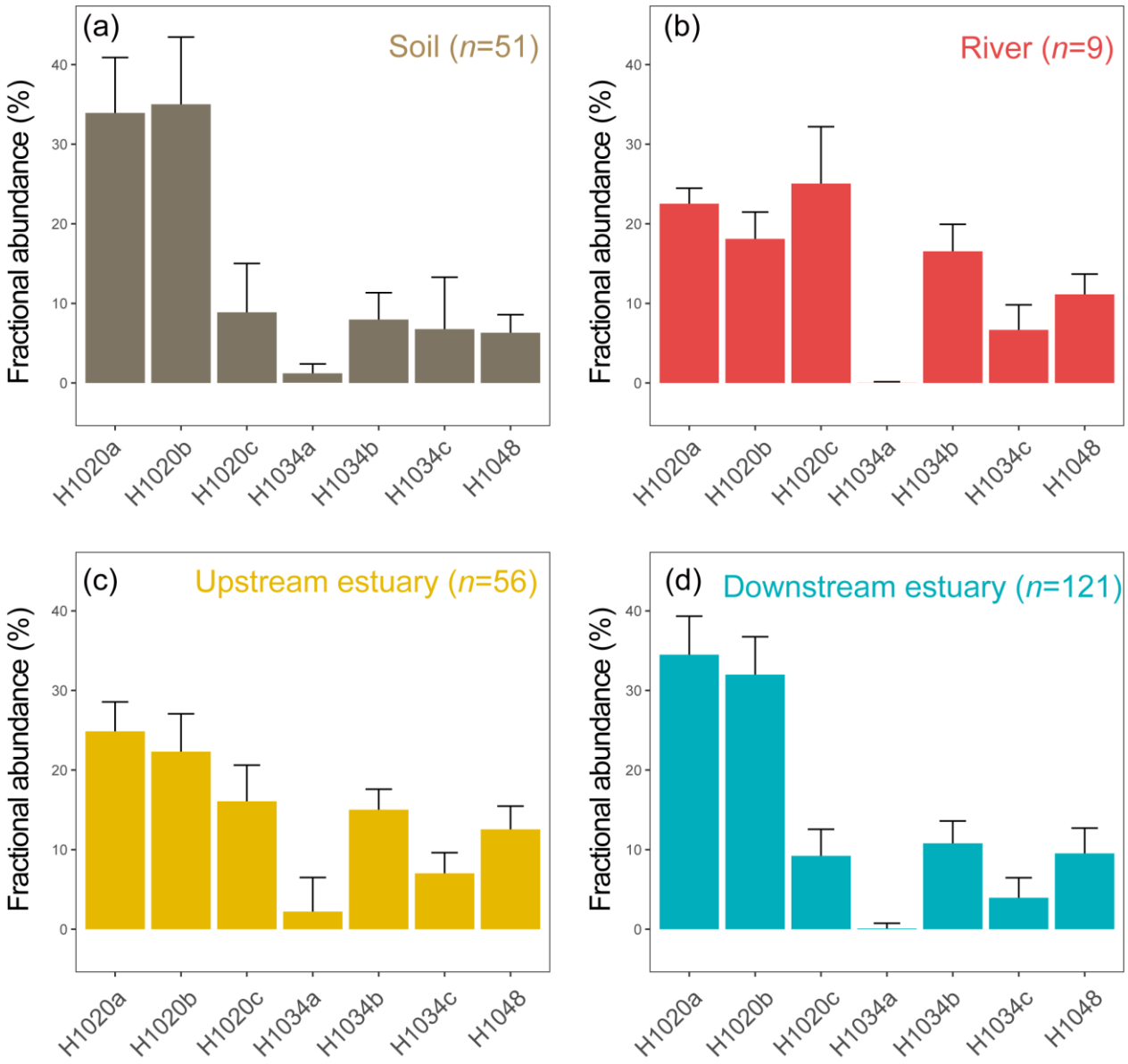

**Supplementary Figure 3-1.** Distribution of brGDGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (n=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin.



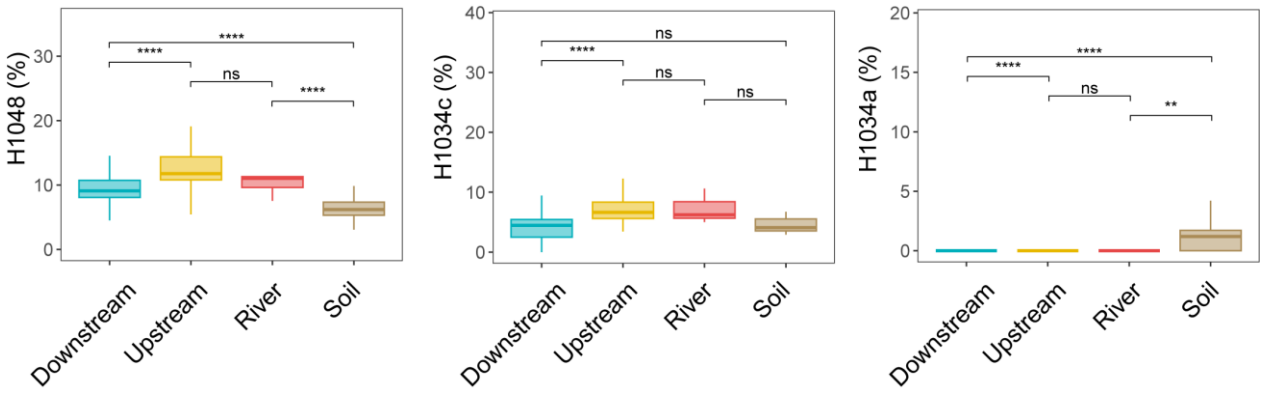

**Supplementary Figure 3-2.** Relative abundance of brGDGTs across the Seine River basin. Box plots of upstream and downstream estuary are composed of SPM and river channel sediments, whereas those of river are composed of SPM. Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).



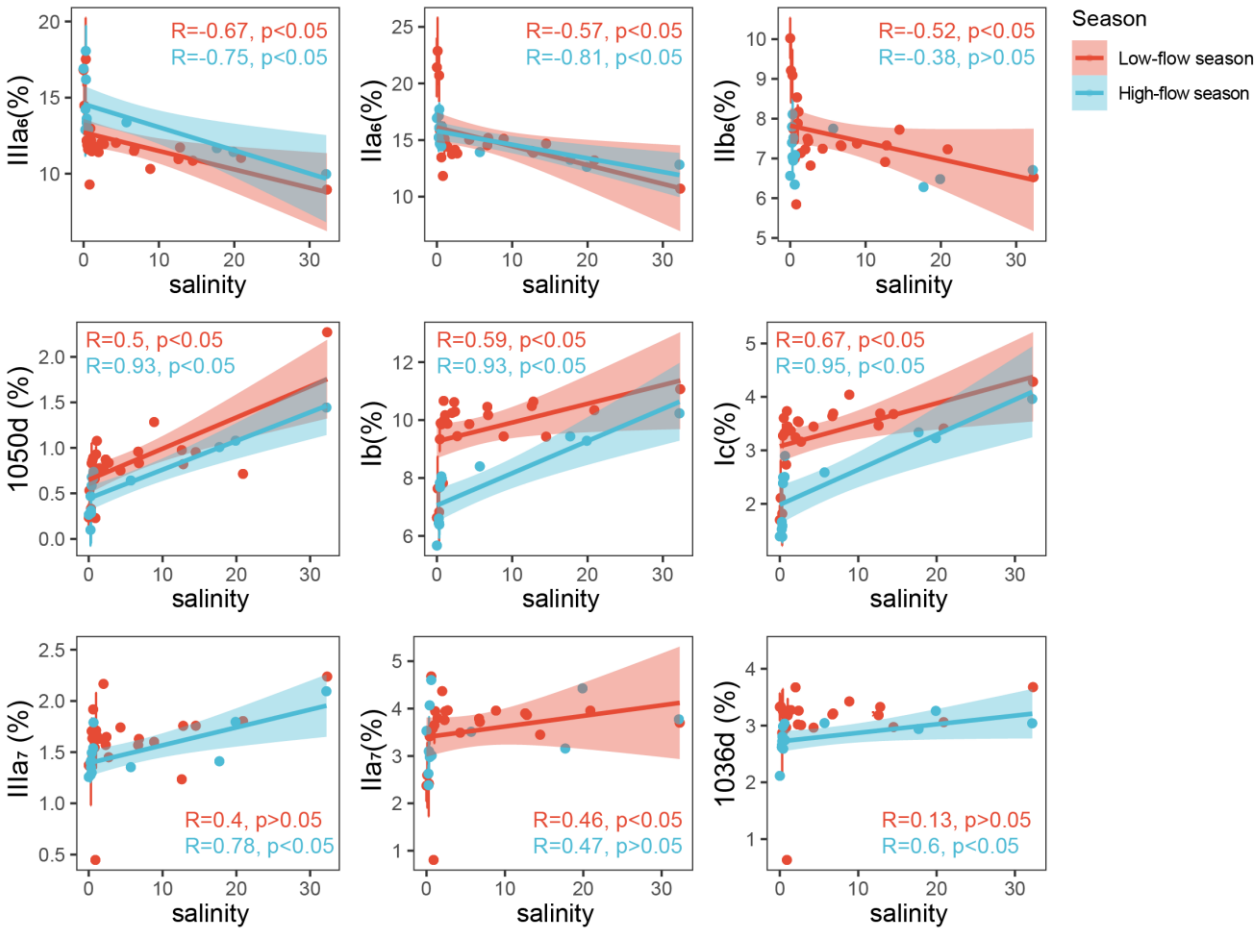

**Supplementary Figure 3-3.** Concentrations (normalized to total organic carbon) of (a) total brGDGTs and (b) total brGMGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (n=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin. Box plots of upstream and downstream estuary samples are based on SPM and river channel sediments, whereas those of river samples are based only on SPM. Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).



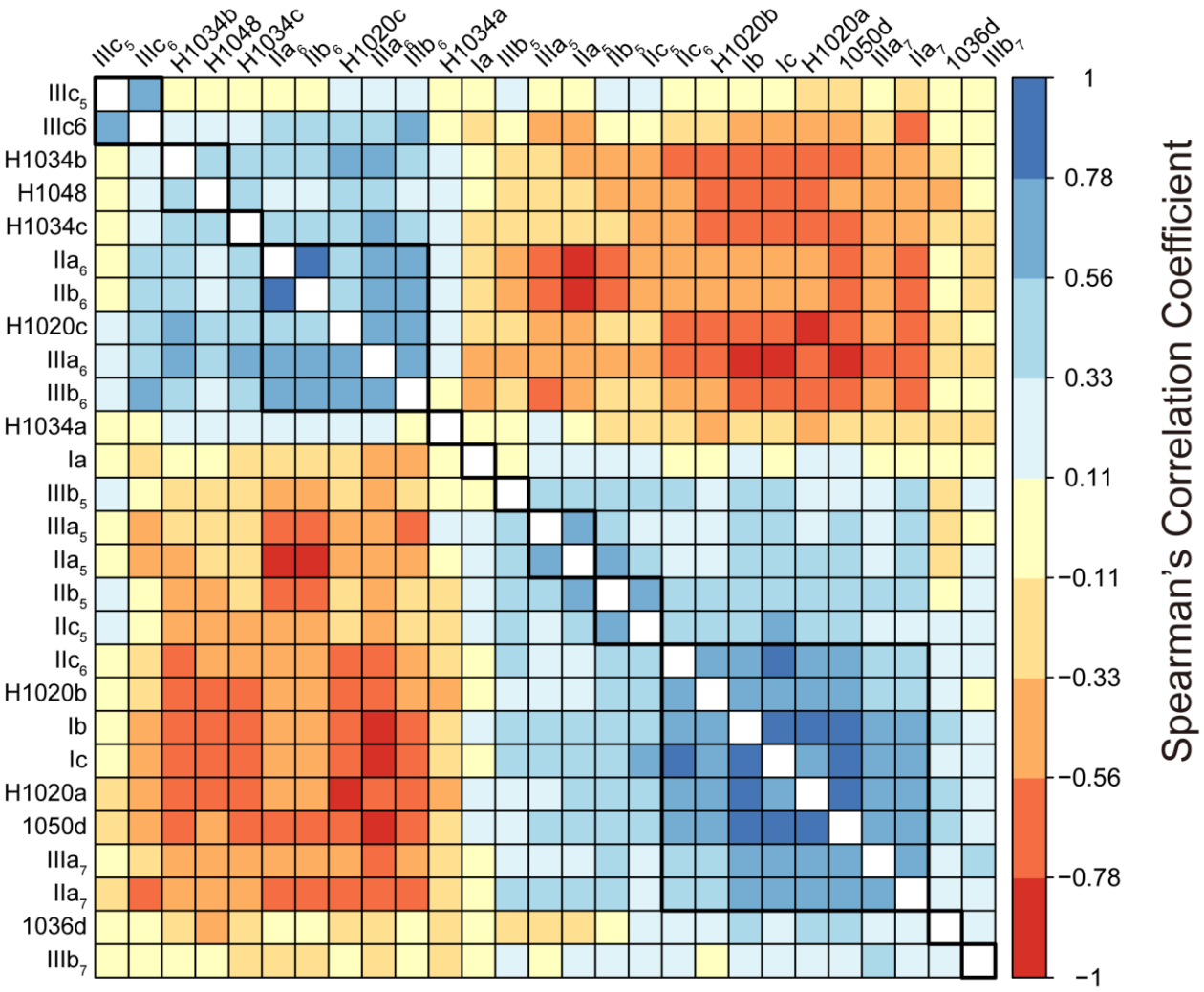

**Supplementary Figure 3-4.** Distribution of brGMGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (n=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin.




**Supplementary Figure 3-5.** Relative abundance of brGMGTs across the Seine River basin. Box plots of upstream and downstream estuary are composed of SPM and sediments, whereas those of river are composed of SPM. Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).




**Supplementary Figure 3-6.** Salinity plotted versus relative abundance of 6-methyl and 7-methyl brGDGTs ($IIIa_6$, $IIa_6$, $IIb_6$, $IIIa_7$ and $IIa_7$) as well as compounds 1050d, 1036d, Ib, and Ic. Shaded area represents 95% confidence intervals. Vertical error bars indicate mean ± s.d for samples with the same salinity. Dataset is composed of SPM.




**Supplementary Figure 3-7.** (a) Correlation plot between fractional abundance of brGDGTs (relative to all brGDGTs) and brGMGTs (relative to all brGMGTs).

# Chapter 4:

## Dynamics of particulate organic matter in a human-impacted estuary influenced by hydroclimate conditions and land use characteristics

This chapter is in preparation for submission to *Chemical Geology*

## Abstract

Estuaries play an important role in regulating Particulate Organic Matter (POM), which is controlled by natural (hydroclimate conditions) and anthropogenic (land use changes) processes. To date, the interactions between these processes and POM dynamics are mainly investigated at the bulk level, hampering our understanding of POM behavior and estuarine functioning. Here, we investigate the spatio-temporal variations of POM characteristics using water samples ($n$=172) collected along the land-sea continuum of a human-impacted estuary (Seine Estuary, NW France) under low-flow and high-flow conditions. POM dynamics are studied at the bulk and molecular levels using elemental and isotopic analysis, as well as lipid biomarkers (sterols, stanols, fatty acids, and *n*-alkanes). Our results show that the dispersal and dynamics of distinct types of POM, are closely related to hydroclimate conditions and land use types. Specifically, anthropogenic POM gradually decreases along the estuary in both high-flow and low-flow seasons, which is related with water discharge and urban land use. Furthermore, phytoplankton blooms and potential priming effect are observed in an agriculturally impacted zone, whereas at high flows, the zone for these processes shifts downstream. Our study suggests that estuaries act as effective filters that dilute sewage contaminants, as well as natural reactors that promote phytoplankton blooms and potentially trigger the priming effect.

*Keywords:* POM; biomarker; Land-ocean continuum; Land use

## 4.1. Introduction

Estuaries are dynamic systems that contribute to over 80% of the global organic carbon burial (Gattuso et al., 1998) and play an important role in controlling the spatio-temporal variabilities of organic matter and associated biogeochemical cycling (Bianchi, 2007; Bianchi and Canuel, 2011). Understanding the sources, transformations and fate of Particulate Organic Matter (POM) in estuaries is crucial in assessing climate change, fisheries management, and biogeochemical impacts (Bianchi, 2011; Canuel et al., 2012; Darnaude, 2005; Cai, 2011). However, investigating the dynamics of estuarine POM is especially challenging due to the multiple sources of the latter and complex biogeochemical processes as well as high variability of the environmental parameters influencing POM characteristics (Bianchi, 2007; Bianchi and Canuel, 2011; Bibi et al., 2020; Goñi et al., 2021).

Estuarine waters are productive and dynamic systems containing organic matter from natural and anthropogenic sources. Dynamics of POM from these sources could be associated with natural processes (i.e. hydroclimate conditions). For example, low discharge increases residence time of nutrient, thus regulating the processing of estuarine POM (Li et al., 2021; Romero et al., 2019). In addition, anthropogenic activities (i.e. land use changes) may alter nutrient inputs and primary productivity, hence significantly influencing estuarine biogeochemical processing (David et al., 2020). To date, the relationships between land use changes, hydroclimate conditions and POM dynamics are primarily studied at the bulk level (Jeong et al., 2023). Further in-depth investigations into estuarine POM dynamics, performed at the molecular level through the use of complementary lipid biomarkers (stanols, sterols, fatty acids, and *n*-alkanes, as detailed in Chapter 1), are essential for a comprehensive understanding of the complex behaviors of POM within

estuaries. Such investigations contribute to the evaluation of estuarine functioning, particularly the role that estuaries play in regulating POM dynamics.

Based on the investigation of bacterial tetraethers and bulk geochemical parameters presented in Chapter 3, which revealed a specific zone characterized by high productivity and intense activity of heterotrophic bacteria, particularly during the low-flow season, we propose two hypotheses. Firstly, we hypothesize that phytoplankton-derived biomarkers are also prevalent in this specific estuarine zone. Secondly, we hypothesize that this preferential production is associated with land use characteristics and hydroclimate conditions. To test these hypotheses and explore how estuaries influence different types of POM under varying hydroclimate conditions and land use characteristics, we investigate the POM dynamics along the Seine Estuary using bulk analysis and complementary biomarkers. Previous POM studies in this estuary has mainly focused on bulk characterization (Etcheber et al., 2007; Savoye et al., 2003); however, more detailed analyses using complementary biomarkers are still needed. The aim of this study is to (i) investigate the spatio-temporal variations of distinct types of POM (i.e. anthropogenic POM, phytoplankton-derived POM, and plant-derived POM) along the land-sea continuum, (ii) explore the interactions between hydroclimate conditions, land use characteristics and distinct types of POM, and (iii) assess the estuarine functioning in terms of POM dynamics (i.e. how estuaries regulate different types of POM).

## 4.2. Material and methods

### 4.2.1 Study area

The Seine River basin (Figure 4-1) is more than 760 kilometers long, passing through the greater Paris region (nearly 12 million people) to the English Channel, which is notable for its high

population density, intensive industrial and agricultural activities (Flipo et al., 2021; Romero et al., 2019). Maximum water discharge is typically observed in winter (>700 $m^3/s$; Figure 4-1c), while minimum discharge is generally observed in summer (<250 $m^3/s$; Figure 4-1c). A dam at Poses (site 5; Figure 4-1a) constitutes the boundary between the Seine River and the Seine Estuary. The upper section of the estuary extends from site 5 (KP 202) to site 11 (KP 278), displaying distinct proportions of riverine POM when compared to the lower section (from site 12 to the coastal region, starting at KP 298) of the estuary, as shown in chapter 3.

The land use data across the Seine River basin was retrieved from the worldwide surface coverage product GLOBELAND30 (http://www.globallandcover.com/) with a resolution of 30 meters in 2020. Eight land use types, including urban, agricultural, forested, water body, shrubland, bareland, grassland and wetland, can be identified in the Seine River basin (Figure 4-1, a-c). To calculate the land use type proportion for each sampling site, a 1km (radius) buffer zone around each site was created using ArcGIS (10.7) software.

### 4.2.2 Sampling

From June 2019 to June 2021, water samples ($n$=156) were collected in high-flow (over 250 $m^3/s$) and low-flow (below 250 $m^3/s$) seasons across distinct land use types of the Seine River basin (Figure 4-1 and Table 4-1). Both sub-surface (ca. 1m depth) and bottom water (2.2-16m depth) samples were retrieved at 5 sites (sites 4, 6, 10, 13 and 15, Figure 4-1a and Table 4-1) using a pump into precleaned FLPE Nalgene carboys (20L). At 4 estuarine sites (sites 6, 10, 13 and 15; Figure 4-1a and Table 4-1), water samples were collected at three tide periods (low tide, mid tide, and high tide). Water samples (0.25-43L) from these sampling sites were immediately filtered by using pre-combusted Whatman GF/F 0.7 µm glass fiber filters. These filters were freeze-dried, scratched, and kept frozen (-20°C) before to recover Suspended Particulate Matter (SPM) and

analyze POM. Additional SPM samples (*n*=16; Table 4-1) used in this study were collected in 2015 and 2016 from the upper and lower section of the estuary (site 5, 7, 13, 15, 17, 18, and 19; Figure 4-1a and Table 4-1) by Thibault et al. (2019).

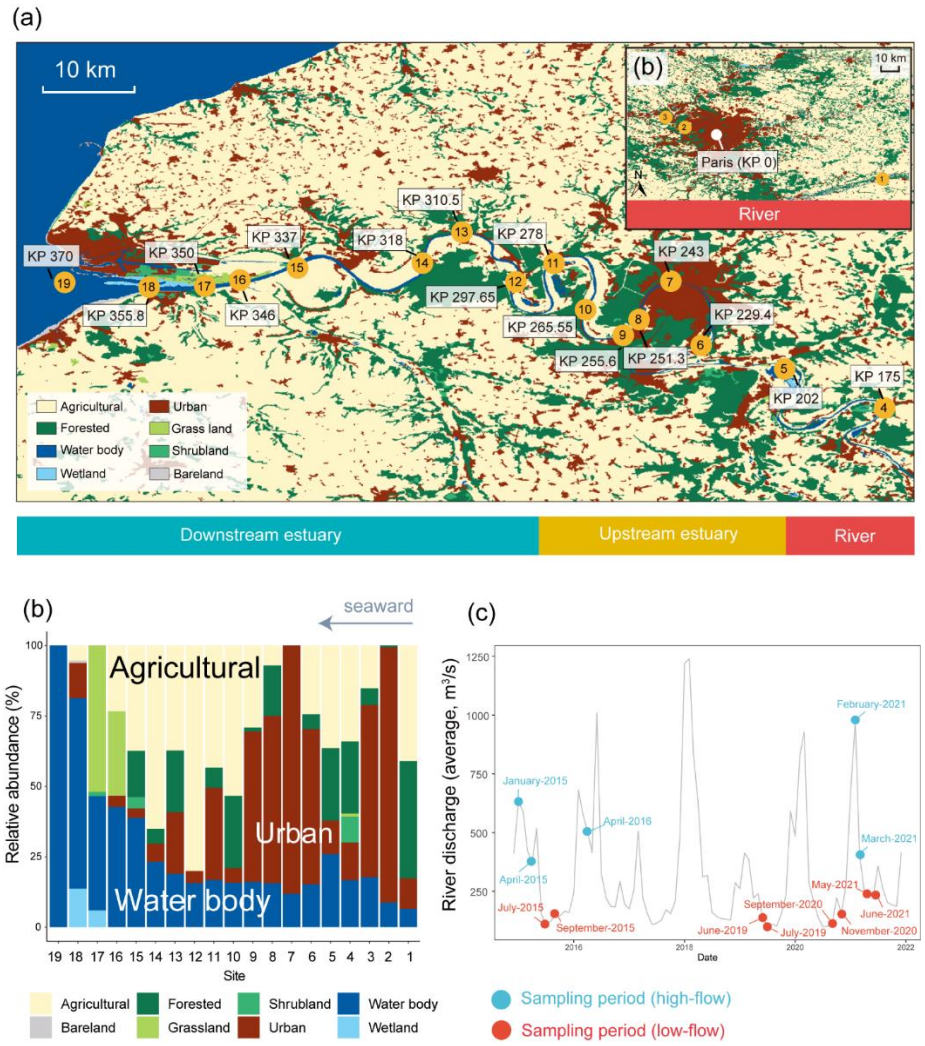



**Figure 4-1.** (a) Map showing the sampling sites (orange bullets) and land use characteristics (agricultural, urban, forested, grass land, water body, shrubland, wetland, and bareland) in the Seine Estuary and downstream part of the Seine River. (b) Map showing the sampling sites (orange bullets) in the upstream section of the Seine River. The white bullet indicates the city of Paris. (c) Relative abundances of distinct land use types along the Seine River basin. (d) Mean monthly water discharge of the Seine River measured at the Paris Austerlitz station from 2015 to 2021 (retrieved from https://www.hydro.eaufrance.fr/). The sampling period is represented by bullets with different colors, with blue bullets representing samples collected during the high-flow (>250 $m^3/s$) season and red ones representing samples collected during the low-flow (<250 $m^3/s$) season. Kilometric Point (KP) indicates the distance in kilometers from the city of Paris (KP 0).

**Table 4-1.** Sampling location

| Site | Name | Longitude (°) | Latitude (°) | KP | Zone | Date |
|---|---|---|---|---|---|---|
| 1 | Marnay sur Seine | 3.56 | 48.51 | -200 | River | 2020-11 |
| 2 | Bougival | 2.13 | 48.87 | 40 | River | 2020-11 |
| 3 | Triel sur Seine | 2 | 48.98 | 80 | River | 2020-11 |
| 4 | Les Andelys | 1.4 | 49.24 | 175 | River | 2019-6; 2019-7; 2020-9; 2021-5; 2021-6 |
| 5 | Poses | 1.24 | 49.31 | 202 | Upstream estuary | 2020-11 |
| 6 | Oissel | 1.1 | 49.34 | 229.4 | Upstream estuary | 2019-6; 2019-7; 2020-9; 2021-5; 2021-6 |
| 7 | Rouen | 1.03 | 49.43 | 243 | Upstream estuary | 2016-4 |
| 8 | Petit Couronne | 1.01 | 49.38 | 251.3 | Upstream estuary | 2020-9; 2021-2; 2021-3 |
| 9 | Grand-Couronne | 0.98 | 49.36 | 255.6 | Upstream estuary | 2019-6 |
| 10 | Val des Leux | 0.92 | 49.4 | 265.55 | Upstream estuary | 2019-6; 2019-7; 2020-9; 2021-5; 2021-6 |
| 11 | Duclair | 0.87 | 49.48 | 278 | Upstream estuary | 2020-9; 2021-2; 2021-3 |
| 12 | Heurtauville | 0.82 | 49.45 | 297.65 | Downstream estuary | 2019-6 |
| 13 | Caudebec | 0.75 | 49.52 | 310.5 | Downstream estuary | 2019-6;2019-7; 2020-9; 2021-2; 2021-3; 2021-5; 2021-6 |
| 14 | Vatteville-La-Rue | 0.67 | 49.47 | 318 | Downstream estuary | 2019-6 |
| 15 | Tancarville | 0.47 | 49.47 | 337 | Downstream estuary | 2019-6;2019-7; 2020-9; 2021-2; 2021-3; 2021-5; 2021-6 |
| 16 | Berville-Sur-Mer | 0.37 | 49.44 | 346 | Downstream estuary | 2019-6 |
| 17 | Fatouville | 0.32 | 49.44 | 350 | Downstream estuary | 2015-4; 2015-7; 2015-9; 2016-4 |
| 18 | Honfleur | 0.23 | 49.43 | 355.8 | Downstream estuary | 2019-6;2020-9;2021-2; 2021-3; 2021-5 |
| 19 | La Carosse | 0.03 | 49.48 | 370 | Downstream estuary | 2015-7; 2016-4 |

### 4.2.3 Measurement of chlorophyll a

An aliquot of water samples collected from May 2021 and June 2021 was filtered using Whatman GF/F 0.7 μm glass fiber filters. These filters (*n*=52) were kept frozen (-20° C) before chlorophyll *a* (Chl *a*) analysis. Chl *a* was extracted from filters by incubating them in 10 ml of 90% acetone for 12 hours at 4°C in the dark. Chl *a* concentrations were measured by a Turner Designs Fluorometer after two centrifugations (1700 g, 5 min), which was based on reference protocol of SNO SOMLIT (Service d'observation du Milieu Littoral) according to Strickland and Parsons (1972). Measurement of Chl *a* was performed at the Laboratoire Ecologie Fonctionnelle et Environnement (Université de Toulouse) as well as at UMR BOREA (Université de Caen Normandie). Chl *a* concentrations of the samples collected from June 2019 to March 2021 were from Chapter 3.

### 4.2.4 Elemental and isotopic analyses

Elemental and isotopic analyses of SPM (*n*=54) collected from May 2021 and June 2021 were performed following Chapter 3. Briefly, 10mL of HCl (3 M) were added to freeze-dried SPM (40mg) with magnetic stirring at room temperature for 2 hours. Samples were then rinsed by ultrapure water and centrifuged until the pH of the supernatant was neutral. Decarbonated samples were kept at −20 °C overnight and then freeze dried for one day. Subsequently, non-decarbonated and decarbonated samples (~6mg) were enclosed in the tin capsule for further measurement. Total Organic Carbon content (TOC) and stable carbon isotopic composition ($\delta^{13}C$) of decarbonated samples were determined using an elemental analyzer coupled with an isotope ratio mass spectrometer (Thermo Fisher Scientific Delta V Advantage) at the ALYSES platform (Sorbonne University / IRD). Given that acidification could impact on the N contents (Ryba and Burgess, 2002), non-decarbonated samples are used for Total Nitrogen (TN) and nitrogen isotope ($\delta^{15}N$)

measurement using the same instrument as for decarbonated samples. The isotopic composition was expressed as relative difference between isotopic ratios in samples and standards (Vienna Pee Dee Belemnite (VPDB) for C and atmospheric $N_2$ for N). Additional elemental and isotopic analyses of SPM collected from January 2015 to March 2021 were retrieved from Chapter 3 and Thibault et al. (2019).

### 4.2.5 Lipid extraction and analyses

SPM (~150 mg, *n*=172) were extracted ultrasonically with dichloromethane (DCM): methanol (MeOH) (5:1, v/v, 3×). The total lipid extracts were then separated into apolar and polar fractions on an activated silica gel column, using 30 mL of heptane, 30 mL of heptane:DCM (1/4, v/v), and 30 mL of DCM/MeOH (1/1, v/v) as eluents. *n*-alkanes are contained in the first (apolar) fraction, whereas fatty acids, sterols and stanols are contained in the third (polar) fraction. An aliquot (6%) of the polar fraction containing sterols and stanols was dried, re-dissolved in DCM and derivatized with a mixture of N,O-bis-(trimethylsilyl)trifluoroacetamide and trimethylchlorosilane (BSTFA + TMCS, 99/1, v/v) at 70 °C for 1h with 5α-cholestane added as the internal standard. An aliquot (40%) of the apolar fraction containing *n*-alkanes was dried and re-dissolved in heptane after addition of *n*-tetracosane-d50 as an internal standard.

Sterols, stanols and fatty acids were analyzed by GC-MS using a Thermo Scientific Trace 1310 gas chromatograph fitted with a Rxi® -5Sil MS column (60 m × 250 µm × 0.25 µm; RESTEK) interfaced to a ISQ 7000 single quadrupole mass spectrometer. 1 µL of the derivatized polar fraction was injected in split mode (10:1) with He as the carrier gas at 2 mL/min. The oven temperature started at 70 °C (held 1 min), increased to 130 °C at 20 °C/min, then to 320 °C (held 25 min) at 4 °C/min. The mass spectrometer was simultaneously operated in full scan mode (m/z 35-700) and selective ion monitoring (SIM) mode *(m/z* 75 for fatty acids, 129 for sterols, 215 for

stanols and 217 for the internal standard), with the transfer line temperature at 320 °C and EI voltage at 45 eV. Sterols, stanols and fatty acids were identified based on their retention time and mass spectra. Data were processed with Chromeleon software.

*n*-alkanes were analyzed with the same instrument and GC capillary column as the polar organic compounds. The oven temperature program was initially at 50 °C and increased to 320 °C (held 30 min) at 4°C/min. 1 µl of the apolar fractions was injected on the same column as the polar fractions in splitless mode. Carrier gas (He) was at a constant flow rate (2 mL/min). The apolar fraction was analyzed in selected ion monitoring (SIM) mode (m/z 57 for n-alkanes and m/z 66 for the internal standard) and in full scan mode (m/z 35–700) simultaneously. The transfer line temperature was set at 320 °C and EI voltage at 45 eV. *n*-alkanes were identified based on their retention time and mass spectra. Chromeleon software was used to process the data.

### 4.2.6 Calculation of molecular proxies

The molecular proxies based on sterols, stanols, fatty acids and *n*-alkanes used in this chapter are summarized in Chapter 1 (Table 1-1). Based on replicate injections ($n$=6), the analytical error was 0.002 for Coprostanol/(Coprostanol+Cholestanol), 0.13 for Brassicasterol (%), 0.01 for $C_{16:1}/C_{16:0}$, 0.15 for CPI, 0.33 for ACL and 0.01 for $P_{aq}$.

### 4.2.7 Statistical analyses

All statistical analyses in this study were performed using R (4.2.1). The Spearman's correlation was utilized to investigate correlations among distinct variables (bulk and molecular proxies, water discharge and land use types). For two independent group comparisons, the unpaired two-samples Wilcoxon test (also known as Mann-Whitney test or Wilcoxon rank sum test) was

used. Distinct levels of significance were distinguished: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. ns (not significant): p-value > 0.05.

Principal Component Analysis (PCA) was performed on the bulk and molecular proxies, land use characteristics, and water discharge using the R packages factoextra and FactoMineR. Different groups (samples from river, upstream estuary and downstream estuary) were highlighted by 95% concentration ellipses.

Spatio-temporal variations of bulk and molecular proxies were assessed based on a locally estimated scatterplot smoothing method (LOESS), which enables the detection of nonlinear data trends and buffers the impact of aberrant data and outliers.

## 4.3. Results

In Chapter 3, samples were categorized into river (before KP 202), upstream (KP 202 to KP 278), and downstream (starting at KP 298) estuary groups, with each of these categories demonstrating distinct proportions of riverine POM (i.e. POM coming from upstream). We start by employing the same categorization scheme to investigate the differences in POM composition along the estuary based on different families of biomarkers (stanols, sterols, fatty acids, and *n*-alkanes).

### 4.3.1 Distribution of stanols

Six stanols (coprostanol, epi-coprostanol, cholestanol, 24-ethyl coprostanol, campestanol, and stigmastanol) were identified in the samples collected across the Seine River basin (Figure 4-2). Stanol distribution was dominated by coprostanol in river and upstream estuary, averaging

40.93 % and 31.72 % of the total stanols, respectively (Figure 4-2). The stigmastanol was present in higher proportions in the downstream estuary (29.17 % on average) compared with river (8.99 % on average) and upstream estuary (21.02 % on average) (Figure 4-2).

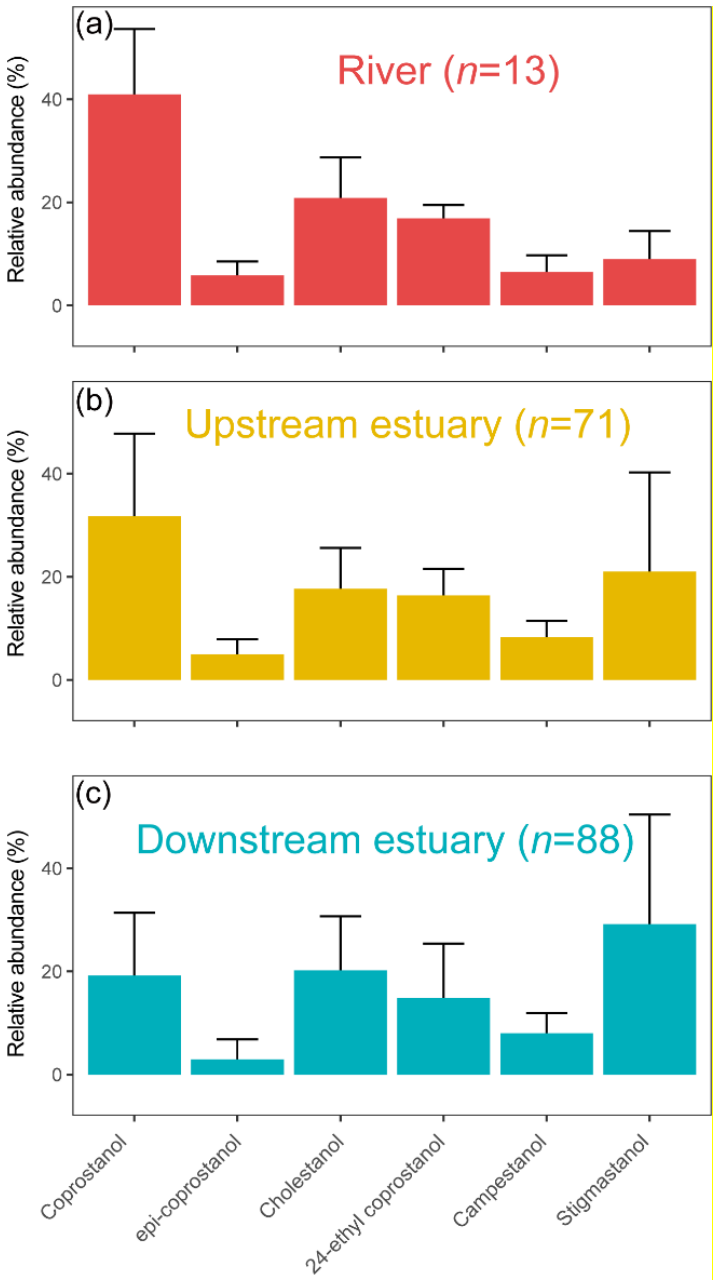



**Figure 4-2.** Relative abundances of the individual stanols for (a) river, (b) upstream estuary, and (c) downstream estuary samples.

### 4.3.2 Distribution of sterols

In the Seine River basin, five sterols (cholesterol, brassicasterol, campesterol, stigmasterol, and sitosterol) were identified (Figure 4-2). In rivers, sterols are mainly dominated by cholesterol, whereas in the upstream and downstream estuaries, they are dominated by sitosterol (Figure 4-3). Brassicasterol is higher in upstream estuary (9.47 ± 7.04 %) compared with river (8.60 ± 3.63 %) and downstream estuary (7.83 ± 7.68 %) samples.

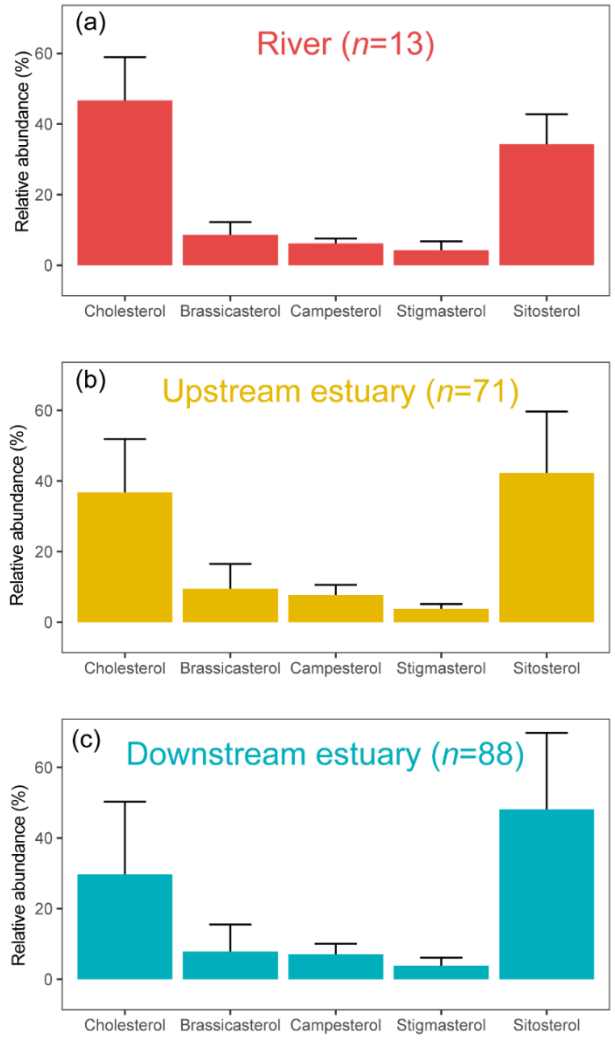



**Figure 4-3.** Relative abundances of the individual sterols for (a) river, (b) upstream estuary, and (c) downstream estuary samples.

### 4.3.3 Distribution of fatty acids

In Seine River basin, fatty acids were dominated by $C_{16:0}$ and $C_{18:0}$ (Figure 4-4). $C_{16:1}$ is higher in downstream estuary (5.15 ± 7.68 %) compared with upstream estuary (5.09 ± 6.11 %) and river (2.76 ± 4.45 %).

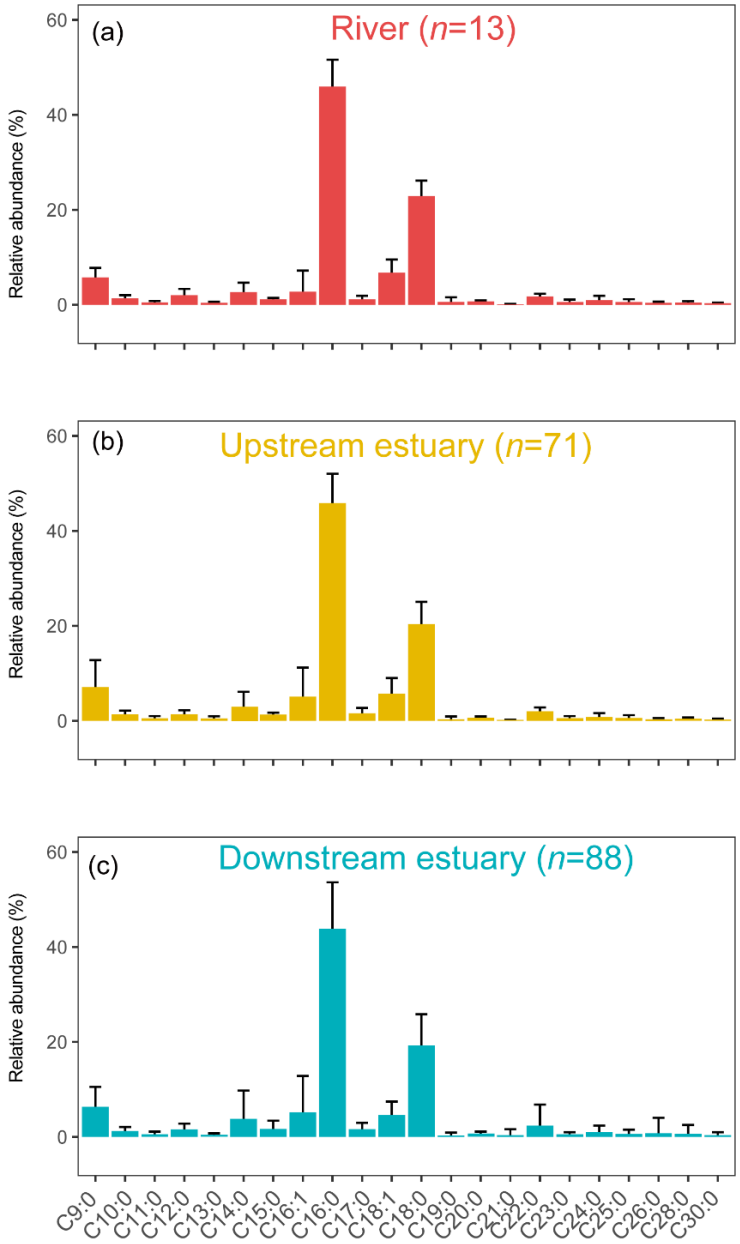



**Figure 4-4.** Relative abundances of the fatty acids for (a) river, (b) upstream estuary, and (c) downstream estuary samples.

### 4.3.4 Distribution of n-alkanes

The *n*-alkanes identified in the Seine River basin have chain lengths ranging from 16 to 35 (Figure 4-5). $C_{29}$ represents the dominating *n*-alkanes throughout the river basin (12.08 ± 6.83 % in river; 12.15 ± 4.46 % in upstream estuary; 12.07 ± 5.50 % in downstream estuary).

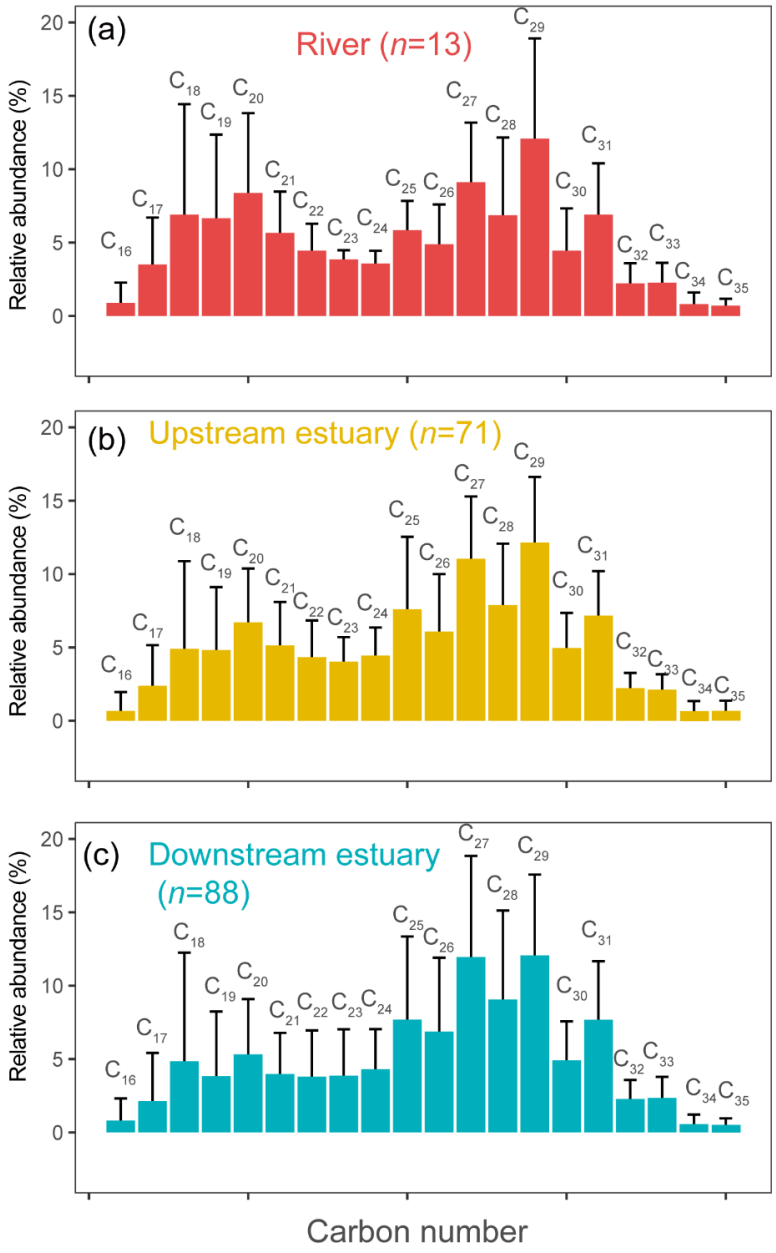



**Figure 4-5.** Relative abundances of the *n*-alkanes for (a) river, (b) upstream estuary, and (c) downstream estuary samples.

## 4.4. Discussion

### 4.4.1 Spatio-temporal variations of anthropogenic POM

As coprostanol can be produced by microbial reduction of cholesterol in the digestive system of humans and higher vertebrates (Venkatesan and Mirsadeghi, 1992), it has been widely used as an anthropogenic biomarker for fecal (sewage) contamination in aquatic systems (Grimalt et al., 1990; He et al., 2018; Rada et al., 2016). Higher relative abundances of coprostanol in the Seine River (40.9 ± 12.7 %) and upstream estuary (31.7 ± 16.0 %) compared with downstream estuary (19.18 ± 12.18 %) suggest higher sewage contributions in upstream regions. The highest coprostanol concentration (33.1 μg/g dry weight) is observed in a region with high urban land use (Site 7; Figure 4-1) during the high-flow season. Such a value is much higher than the threshold of sewage contamination (0.7 μg/g dry weight) (Rada et al., 2016), suggesting that the Seine Estuary is especially subject to sewage contamination in this area with high portions of urban land use.

Due to the potential *in situ* production of coprostanol in anoxic sediments, Grimalt et al. (1990) proposed that there are limits in using coprostanol concentration as a robust indicator to trace sewage contamination, and presented the coprostanol/(coprostanol + cholestanol) ratio to solve this problem. The spatio-temporal variations of sewage contamination in the Seine Estuary is further assessed by this diagnostic ratio, with higher than 0.7 as the criteria for sewage inputs (Grimalt et al., 1990). This ratio shows decreasing trends from upstream to downstream in high-flow and low-flow conditions, suggesting a dilution of sewage contamination during the mixing of riverine and marine waters regardless of the seasonality (Figure 4-6a). Indeed, the highly urbanized area in the upstream estuary (Figure 4-1, a-b) is more likely to experience high levels of sewage contamination because it is closer to possible pollution sources such as industrial and urban wastewater (Touron et al., 2007). The contaminants from the urban region can be diluted as they

move downstream and mix with the seawater, where there is less urban land use and more water body (Figure 4-1b). This suggests that estuaries act as effective natural filters and buffers for anthropogenic contaminants (Celis-Hernandez et al., 2021). The purification capacity of estuaries has been observed in many other estuaries worldwide, such as the Xiaoqing River-Laizhou Bay system in China, where the seaward decreasing trend of coprostanol / (coprostanol + cholestanol) was also noticed (He et al., 2018).

In addition to a clear spatial variation, coprostanol / (coprostanol + cholestanol) also shows seasonal variabilities. This ratio is significantly higher in the high-flow season (0.7 ± 0.1) than in the low-flow season (0.5± 0.2) ($p$<0.05, Wilcoxon test; Figure 4-6c). Greater sewage contamination in the water column at high flows could be explained by different hydrodynamic conditions. During the high-flow season, the volume of water may exceed the capacity of the sewage treatment plants, which may lead to the release of untreated sewage into the water body (Al Aukidy and Verlicchi, 2017). This could explain the seasonal variations of the coprostanol / (coprostanol + cholestanol) ratio in the Seine Estuary.

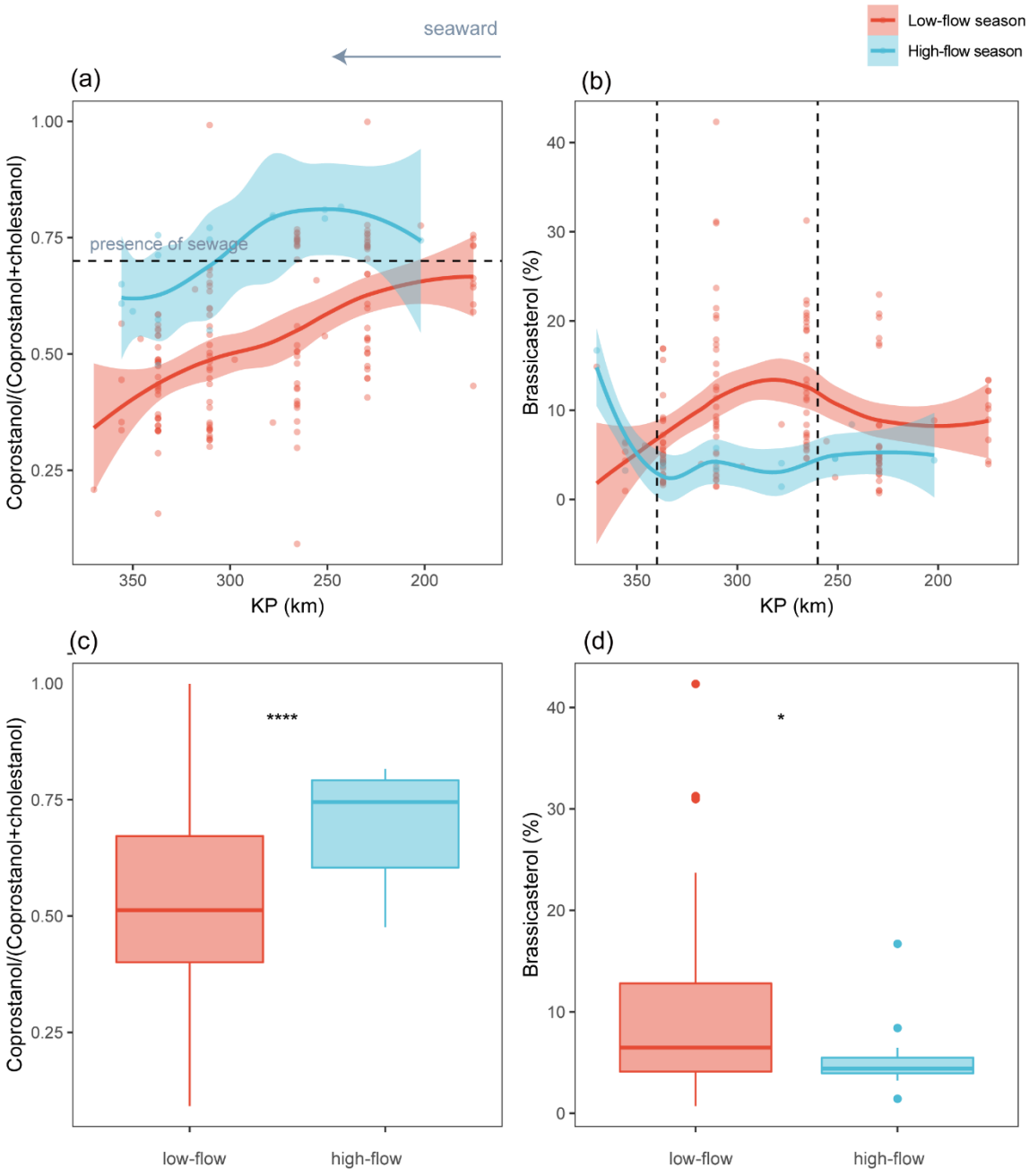



**Figure 4-6** Spatio-temporal variations of proxies based on sterols and stanols, including (a) Coprostanol/(Coprostanol+Cholestanol) and (b) Brassicasterol (%). Kilometric Point (KP) represents the distance in kilometers from the city of Paris (KP 0). The trends showing proxy variations from site 4 (KP 175) to site 19 (KP 370) were based on locally estimated scatterplot smoothing (LOESS), with the shaded area representing 95% confidence intervals. Box plots comparing the indices based on sterols and stanols, including (c) Coprostanol/(Coprostanol+Cholestanol) and (d) Brassicasterol (%) between low-flow (<250 $m^3/s$ - red) and high-flow (>250 $m^3/s$ - blue) seasons. Statistical testing was performed by using a Wilcoxon test (*$p < 0.05$ and ****$p < 0.0001$).

### 4.4.2 Spatio-temporal variations of phytoplankton-derived POM

Sterols are produced by different types of algae and plants (Rontani et al., 2018; Saeidnia et al., 2014; Volkman, 1986; Xiao et al., 2015). For example, campesterol, stigmasterol, and sitosterol are typical phytosterols, and they have been found abundantly in terrestrial plants (Saeidnia et al., 2014). In addition, cholesterol is considered as a non-specific compound, which is found in various organisms, including animals, plants, and phytoplankton (Volkman, 1986). Compared with cholesterol that have multiple sources, brassicasterol is often a predominant sterol of diatoms, a group of phytoplankton commonly found in marine and freshwater environments (Gladu et al., 1990; Rampen et al., 2010), although it has also been attributed to many groups of marine phytoplankton (Volkman, 2003).

Spatially, the relative abundances of brassicasterol reach their high levels at the interface between the upstream and downstream estuary ($260 < KP < 340$) during the low-flow season (Figure 4-6b). Brassicasterol relative abundances are significantly higher at low flows than at high flows ($p<0.05$, Wilcoxon test; Figure 4-6d), which implies that diatom-derived POM is particularly accumulated during that period. The low flow period primarily occurs during spring and summer (Figure 4-1c), which could provide favorable conditions for the growth of phytoplankton.

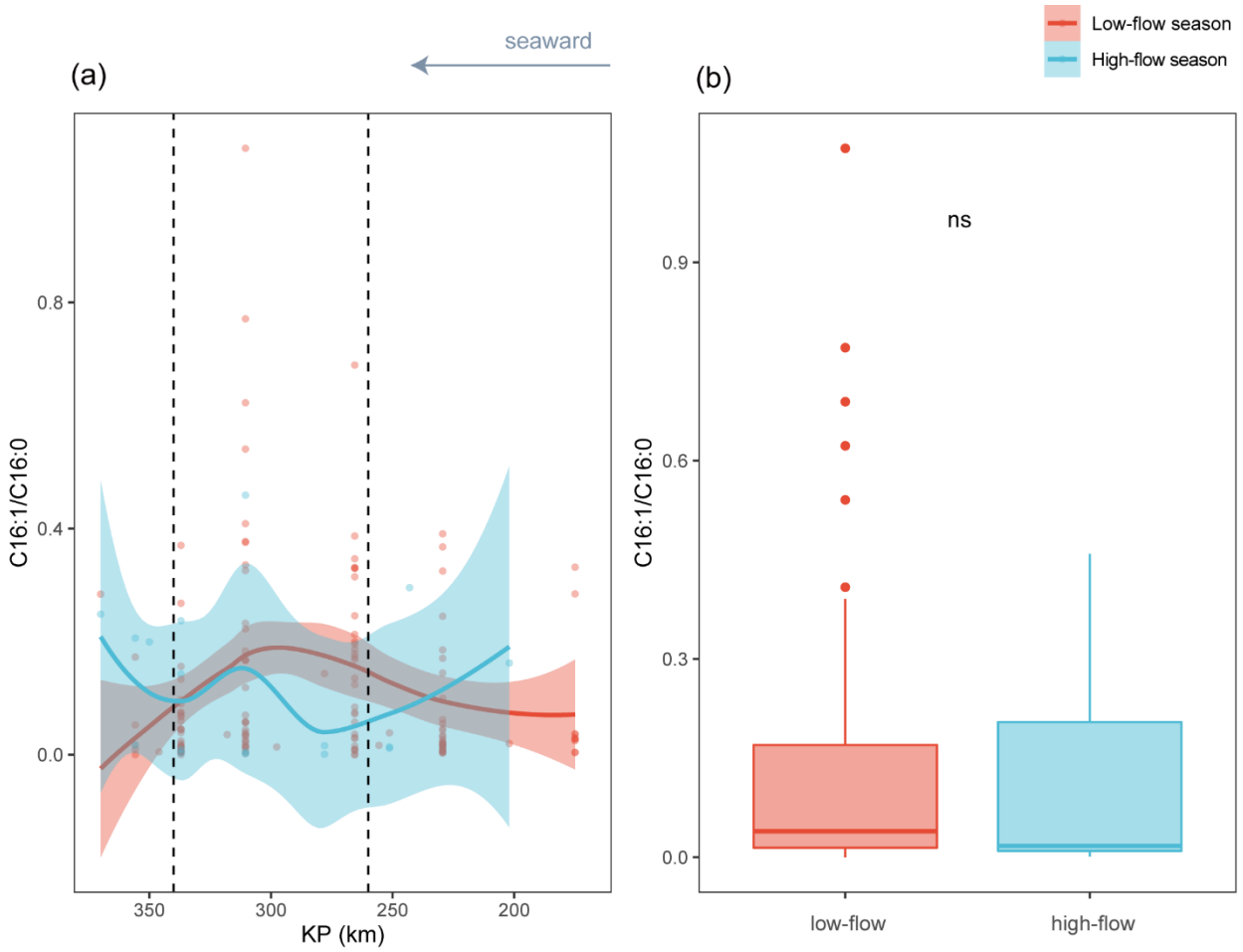



**Figure 4-7.** (a) Spatio-temporal variations of C16:1/C16:0. (b) Box plots comparing the C16:1/C16:0 between low-flow (<250 $m^3/s$ - red) and high-flow (>250 $m^3/s$ - blue) seasons. Statistical testing was performed by using a Wilcoxon test (ns, not significant, $p$ >0.05).

The contribution of diatoms can also be identified by specific fatty acids (Figure 4-7). As diatoms are characterized by high values of $C_{16:1}/C_{16:0}$, this ratio has been used as a general diatom biomarker (Budge et al., 2001; Claustre et al., 1989). In the Seine Estuary, no significant difference in the $C_{16:1}/C_{16:0}$ ratio is observed between high flows and low flows (Figure 4-7c). This lack of difference could be explained by the presence of multiple sources of fatty acids, particularly $C_{16:0}$, which might dilute the impact of seasonal variations. However, the spatial distribution of this ratio and the relative abundance of brassicasterol seems to be linked, both exhibiting a peak at the interface (260 < KP < 340) between the upstream and downstream areas of the estuary when flow

levels are low (Figure 4-6, 4-7). Similar variations are observed in the bulk parameters (TOC, TN and $\delta^{15}N$) and a phytoplankton biomass proxy (Chl *a*) in this region (Supplementary Figure 4-1). The peak for these parameters observed at low flow seems to be shifted about 10 km downstream at high flow (Figure 4-6, 4-7, and Supplementary Figure 4-1). This shift could be attributed to the flushing of phytoplankton biomass further downstream by the increased water discharge.

The fact that all the above-mentioned parameters peak ($260 < KP < 340$) during the low-flow season could be explained by various processes. This zone ($260 < KP < 340$) represents an agriculturally impacted area, which is characterized by high proportions of agricultural land use (Sites 9-15; Figure 4-1b). Intense agricultural activities in the Seine River basin may release significant amounts of organic fertilisers, manure, as well as urban and industrial wastewater into waters (Billen et al., 2021; Romero et al., 2022). Nitrite derived from these sources could have relatively higher $\delta^{15}N$ values (10-25‰) (Andrisoa et al., 2019; Leavitt et al., 2006). During the low-flow season, the residence time of the water masses would increase, which may extend the nutrient retention (Li et al., 2021; Romero et al., 2019). Hence, the nitrate with elevated $\delta^{15}N$ values (Supplementary Figure 4-1) can be extensively assimilated by phytoplankton, further triggering phytoplankton blooms. Indeed, the development of phytoplankton blooms in this region is supported by elevated levels of phytoplankton biomass (chl *a*, Supplementary Figure 4-1d), TOC (Supplementary Figure 4-1a), TN (Supplementary Figure 4-1b), and phytoplankton biomarkers (brassicasterol and $C_{16:1}/C_{16:0}$; Figures 4-6, 4-7). During downstream transport, these parameters decrease, which could be attributed to the dilution with seawater. This dilution effect is more pronounced when the water discharge is low, during which the seawater mixes well with freshwater (Kolb et al., 2022; Ralston and Geyer, 2019).

On the other hand, the phytoplankton-derived POM (based on brassicasterol) remains constant throughout most areas of the estuary during the high-flow period, with a significant rise

in the coastal region (Figure 4-6b). Under high-flow conditions, nutrients derived from agricultural activities can be effectively transported into downstream waters (Xia et al., 2020). As the nutrient-rich river water flows downstream and reaches the coastal region, it brings large amounts of nutrients, potentially fueling phytoplankton growth in the coastal waters.

### 4.4.3 Spatio-temporal variations of plant-derived POM

Generally, middle-chain *n*-alkanes ($C_{20}$-$C_{25}$) are enriched in aquatic plants (submerged and floating aquatic macrophytes), whereas long-chain *n*-alkanes (C>25) with a strong odd-to-even carbon preference are predominant in terrestrial higher plants (Cranwell, 1984; Ficken et al., 2000; Silva et al., 2012). Hence, plant-derived POM could be identified based on *n*-alkane distributions (Derrien et al., 2017). Under low-flow conditions, the average values of ACL are >25, suggesting a predominance of long chain *n*-alkanes at low flows. ACL shows limited spatial variations (Figure 4-8a), suggesting the ability of plant biomarkers to transport for long distances. This could be attributed to potential associations with clay particles (Keil et al., 1997; Yedema et al., 2023). Furthermore, the restricted spatial variations of ACL could potentially be attributed to the consistent plant inputs all along the estuary. During the high-flow season, substantial fluctuations in ACL are observed (Figure 4-8a), implying dynamic changes in the sources and/or transformations of *n*-alkanes. This decrease can be explained by higher inputs of aquatic-plant derived *n*-alkanes as reflected by high levels of $P_{aq}$.

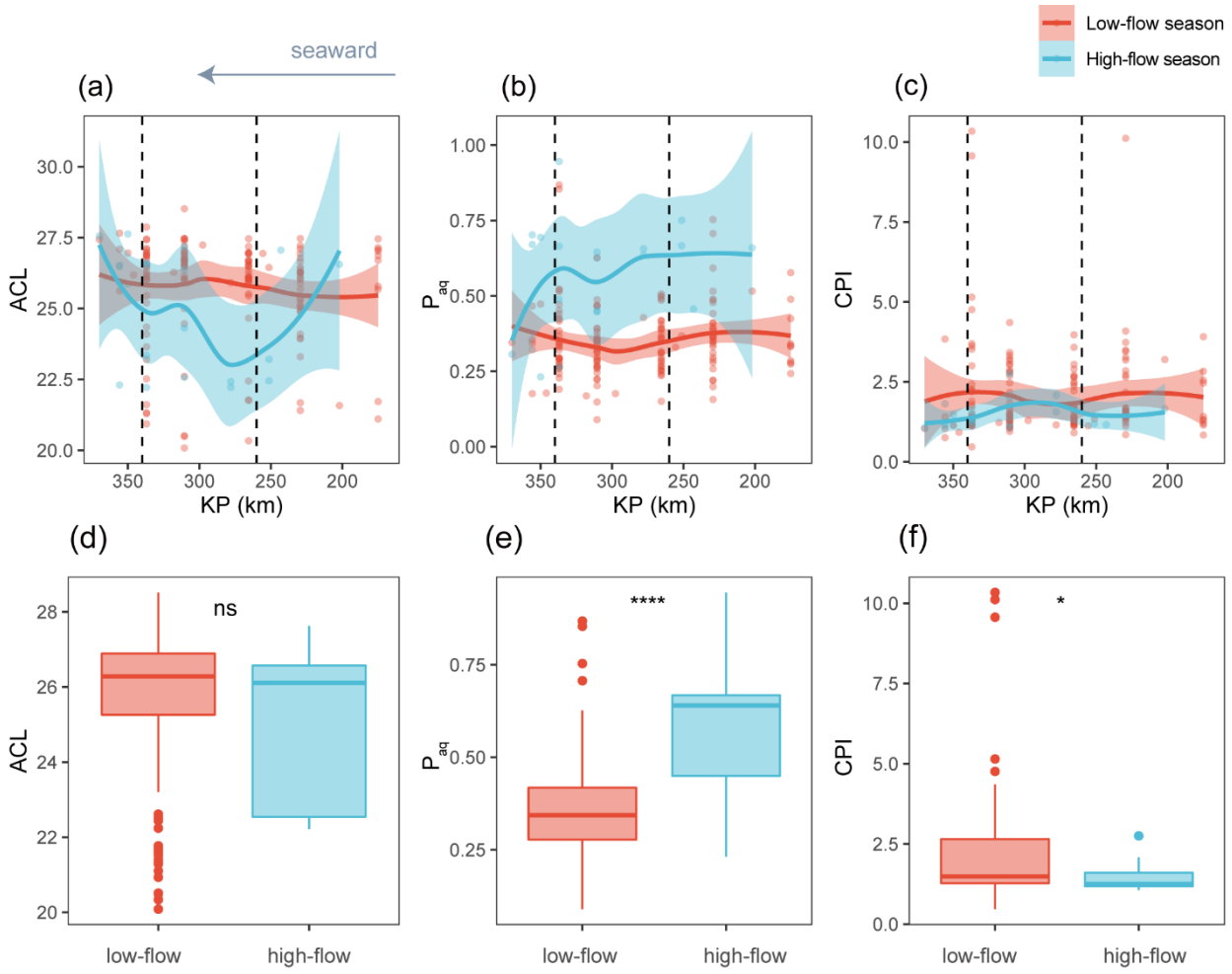



**Figure 4-8.** Spatio-temporal variations of proxies based on *n*-alkanes, including (a) ACL, (b) $P_{aq}$ and (c) CPI. Kilometric Point (KP) represents the distance in kilometers from the city of Paris (KP 0). The trends showing proxy variations from site 4 (KP 175) to site 19 (KP 370) were based on locally estimated scatterplot smoothing (LOESS), with the shaded area representing 95% confidence intervals. Box plots comparing the indices based on *n*-alkanes, including (d) ACL, (e) $P_{aq}$, and (f) CPI between low-flow (<250 $m^3/s$ - red) and high-flow (>250 $m^3/s$ - blue) seasons. Statistical testing was performed by using a Wilcoxon test (*$p < 0.05$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

To differentiate distinct types of plants (terrestrial or aquatic), the aquatic proxy ($P_{aq}$) was proposed (Ficken et al., 2000; Sikes et al., 2009), based on the relative proportion of two middle-chain *n*-alkanes ($C_{23}$ and $C_{25}$) to two long-chain *n*-alkanes ($C_{29}$ and $C_{31}$). Specifically, $0.01<P_{aq}<0.25$ reflects a predominance of *n*-alkanes derived from terrestrial plants, 0.4–0.6 from emergent aquatic plants and >0.6 from submerged aquatic plants (Ficken et al., 2000; Sikes et al.,

2009). In the Seine Estuary, $P_{aq}$ is significantly higher in the high-flow season (0.6±0.2) compared with the low-flow season (0.4±0.1) ($p$<0.05, Wilcoxon test; Fig. 3e). Hence, *n*-alkanes are mostly contributed by aquatic plants at high flows and a mix of terrestrial and aquatic plants at low flows. Greater inputs from aquatic plants at high flows were also observed in other estuaries, such as the Danzhou Bay (China) (Chu et al., 2020). It should be noted that, other processes, such as preferential degradation of middle-chain alkanes, could also explain the decrease of $P_{aq}$ values.

Microbial degradation processes can notably be evaluated by Carbon Preference Index (CPI) (Derrien et al., 2017; Martens et al., 2023). Generally, higher CPI values (>5) are characteristic of well-preserved (non-degraded) long chain *n*-alkanes, whereas lower CPI values suggest intense degradation of *n*-alkanes (Bray and Evans, 1961; Cranwell, 1981; Meyers and Ishiwatari, 1993). Low CPI values (<5) are typically observed in the Seine River basin, which indicates that the plant-derived POM is subject to substantial microbial degradation across the river basin. Enhanced microbial degradation activities (as evidenced by a slight decrease in CPI) are particularly observed during the low-flow season at the interface (260 < KP < 340) between upstream and downstream estuaries (Figure 4-8c), where a phytoplankton bloom occurs, as explained in section 4.4.2. Indeed, high heterotrophic bacterial activities have been found, as evidenced by the abundant heterotrophic bacterial-derived lipids in this zone under low-flow conditions shown in chapter 3. These heterotrophic microbes may be fueled by the labile organic matter released from the phytoplankton (Bachi et al., 2023). Active heterotrophs can further enhance the remineralization of the plant-derived organic matter (Halvorson et al., 2019). This process is termed as priming effect (Bianchi, 2011; Guenet et al., 2010), which refers to the increased decomposition rate of recalcitrant organic matter by the addition of labile organic matter. The priming effect may be a widespread biogeochemical process across a number of settings ranging from freshwater to the ocean (Bianchi, 2011; Bianchi et al., 2015; Guenet et al., 2010;

Ward et al., 2016). In the Seine Estuary, it is likely that middle chain alkanes are preferentially degraded, considering the slight decrease of $P_{aq}$ in this zone (260 < KP < 340) under low-flow conditions (Figure 4-8). Hence, slightly lower values of $P_{aq}$, CPI, and high levels of phytoplankton biomarkers (brassicasterol and $C_{16:1}/C_{16:0}$) in this zone could potentially indicate the occurrence of the priming effect. Furthermore, as detailed in 4.4.2, higher levels of brassicasterol are observed in coastal waters at high flows (Figure 4-6b). During the high-flow period, the most downstream part of the estuary (KP>340) is characterized by low $P_{aq}$ levels (Figure 4-8b), which most likely indicates a priming effect. Overall, our results show that priming effect possibly occurs within the agriculturally impacted area during the low-flow period, whereas at high flows, the zone for this process moves further downstream (Figures 4-6, 4-7, 4-8).

### 4.4.4 POM dynamics associated with hydroclimate conditions and land use characteristics

In the Seine Estuary, anthropogenic POM, phytoplankton-derived POM, and plant-derived POM show distinct spatial and temporal patterns (Figure 4-6, 4-7, 4-8), which are potentially linked to hydroclimate conditions and land use characteristics. Phytoplankton-derived POM especially accumulated in an agriculturally impacted region at the interface between upstream and downstream estuary (260<KP<340) during low flows, as reflected by high levels of molecular proxies (brassicasterol and $C_{16:1}/C_{16:0}$; Figure 4-6, 4-7) as well as bulk parameters (TN, $\delta^{15}N$, chl *a*, and TOC; Supplementary Figure 4-1). Such characteristics could further divide the estuary into three distinct zones: Zone I (KP<260), Zone II (260<KP<340), and Zone III (KP>340). These zones are distinguished by contrasting land use characteristics and POM dynamics as shown by principal component analysis (PCA) (Figure 4-9).

The first principal component (PC1) explains 20.7% of the variance, with strong negative loadings for agricultural land use and several phytoplankton-related proxies (i.e. brassicasterol and

Chl *a*) (Figure 4-9). Hence, PC1 effectively separates samples based on distinct levels of phytoplankton biomarkers and agricultural land use (Figure 4-9). Zone II with negative values of PC1 indicates a specific region with accumulation of phytoplankton biomass as well as a potential priming effect under low-flow conditions, characterized by high portions of agricultural land use, low $P_{aq}$, low CPI, low discharge, and high levels of phytoplankton biomarkers (Figure 4-9). This highlights the potential influence of agricultural activities on the phytoplankton blooms and related biogeochemical processes within Zone II especially at low flows. Indeed, increased agricultural activities are often associated with higher nutrient inputs and subsequent blooms of phytoplankton (Michael Beman et al., 2005). This is consistent with positive correlations observed between agricultural land use and phytoplankton biomarkers in the Seine Estuary (Supplementary Figure 4-3). On the other hand, the second principal component (PC2) explains 14.4 % of the variance, with significant negative loadings for urban land use and sewage biomarker, and water discharge (Figure 4-9). This indicates that PC2 mainly separates samples with different levels of sewage contamination, hydrodynamic conditions, and urban land use. Samples with negative PC2 values are mainly from Zone I with higher portions of urban land use (Figure 4-9). This implies that Zone I is characterized by intense urban land use and substantial contributions from anthropogenic (sewage-derived) POM. Additionally, Zone III separates well with Zone I in the PCA biplot, with higher portions of water body and enriched $\delta^{13}C$ (Figure 4-9). This suggests that Zone III is characterized by high proportions of water body, with low contributions from anthropogenic (sewage) POM.

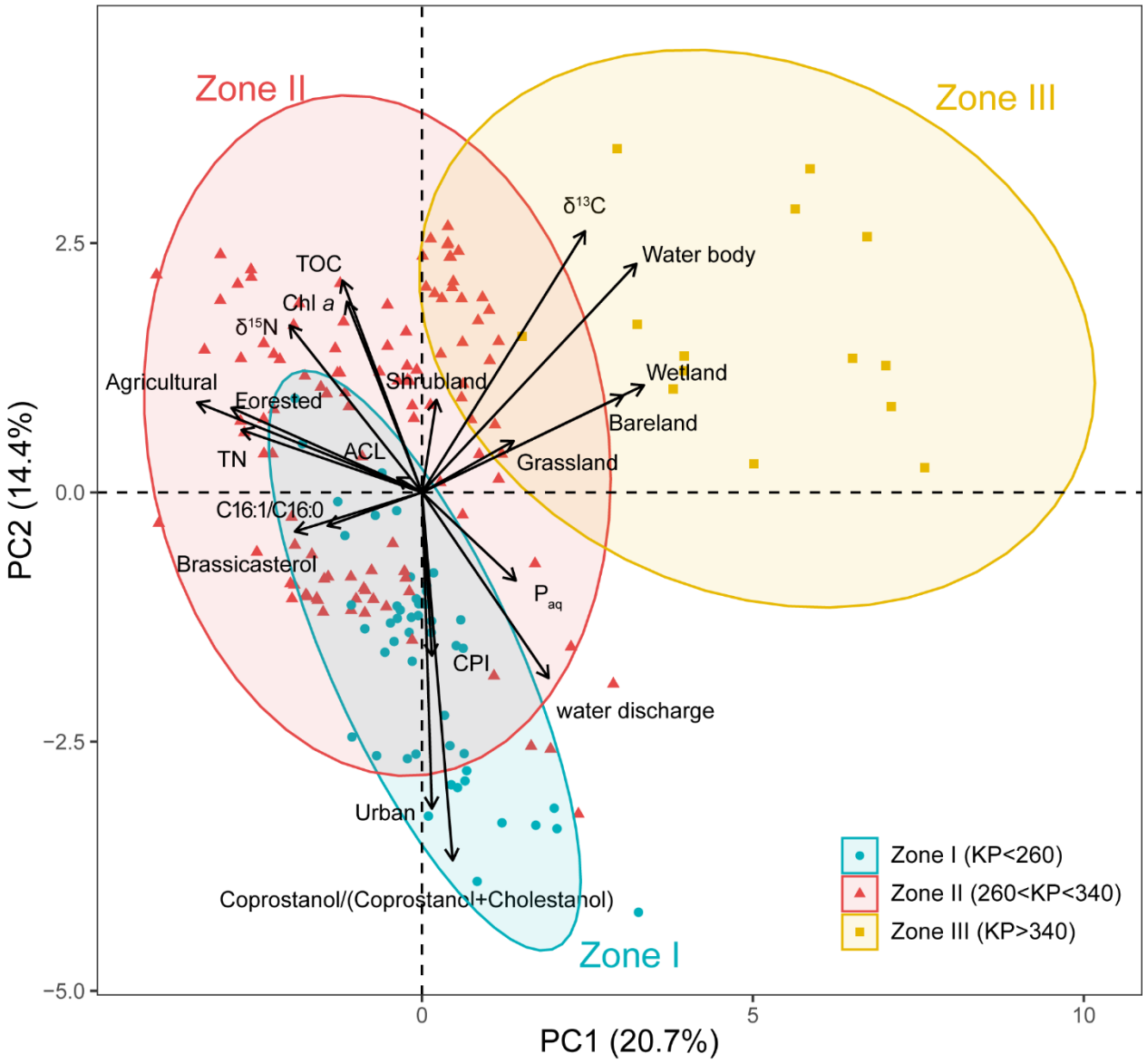



**Figure 4-9.** PCA analysis of distinct (bulk and molecular) proxies, water discharge and land use types. Samples collected in different zones were highlighted with 95% concentration ellipses.

Overall, the PCA results (Figure 4-9) and variations of different types of POM (Figure 4-6, 4-7, 4-8, and Supplementary Figure 4-1) in the Seine Estuary imply that distinct pools of POM are subjected to contrasting dynamics and transformations, which is closely linked to hydroclimate conditions and land use characteristics. Estuarine zonation is identified considering POM dynamics, natural and anthropogenic factors. Specifically, Zone I has high portions of urban land

use and significant contributions from anthropogenic (sewage-derived) POM in both high-flow and low-flow conditions. Zone II is mostly characterized by high portions of agricultural land use, with phytoplankton bloom and potential priming effect occurring in the low-flow season. Zone III is characterized by high portions of water body and is representative of costal environments, with low anthropogenic (sewage-derived) POM contributions. Phytoplankton bloom and potential priming effect occur in this zone in the high-flow season. Hence, the biogeochemical processes within these zones can be categorized into low-flow and high-flow scenarios (Figure 4-10). Estuarine functioning in terms of POM dynamics is further assessed.

During the low-flow scenario, reduced water discharge leads to an increase in water residence time. This prolonged residence time enhances retention of nutrients derived from agricultural activities in Zone II. Estuaries thus play a crucial role as reactors, facilitating biogeochemical uptake of nutrients, thereby promoting phytoplankton growth and potentially triggering priming effect. Additionally, anthropogenic POM (sewage contamination) derived from urban area in Zone I decreases gradually along the estuary. This indicates that estuaries serve as natural filters/buffers, diluting sewage contaminants in downstream region (Zone III).

During the high-flow scenario, the increased water discharge effectively flushes nutrient-rich waters into the coastal region (Zone III), leading to phytoplankton blooms and possibly triggering priming effect in this area. Estuaries act as biogeochemical reactors particularly in this zone at high flows. Moreover, sewage contaminations derived from Zone I become particularly pronounced during high-flow scenarios. In response, estuaries play an important role as effective filters/buffers, mitigating sewage contamination levels in downstream areas (Zone III).

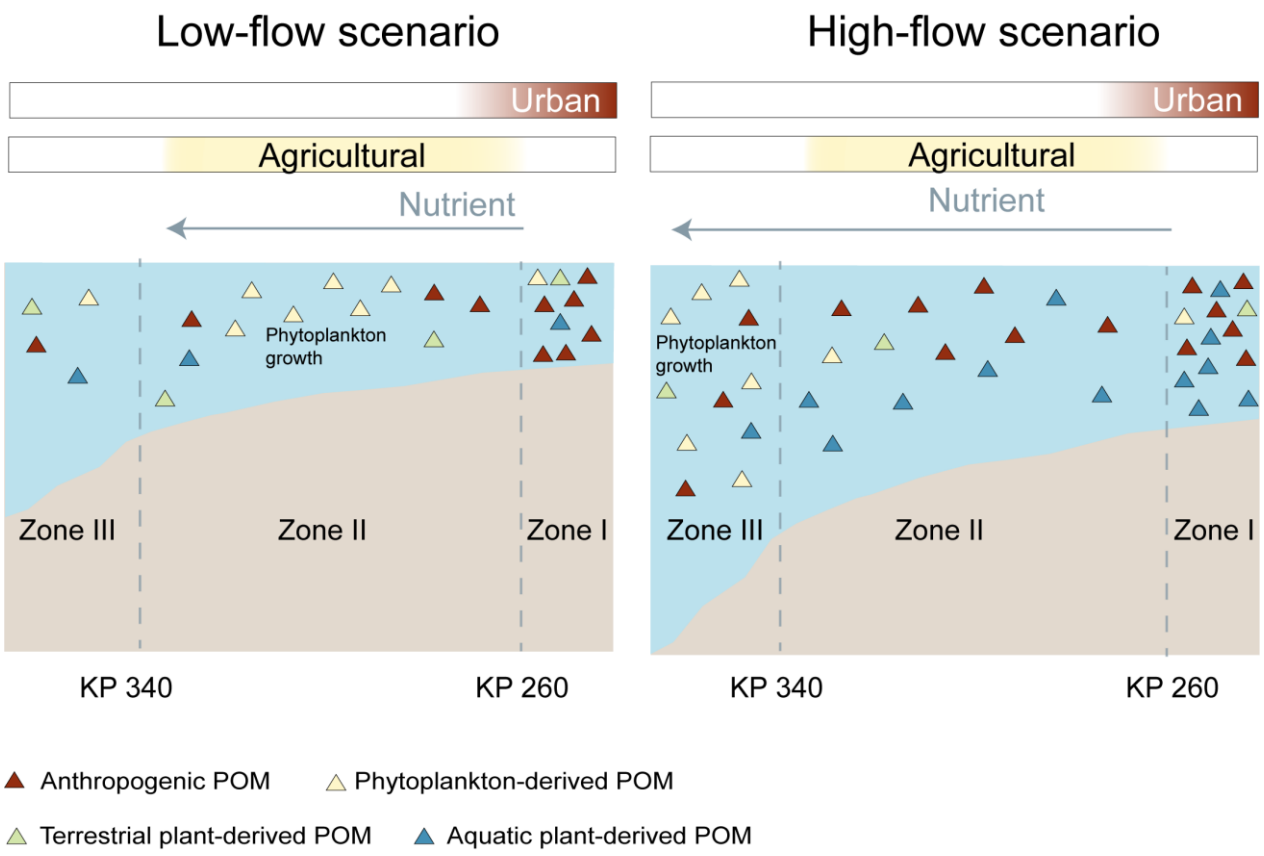


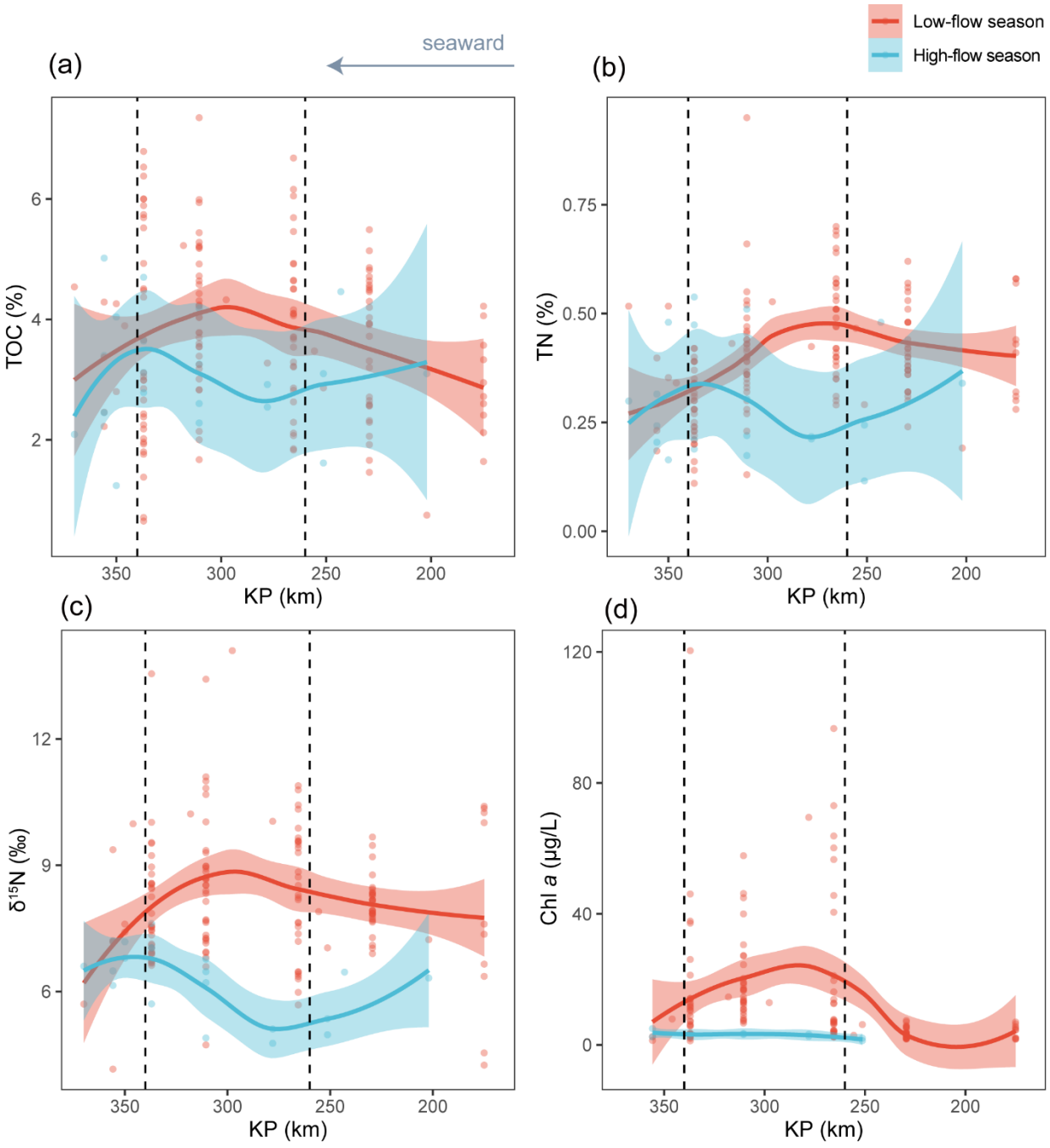

**Figure 4-10.** Schematic diagrams showing the biogeochemical functioning of the Seine Estuary in terms of POM dynamics in low-flow and high-flow scenarios.

## 4.5. Conclusion

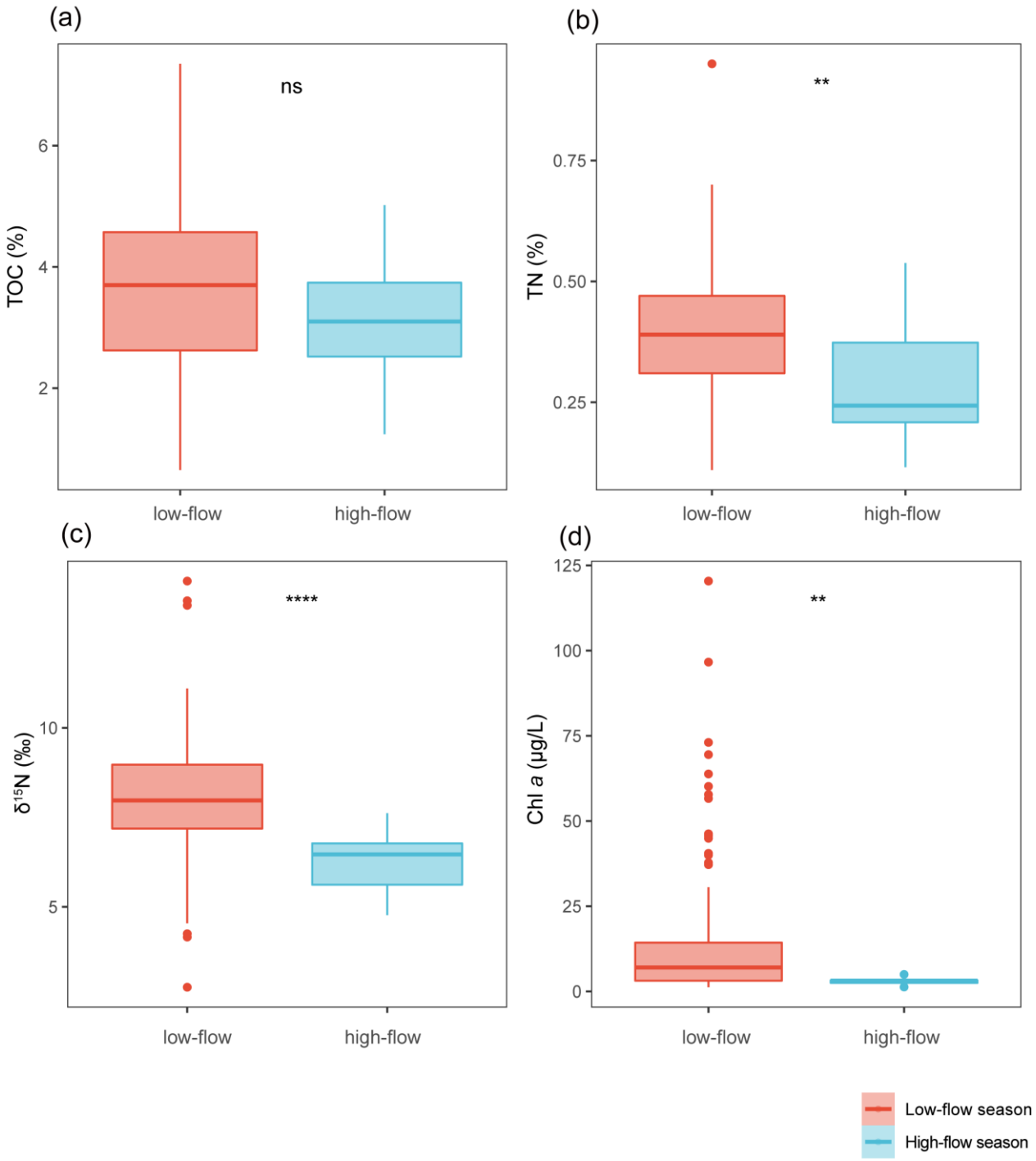
In this study, distributions of sterols, stanols, fatty acids, and *n*-alkanes were investigated in SPM (*n*=172) collected in high-flow and low-flow seasons across the Seine River basin. Spatio-temporal variations of anthropogenic POM, phytoplankton-derived POM, and plant-derived POM indicate that the different pools of POM (based on their sources) are subjected to contrasting dynamics and transformations. The dispersal and dynamics of these POM pools are closely related to hydroclimate conditions and land use types. Specifically, sewage-derived POM is positively correlated with water discharge and urban land use. The proportion of this anthropogenic carbon pool gradually decreases along the estuary in both high-flow and low-flow periods, implying that

estuaries act as effective filters/buffers, diluting sewage contaminations. During low-flow conditions in regions dominated by agricultural land use, phytoplankton blooms and potential priming effects occur; however, during high flows, these processes shift downstream, indicating that estuaries also act as biogeochemical reactors, stimulating phytoplankton blooms and possibly initiating priming effects. Finally, we propose a conceptual model to assess the functioning of estuarine ecosystems in high-flow and low-flow scenarios, which is crucial for sustainable estuarine management.

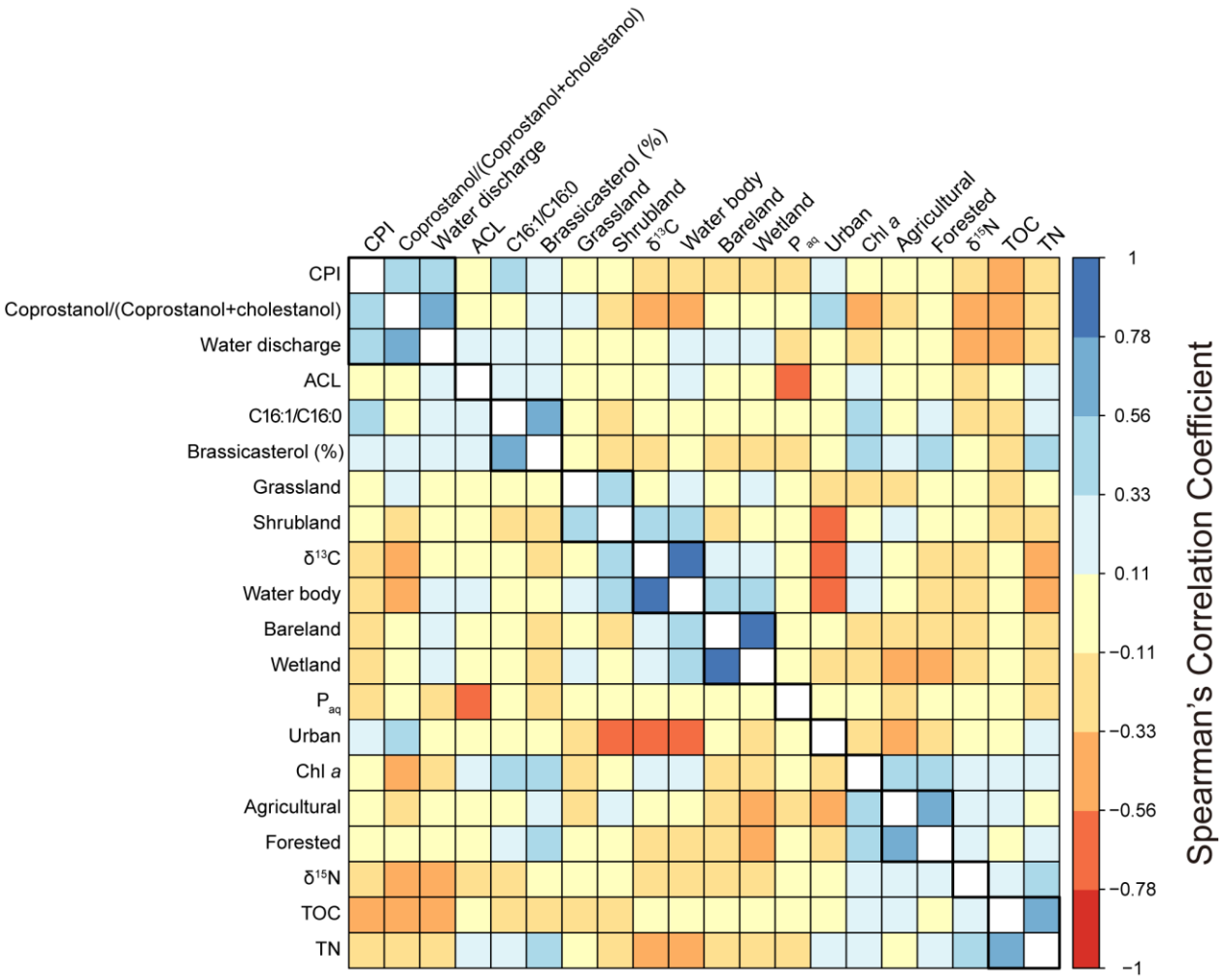
## 4.6 Annexes




**Supplementary Figure 4-1.** Spatio-temporal variations of bulk geochemical indices, including (a) TOC (%), (b) TN, (c) $\delta^{15}N$ (‰), and (d) Chla (µg/L). Kilometric Point (KP) represents the distance in kilometers from the city of Paris (KP 0). The trends showing proxy variations from site 4 (KP 175) to site 19 (KP 370) were based on locally estimated scatterplot smoothing (LOESS), with the shaded area representing 95% confidence intervals.




**Supplementary Figure 4-2.** Box plots comparing the bulk geochemical indices, including (a) TOC (%), (b) TN (%), (c) $\delta^{15}N$ (‰), and (d) Chla (µg/L) between low-flow (<250 $m^3/s$ - red) and high-flow (>250 $m^3/s$ - blue) seasons. Statistical testing was performed by using a Wilcoxon test (**$p < 0.01$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).




**Supplementary Figure 4-3.** Correlation plot between distinct (bulk and molecular) proxies, water discharge and land use types.

# Chapter 5:

## Disentangling dissolved organic matter composition by unsupervised and supervised machine learning

This chapter is in preparation for submission to *Science of The Total Environment*

## Abstract

Disentangling Dissolved Organic Matter (DOM) composition in estuaries is a major environmental concern, as the DOM properties are closely linked to critical biogeochemical cycling. However, tracing spatio-temporal variations of estuarine DOM is challenging due to multiple sources and complex transformation processes. Here, we investigate the dynamics of estuarine DOM using cutting-edge machine learning algorithms and explainable artificial intelligence. To this aim, we collected surface water samples ($n$=249) from a human-impacted estuary with intense industrialization and urbanization in France (Seine Estuary) across distinct land use characteristics in contrasting hydrological conditions. We then applied unsupervised and supervised machine learning to DOM optical properties, which were determined by UV–visible absorbance and Excitation-Emission Matrix (EEM) fluorescence spectroscopy combined parallel factor analysis (PARAFAC). Our results show that unsupervised machine learning (K-means clustering) captures the variabilities of DOM, identifying three estuarine zones based on pronounced spatial variations of several DOM optical parameters (relative abundances of PARAFAC components C2 (terrestrial origin), C3 (microbial/biological origin), C5 (protein-like substances), and fluorescence index (FI) especially within two clusters. Supervised machine learning (Light Gradient Boosted Machine, LightGBM) further validates the rationality of the defined zonation. Subsequently, explainable artificial intelligence based on SHapley Additive exPlanations (SHAP) analysis shows that DOM in each zone has specific characteristics. Our

model indicates that DOM in the Seine Estuary is primarily influenced by aromatic material and autochthonous contribution in the upper estuary (Zone I; Kilometre Point (KP) <260). The dominant contribution to DOM in the mid-estuary (Zone II; 260<KP<340) comes from autochthonous and aromatic material as well as transformation and (photo)degradation products. Lower estuary (Zone III; KP>340) is mainly characterized by aromatic DOM (subject to photodegradation), low molecular weight compounds, autochthonous DOM, as well as transformation and (photo)degradation products. Overall, this study presents a workflow for disentangling the composition of DOM, tracing its variability and dynamics along the land-to-sea continuum, and elucidating the involved processes.

*Keywords:* DOM; Fluorescence; Land-ocean continuum; Machine learning; Land use

## 5.1. Introduction

Estuaries refer to transition zones between freshwater and marine systems and play an important role in carbon cycling (Canuel and Hardison, 2016). Global rivers export a Dissolved Organic Carbon (DOC) flux of ~130 $TgC.yr^{-1}$ to the ocean for the past two decades (Fabre et al., 2020). Dissolved Organic Matter (DOM) accounts for an important reservoir of organic carbon on Earth. Within estuaries, DOM originates from distinct sources, which encompass allochthonous DOM from soils and plant litters, autochthonous DOM produced by biomass in the water column,

as well as anthropogenic DOM from agricultural runoff, industrial and urban effluents (Li et al., 2023; Xie et al., 2018; Zhou et al., 2021). DOM is linked to critical biogeochemical cycling in aquatic systems, including greenhouse gas emissions (Amaral et al., 2021; Begum et al., 2023), complexation with pollutants and metals (Jiang et al., 2017; Mori et al., 2019), and trophic network (Liu et al., 2023). Therefore, investigating the dynamics of estuarine DOM is a major environmental concern, which has recently received significant attention (Hounshell et al., 2022; Tang and Wang, 2022; Thibault et al., 2019; Vidal et al., 2023). However, constraining the sources and investigating the fate of estuarine DOM is still challenging, given its multiple sources, complex in-estuary processing (i.e. microbial and photochemical alterations), and varying factors (i.e. hydroclimate conditions and land use characteristics) (Asmala et al., 2013; Vidal et al., 2023; Zhang et al., 2022). A significant research gap is the lack of quantitative approach to disentangle different types of DOM and assess main DOM characteristics within specific estuarine area, which is important for understanding the ecological functioning of estuaries.

Advanced statistical techniques, such as machine learning algorithms, have been applied for pattern recognition and analysis of complex environmental datasets dealing with DOM (Harjung et al., 2023; Liao et al., 2023; Yi et al., 2023; Zhao et al., 2023). Generally, machine learning can be classified into unsupervised and supervised learning. Unsupervised learning algorithms are used to uncover the inherent traits and hidden patterns within unlabeled (data points without corresponding labels) dataset (Huang et al., 2021). On the other hand, supervised learning algorithms are typically applied to handle labeled (each data point with a corresponding label)

dataset and make predictions by regression (regression tasks; prediction of a continuous numeric value) or by classification (classification tasks; assignment of input data into predefined categories). Explainable artificial intelligence is an emerging approach to provide a reasonable interpretation of the black box machine learning model by evaluating the importance of each input variable (Liao et al., 2023; Park et al., 2022; Zhao et al., 2023). The machine learning and explainable artificial intelligence have been successfully used for regression tasks within the field of environmental science. For example, complex correlations between DOM properties and apparent quantum yields of photochemically produced reactive intermediates are identified by these approaches (Liao et al., 2023). However, to the best of our knowledge, the use of machine learning and explainable artificial intelligence for classification tasks, notably disentangling DOM composition and identifying main DOM characteristics within specific estuarine zones, is currently lacking and needs to be explored.

Previous studies (Butturini et al., 2016; S. Chen et al., 2021; Hu et al., 2022; Singh et al., 2019) showed that DOM characteristics are highly dynamic under varying hydroclimate conditions (i.e. temperature and water discharge) and land use characteristics (i.e. urban and agricultural land use), making it challenging to identify spatio-temporal variations of DOM and to evaluate the estuarine functioning (e.g. the role that estuaries play in regulating DOM dynamics). Advanced statistical approaches, including unsupervised, supervised machine learning, and explainable artificial intelligence may help to disentangle DOM composition and to identify main DOM characteristics in specific estuarine regions. The hypothesis was tested by investigating the DOM

optical properties of the surface water samples (*n*=249) in low-flow and high-flow periods from the Seine Estuary (NW France), which spans urbanized, industrialized, and agricultural regions. This estuary is highly stressed by both natural fluctuations and anthropogenic pressures, making it representative of human-impacted estuaries, and it is one of the most contaminated hydrosystems in the Northern hemisphere. The aim of the present study is to (ii) investigate the dynamics of DOM along the land-sea continuum, (ii) identify the main DOM characteristics in specific estuarine zones, and (iii) evaluate the potential of machine learning and explainable artificial intelligence to disentangle the heterogeneity of DOM, tracing its dynamics and provide biogeochemical interpretations.

## 5.2. Materials and methods

### 5.2.1 Study area and sampling

The Seine Estuary (NW France, Figure 5-1a) is approximately 160 km in length, occupying an area of 50 $km^2$, and is characterized as a macrotidal estuary based on its small depth, high tidal range and morphology (Grasso et al., 2018; Romero et al., 2019). The average monthly water discharge of the Seine River from 2019 to 2022 at the Paris Austerlitz station (F700 0001 03, retrieved from HydroPortail - https://www.hydro.eaufrance.fr/) is generally higher in winter (above 250 $m^3/s$, Figure 5-1b) and lower in the other seasons (below 250 $m^3/s$, Figure 5-1b). Samples collected during the 5 high-flow campaigns are represented by the 'high-flow period' (in blue, Figure 5-1b), and during all other periods, represented by the 'low-flow period' (in red, Figure 5-

1b).The Seine Estuary acts as the outlet for the Seine drainage basin and stands out for its high population density (200 inhabitants per square kilometer on average), as well as its substantial agricultural and industrial activities (Romero et al., 2019). Eight land use types were identified in the Seine Estuary, categorized as agricultural, forested, shrubland, water body, bareland, grassland, urban (industrial zones included), and wetland (Figure 5-1a, c). The land use information was obtained from the global surface coverage product GLOBELAND30 (http://www.globallandcover.com/) with 30 m resolution recorded in 2020. Seawater and inland water body are grouped together under the single category of "water body." A 1km (radius) buffer zone surrounding each sampling site was constructed to calculate the land use type proportions using the ArcGIS (10.7) software. A 1 km buffer is selected because it is capable of capturing the influence of nearby land use patterns on the DOM characteristic within the water column (Hu et al., 2016; Zhang et al., 2023).

During 19 monitoring campaigns from June 2019 to November 2022, sub-surface water (ca. 1m depth; 20mL) samples ($n$=249) were collected at 15 locations in contrasted seasons across the Seine Estuary with distinct land cover regimes (Figure 5-1 and Table 5-1). For these sites, water samples were immediately filtered through pre-combusted (450 °C) 0.7 µm glass fiber filters (GF/F Whatman) on board and stored in darkness at 4 °C until analysis.

**Table 5-1.** Sampling sites

| Site | Name | Longitude (°) | Latitude (°) | KP (km) | Number of samples |
|---|---|---|---|---|---|
| 0 | Balise A | 0.110671 | 49.431828 | 360.8 | 10 |
| 1 | Honfleur | 0.232682 | 49.432638 | 355.8 | 19 |
| 2 | Berville-Sur-Mer | 0.3682 | 49.441587 | 346 | 19 |
| 3 | Tancarville | 0.463442 | 49.472351 | 337 | 19 |
| 4 | Petitville | 0.577669 | 49.435988 | 326.6 | 19 |
| 5 | Vatteville-La-Rue | 0.66614 | 49.472695 | 318 | 19 |
| 6 | Caudebec | 0.72753 | 49.522585 | 310.5 | 19 |
| 7 | Le Trait | 0.776177 | 49.483864 | 303 | 17 |
| 8 | Heurtauville | 0.816867 | 49.447614 | 297.65 | 18 |
| 9 | Duclair | 0.873297 | 49.478666 | 278 | 18 |
| 10 | La Bouille | 0.934366 | 49.35228 | 259.7 | 18 |
| 11 | Haulot Sur Seine | 0.98475 | 49.356683 | 255.6 | 18 |
| 12 | Petit Couronne | 1.008118 | 49.379279 | 251.3 | 17 |
| 13 | Le Grand Quevilly | 1.030269 | 49.432815 | 246.6 | 18 |
| 14 | Rouen | 1.06979 | 49.4428698 | 243 | 1 |

### 5.2.2 DOC concentration measurement

The DOC concentrations were determined using an aliquot of water sample (50 µL), which was acidified with 0.75 µL HCl (2 mol/L) and analyzed in Non-Purgeable Organic Carbon (NPOC) mode using Total Organic Carbon Analyzer (Shimadzu, Tokyo, Japan). 3 replicate analyses were made for each sample. The average value is reported here, with the relative standard deviation below 1%.

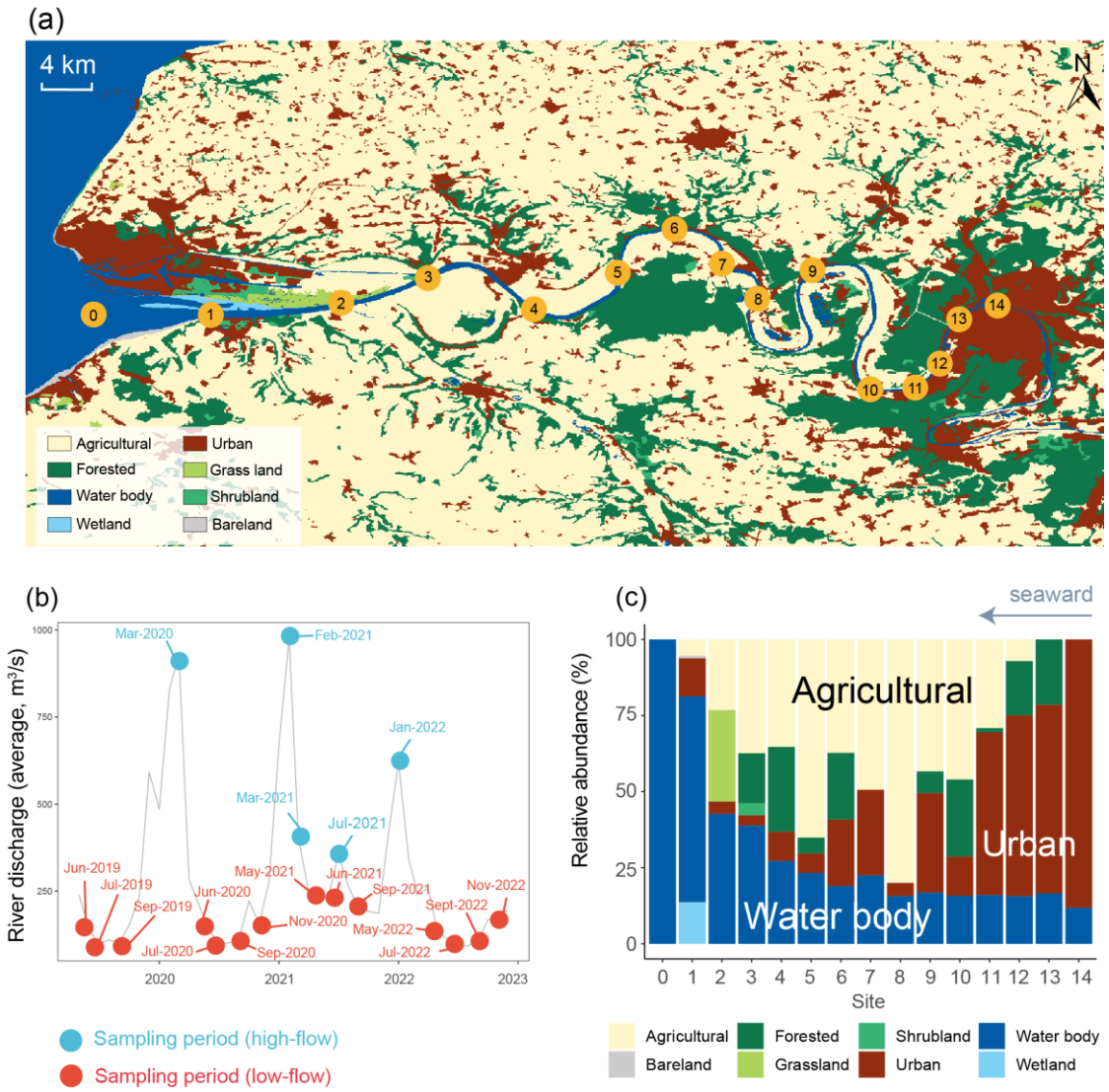



**Figure 5-1.** (a) Map of the study area (Seine Estuary) showing the land use classification (agricultural, urban, forested, grass land, water body, shrubland, wetland, and bareland), with orange bullets representing sampling sites. The land use information was retrieved from the global surface coverage product GLOBELAND30 (http://www.globallandcover.com/). Seawater and inland water body are combined into a single category of "water body". Industrial regions are included in "urban". (b) Water discharge (mean monthly) of the Seine River from 2019 to 2022 measured at the Paris Austerlitz station (data retrieved from https://www.hydro.eaufrance.fr/). Sampling period of this study is shown by bullets with different color. The red bullets represent samples were collected in the low-flow (<250 $m^3/s$) period and the blue bullets denote samples were collected in the high-flow (>250 $m^3/s$) period. (c) Variation of the land use relative abundances along the Seine Estuary.

### 5.2.3 Spectroscopic analyses

The spectroscopic analyses (absorbance and fluorescence) were carried out in a Hellma Suprasil® quartz cell with a path length of 1 cm. The UV–Visible absorbance spectrum of water samples ($n = 249$) was recorded using a Jasco® V-760 spectrophotometer. The absorbance spectra for water samples were acquired between 210 nm and 700 nm at 200 nm/min. The absorbance spectrum of the ultrapure water blank daily acquired was subtracted from the spectrum of each sample. Samples were diluted with ultrapure water when the maximum of absorbance was above 0.1 to avoid an inner-filtering effect in subsequent fluorescence analyses. The variation of DOM molecular weight was assessed by spectral Slope Ratio (SR), which corresponds to the ratio of the slope for wavelengths in the 275–295 nm region to that in the 350–400 nm region, with higher values indicating a lower DOM molecular weight (Helms et al., 2008). The specific ultraviolet absorbance at 254 nm ($SUVA_{254}$), with high values indicating greater aromatic content (Weishaar et al., 2003), was calculated as follows and expressed in L mg-$C^{-1}$ $m^{-1}$:

$$SUVA_{254} = Abs_{254} / (L \times DOC) \quad (1)$$

In Eq. 1, $Abs_{254}$ is the measured absorbance at 254 nm, L is the path length (m), and DOC is the dissolved organic carbon concentration (mg/L).

The excitation-emission matrix (EEM) fluorescence spectra ($n = 249$) were obtained between the wavelengths 240–800 nm at excitation (2 s integration time, 5 nm intervals) and 245-830 nm at emission (high CCD detector gain, 1 pixel (ca. 0.58 nm intervals)), using an Aqualog

spectrofluorometer (Horiba Scientific, France) equipped with a xenon lamp (150W), a double monochromator at excitation, and a CCD detector. To eliminate Rayleigh and Raman scatter peaks, each sample EEM spectrum was subtracted from the ultrapure water blank EEM spectrum daily acquired. The area of the Raman scattering peak of ultrapure water is calculated daily at the excitation of 350 nm and allows the spectra to be normalized. The fluorescence intensities are thus expressed in Raman Units (RU). The EEM spectra were then processed to record fluorescence intensities and calculate fluorescence indices, including γ/α (Huguet et al., 2009; Parlanti et al., 2000), FI (McKnight et al., 2001), BIX (Huguet et al., 2009) and HIX (Zsolnay et al., 1999), using the TreatEEM software (Omanović et al. 2023). Fluorescence indices are summarized in Chapter 1 (Table 1-4).

### 5.2.4 Parallel factor analysis (PARAFAC)

The 3D EEM fluorescence spectra can be decomposed by a multi-way PARAllel FACtor analysis (PARAFAC) into independent underlying fluorescent components (Stedmon et al., 2003b). This statistical method helps to identify the fluorophores contributing to the overall spectral dataset and to estimate their relative contribution to the total DOM fluorescence. The PARAFAC model was carried out using the DOM Fluor toolbox (version 1.7) in Matlab R2021b and run for 4 to 8 components with non-negativity constraints (Stedmon and Bro, 2008). A six-component model was validated after split-half validation analysis and examination of the residuals (Murphy et al., 2013; Stedmon and Bro, 2008). The spectral characteristics of the components determined by

PARAFAC were further compared to those identified in other environments by an online spectral library (Openfluor) (Murphy et al., 2014). The similarity between the six components determined in this study and those in the database was measured using Tucker's congruence coefficient, with criteria set at 95%.

### 5.2.5 Unsupervised machine learning

K-means clustering, an unsupervised machine learning technique, was used to find clusters (group together similar samples) based on DOM optical parameters in an unlabeled dataset ($n$=249). K-means clustering was performed using the KMeans function from the cluster module of the scikit-learn library (https://github.com/scikitlearn/) (Pedregosa et al., 2011) in Python 3.9.16. The optimal number of clusters (K) was chosen using the elbow method.

### 5.2.6 Supervised machine learning

A cutting-edge supervised machine learning algorithm (Light Gradient Boosting Machine, LightGBM) (Ke et al., 2017) was trained for classifying the estuarine zones based on DOM optical properties in a labelled dataset (i.e. zonation of the estuary), implemented with the LightGBM package (https://lightgbm.readthedocs.io) in Pyhton (3.9.16). This algorithm is a tree-based gradient boosting framework, which works by combining weak decision trees to create a strong model (Ke et al., 2017). Specifically, it starts by constructing a single decision tree based on the input data that predicts the target variable. Further decision trees are added to the model iteratively,

with each tree aiming to correct the errors of the previous tree. It has been widely used across many domains because of its high accuracy, high training speed, and low memory consumption (Aiken et al., 2022; Alova et al., 2021; Ke et al., 2017).

We split our dataset into 75% training dataset and 25% test set randomly. The training set is used for fitting the machine learning model, whereas the test set (independent set of new data that has never been used in training) is used to evaluate the model performance. We further used a standard framework (Synthetic Minority Oversampling Technique, SMOTE) to solve the class imbalance problem, which occurs when one class (zone) contains significantly fewer samples than the other classes (Chawla et al., 2002). This technique is used to oversample an imbalanced training set, implemented with the imblearn library (Lemaître et al., 2017) (https://github.com/scikit-learn-contrib/imbalanced-learn) in Python (version 3.9.16).

### 5.2.7 Evaluation of the supervised machine learning model

We run 10-fold cross-validation experiments to avoid overfitting and assessed the model performance. With the 10-fold cross-validation, we divided our training set into 10 parts randomly. The model was trained using nine of these parts and tested with the remaining one. This procedure was repeated ten times.

The performance of the machine learning classification model was also evaluated by the recall (sensitivity of model prediction), precision (hitting ratio of positive predictions), AU-ROC (area under the receiver operating characteristic curve), and AU-PRC (area under the precision-

recall curve). In the Receiver Operating Characteristic (ROC) curve, True Positive (TP; the number of positive samples correctly classified) rate was plotted against the False Positive (FP; the number of negative samples wrongly classified as positive) rate for distinct thresholds. The ROC curve demonstrates how well the classification model distinguishes between classes (zones), with higher AU-ROC indicating a better model performance. In addition, the Precision-Recall Curve (PRC) was used to show the tradeoff between precision and recall for distinct thresholds. The PRC is a graph displaying recall values on the x-axis and precision values on the y-axis. It is typically used when classes (zonation of the estuary) are significantly imbalanced, with higher AU-PRC elucidating a better classifier performance.

### 5.2.8 Explainable Artificial Intelligence

The explainable artificial intelligence framework (SHapley Additive exPlanations, SHAP) is a game theoretical approach, which is used for interpreting black-box models such as gradient-boosting machines (Lundberg et al., 2020). The SHAP method was used in this study to evaluate the weight/importance of distinct features (individual DOM optical properties) in the trained machine learning model, with high SHAP values indicating a stronger positive influence of that feature on the specific prediction, implemented with the SHAP package (https://github.com/slundberg/shap) in Python (3.9.16). The main DOM characteristics in each class (zonation of the estuary) were further identified.

### 5.2.9 Other statistical analyses

Other statistical analyses were performed using R (version 4.2.1). Due to the non-normal distribution of our dataset ($p < 0.05$; Shapiro–Wilk normality test), non-parametric statistical tests were performed in this study. The unpaired two-sample Wilcoxon test (also known as Wilcoxon rank sum test or Mann-Whitney test) was used for comparing two independent groups, while Spearman's correlation was used to explore correlation patterns among distinct variables. The significance level is based on p-value and denoted by distinct symbols: * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$, **** represents $p < 0.0001$, and "ns" represents not significant, with $p > 0.05$.

Principal Component Analysis (PCA) was performed based on optical parameters with the R packages factoextra and FactoMineR. Samples clustered in distinct groups were highlighted with 95% concentration ellipses. The significance of separation of different groups (clusters) was further assessed by permutational multivariate analysis of variance using distance matrices (Adonis test, 999 permutations), which was implemented using the adonis2 function of the R package vegan.

Redundancy analysis (RDA) was used to investigate the impact of land use and water discharge on DOM optical proxies and was performed with the R package vegan. Straight or small angles (close to 180° or 0°) indicate negative or positive correlations, respectively, whereas right angles (90°) imply a lack of linear correlations between variables. Thus, the DOM optical parameters that are close to each other in the RDA are strongly correlated and can be considered

as responding to environmental factors similarly. We further used a hierarchical portioning approach to calculate the individual importance of explanatory variables on response variables (optical properties), which can generate an unordered assessment of individual importance. This approach was implemented with the R package rdacca.hp (Lai et al., 2022).

A locally estimated scatterplot smoothing (LOESS) method was used to investigate the spatio-temporal variations of optical and environmental parameters. It can capture the nonlinear pattern of the dataset and buffer the outliers. This method was carried out using the smooth function from the R package ggplot2. The colored region represented the 95% confidence intervals for each group (cluster).

## 5.3. Results and discussion

### 5.3.1 Complexity of DOM characterization across the land-sea continuum

The final six-component model determined by PARAFAC analysis (C1-C6; Figure 5-2 and Supplementary Figure 5-1) accounted for>99.6% of the measured spectral variation for the 249 surface water samples from the Seine Estuary collected between June 2019 and November 2022. A summary of spectral characteristics, potential origin of these components and number of PARAFAC models matched with Tucker congruence coefficient of over 0.95 on the excitation and emission spectra simultaneously in the online spectral database (OpenFluor) are provided in Table 5-2.

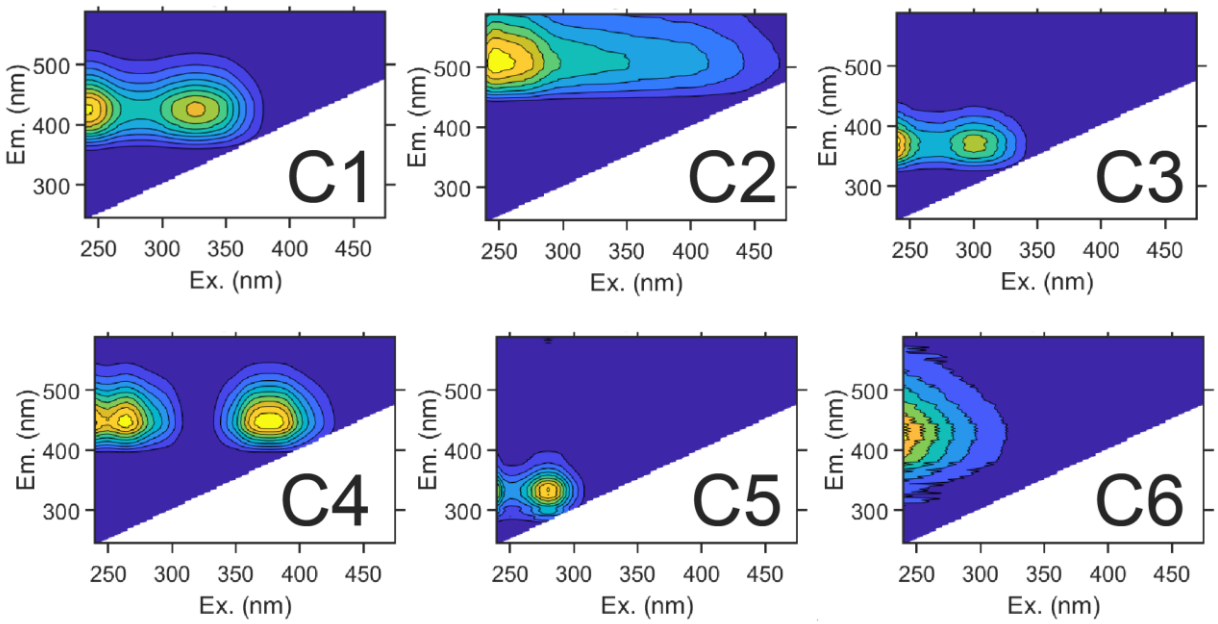



**Figure 5-2.** Contour plots of the six components determined by PARAFAC for surface water samples (*n*=249) collected in the Seine Estuary during 19 monitoring campaigns from June 2019 to November 2022.

**Table 5-2**. Spectral characteristics of the six PARAFAC components

| Component | Ex(max)/Em(max) | Potential origins and characteristics | References | Number of OpenFluor matches |
|---|---|---|---|---|
| C1 | 240(325)/429 | Terrestrial substances[a,b,c]; Common in freshwater[a,b,c]; Biologically degraded and produced[d]; Subject to photodegradation[d]; Exported from agricultural catchments[e] | ([a]Weigelhofer et al., 2020; [b]Stedmon and Markager, 2005; [c]Yamashita et al., 2011; [d]Zhuang et al., 2022; [e]Graeber et al., 2012) | 82 |
| C2 | 245/510 | Terrestrial substances[b,f,g]; Common in freshwater[b,f,g]; High-molecular-weight and aromatic compounds[b,h]; Subject to photodegradation[d]; Terrestrial components in coastal environments[i,j] | ([f]Lambert et al., 2016; [g]Wünsch et al., 2017; [h]Yamashita et al., 2008; [i]Kowalczuk et al., 2009; [j]Yamashita et al., 2011b) | 64 |
| C3 | 240(300)/374 | Biological/Microbial origin[e,k,l] | ([k]Parlanti et al., 2000; [l]Williams et al., 2010) | 28 |
| C4 | 265(375)/445 | Bacterial origin[m,n,o]; Terrestrial origin in agricultural aeras[p] | ([m]Fox et al., 2017; [n]Fox et al., 2021; [o]Lee et al., 2018; [p]Amaral et al., 2020) | 13 |
| C5 | 280/335 | Protein-like substances[b,k,q,r] | ([q]Catalán et al., 2021; [r]Kim et al., 2022) | 83 |
| C6 | 240/431 | Transformation and degradation product[s]; Photodegradation product[t,u,v] | ([s]Osburn et al., 2017; [t]Stedmon et al., 2007; [u]De Francesco and Guéguen, 2021; [v]Ishii and Boyer, 2012) | 43 |

OpenFluor comparison conducted on 18/07/2023

C1 [Excitation/Emission (Ex/Em) maxima: 240(325)/429 nm] and C2 [Ex/Em maxima: 245/510 nm] are fluorophores commonly observed in freshwaters and attributed to terrigenous, aromatic and hydrophobic DOM (Lambert et al., 2016; Stedmon and Markager, 2005; Weigelhofer et al., 2020). Component C1 was also described as terrestrial DOM exported from agricultural catchments (Graeber et al., 2012) and components similar to C2 were attributed to terrestrial components in coastal environments (Kowalczuk et al., 2009; Yamashita et al., 2011, 2008). It should be noted that both C1 and C2 can be photodegraded (Zhuang et al., 2022), while only C2 could be produced photochemically (Ishii and Boyer, 2012). C3 [Ex/Em maxima: 240(300)/374 nm] is categorized as a material recently produced autochthonously or transformed by biological/microbial activity (Graeber et al., 2012; Guéguen et al., 2014; Parlanti et al., 2000). C4 [Ex/Em maxima: 265(375)/445 nm] has often been reported as a component from terrestrial sources (Murphy et al., 2013; Stedmon et al., 2003b), particularly in agricultural areas (Amaral et al., 2020). The EEM spectrum of this component is actually similar to that of the siderophore pyoverdine (Cornu et al., 2022; Dartnell et al., 2013), which is an extracellular metabolite produced notably by the bacterium *Pseudomonas aeruginosa* (Fox et al., 2017) mainly observed in places subject to human activities (Crone et al., 2020; Pirnay et al., 2005). C5 [Ex/Em maxima: 280/335 nm] is closely related to proteins or amino acids, and associated with biological activity (Catalán et al., 2021; Hambly et al., 2015; Huguet et al., 2009; Kim et al., 2022). C6 [Ex/Em maxima: 240/431 nm] is attributed to DOM transformation/degradation residue (Osburn et al., 2017) and has been identified as a common product of photodegradation in various environments (DeFrancesco and Guéguen, 2021; Stedmon et al., 2007). It is thought to be aromatic product of photochemical degradation that is resistant to further photodegradation (Ishii and Boyer, 2012).

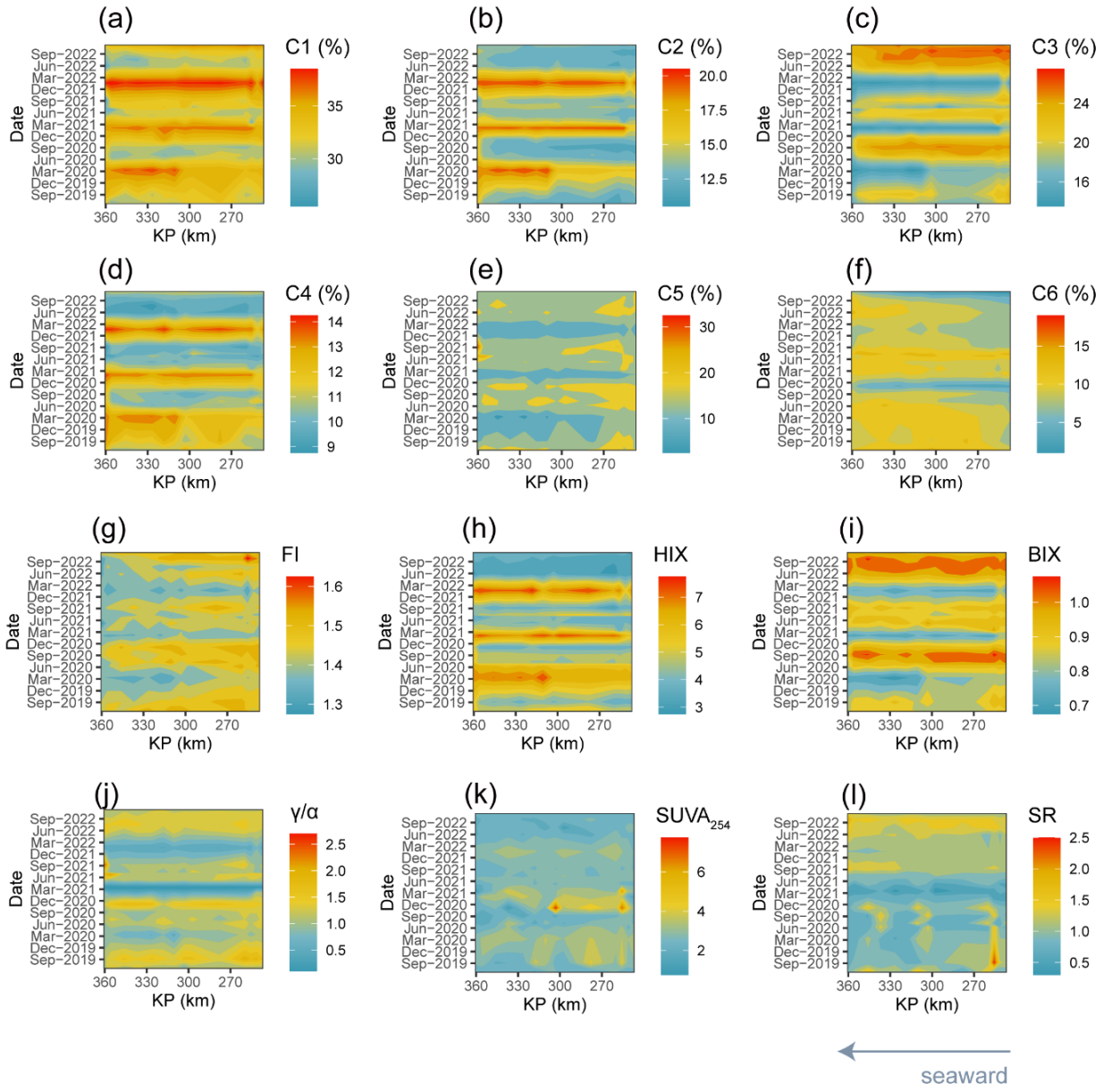



**Figure 5-3.** Contour plots showing the spatial and temporal variations of the relative percentage of the six PARAFAC components: (a) C1, (b) C2, (c) C3, (d) C4,(e) C5, (f) C6; the fluorescence indices (g) fluorescence index – FI, (h) humification index – HIX, (i) biological index – BIX, (j) fluorescence intensity ratio γ/α; and the absorbance indices (k) specific UV absorbance - $SUVA_{254}$, (l) spectral slope ratio – SR; for the samples collected in the Seine Estuary from upstream (kilometre point (KP) 246) to downstream (KP 361) during 19 campaigns between June 2019 and November 2022 (*n*=249).

The relative abundances of these components (fluorescence intensity of each component to the total fluorescence intensity for each sample) show significant temporal variations in the Seine Estuary (Figure 5-3). Specifically, average relative proportions of C1, C2 and C4 are much higher in the high-flow periods (above 250 $m^3/s$) than in the low-flow periods (below 250 $m^3/s$; $p<0.05$; Figure 5-4). Such variability linked to hydrological conditions is also shown by the intensities of these components, absorbance (SR and $SUVA_{254}$) and fluorescence (HIX, BIX and FI) indices (Figure 5-3, 5-4 and Supplementary Figure 5-2).

The distinct hydrological conditions may lead to higher proportions of aromatic compounds with high molecular weight at high flows in the Seine Estuary. Indeed, we observe that all aromatic/terrestrial indicators (intensities as well as relative abundances of C1, C2, C4; $SUVA_{254}$ and HIX) significantly increase with increasing water discharge in the Seine Estuary ($p<0.05$, Spearman's correlation; Supplementary Figure 5-3, 5-4). Such significant relationship indicates that higher proportions of terrestrial/aromatic DOM are flushed into the water column due to enhanced precipitation during high-flow events, as previously observed in other land-sea continuums such as the Neuse River Estuary (Hounshell et al., 2022) and the Changjiang River Estuary (Zhang et al., 2022). Moreover, the intensities and relative abundances of photogenerated aromatic component (C6) are significantly higher at high flows compared to the low-flow period ($p<0.05$, Wilcoxon test; Figure 5-4f and Supplementary Figure 5-2f). This photodegradation product has high aromaticity and is suggested to be produced from terrestrial material (Du et al., 2016; Grunert et al., 2021; Ishii and Boyer, 2012). Therefore, the significantly higher portion of C6 during high flows could be attributed to the greater amount of terrestrial DOM that is flushed into the Seine Estuary, providing more material for photochemical alteration. This is particularly true for the flood period of July 2021, with an increase in terrestrial DOM inputs concomitantly with higher levels of solar irradiation in summer (Figure 5-3). In addition, highest intensities and

relative proportions of C6 (photoproducts) are observed during periods of low flow in summer (June 2019; Supplementary Figure 5-5f and 5-6f). This further indicates that seasonal variations may also regulate DOM dynamics, in addition to hydrological conditions.

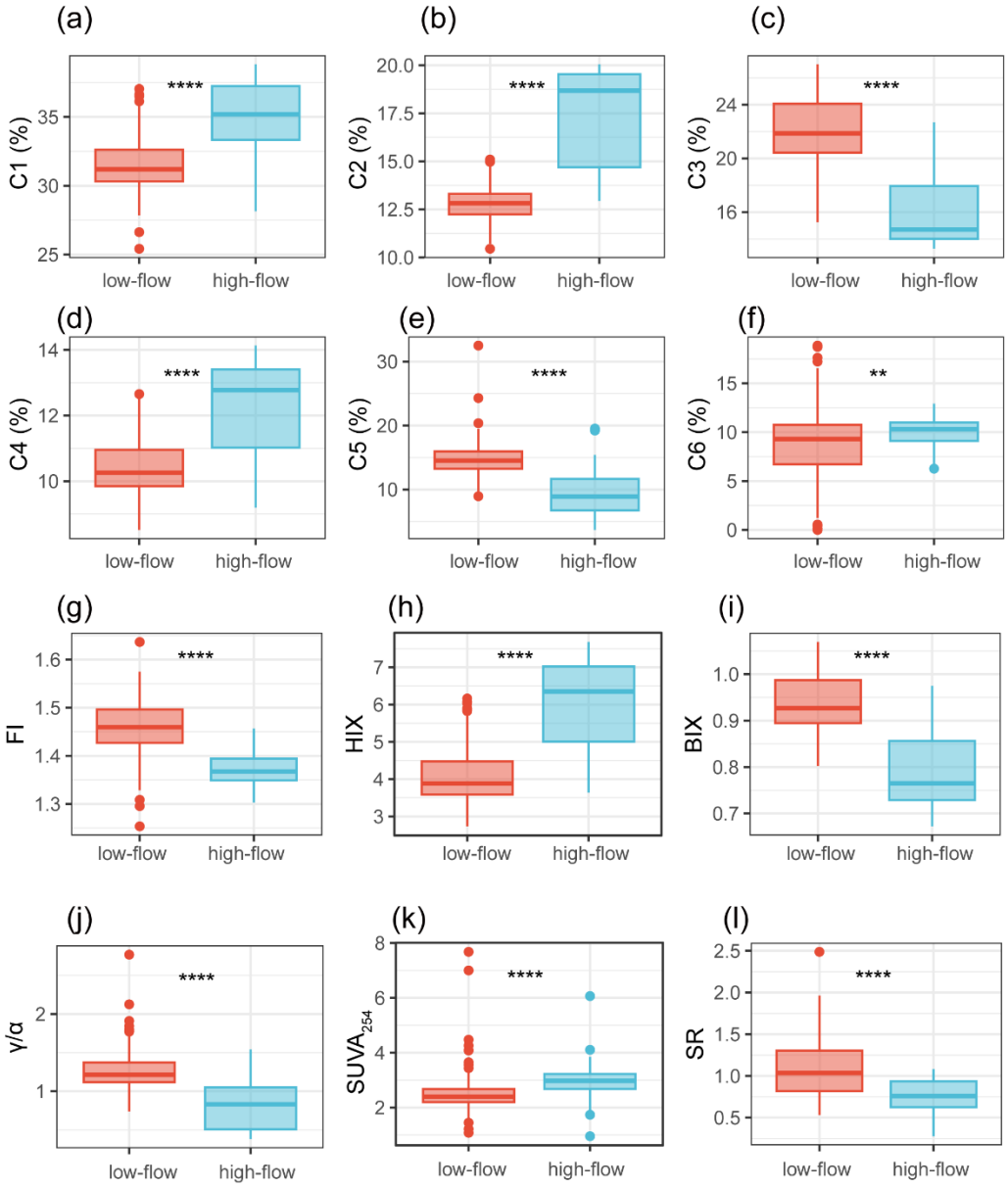



**Figure 5-4**. Box plots comparing the DOM optical parameters between high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) seasons. Statistical testing was performed using a Wilcoxon test (**$p < 0.01$; ****$p < 0.0001$).

Indeed, the RDA triplot shows that the first axis (24.24 % of the variability) was mainly linked to temperature and water discharge, highlighting the impact of seasonal and hydrological

conditions on the DOM spectral characteristics in the Seine Estuary (Figure 5-5). Hierarchical partitioning further elucidates that temperature and water discharge are the most important parameter controlling the DOM characteristics in the Seine Estuary (13.8 % and 13.15% of the variance, respectively; Figure 5-5 and Supplementary Table 5-1). In contrast to previous studies in other river watersheds which showed that land use can significantly control the DOM dynamics (Bhattacharya and Osburn, 2020; Hu et al., 2022), there is a minor impact of land use type in the Seine Estuary, with urban land use accounting for 1.43 % of the total variance (Figure 5-5), which suggests that the influence of land use characteristics on the DOM optical compositions is likely masked to some extent by the dominating natural driver (i.e. hydroclimate conditions).

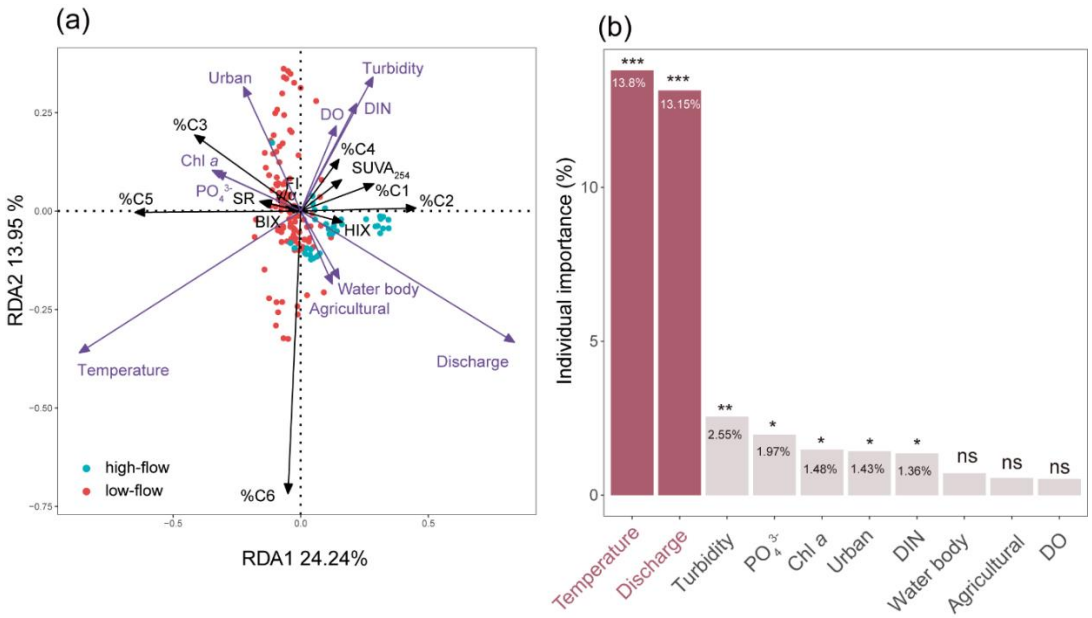



**Figure 5-5.** (a) RDA analysis between available environmental variables (purple arrows) and DOM optical parameters (black arrows). Samples are colored according to hydrological conditions, including high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) periods. (b) The individual importance of different environmental variables explaining the variation in DOM optical parameters were assessed by hierarchical partitioning analysis. Significance level is indicated by asterisks: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; ns, not significant, $p > 0.05$. *p*-values are from permutation tests (999 randomizations). Physical parameters (turbidity, temperature, dissolved oxygen - DO), inorganic nutrients, and Chlorophyll *a* (Chl *a*) are measured by Serre-Fredj et al. (2023). Dissolved Inorganic Nitrogen (DIN) = $NO_3^-$-N + $NH_4^+$-N + $NO_2^-$-N.

In addition to regulating the terrestrial/aromatic DOM, hydroclimate conditions can also influence the production and reactivity of autochthonous DOM (Butturini et al., 2016; Ejarque et al., 2017; Hounshell et al., 2022; Li et al., 2023). For example, protein-like DOM is strongly produced in the Tordera River during the drought condition (Butturini et al., 2016). Low precipitation can increase the water residence time, which allows for considerable production and degradation of autochthonous DOM (Singh et al., 2019). The river thus become a reactive system for autochthonous DOM at low flows (Butturini et al., 2016). In the Seine Estuary, optical parameters for autochthonous contributions (intensity and relative abundances of C3 and C5; FI, BIX, and $\gamma/\alpha$) are significantly higher in the low-flow condition ($p<0.05$, Wilcoxon test; Figure 5-4 and Supplementary Figure 5-2). In addition, these indicators significantly decrease with increasing water discharge ($p<0.05$, Spearman's correlation; Supplementary Figure 5-3, 5-4). This shows that DOM in the low-flow condition is largely comprised of autochthonous material.

During the low-flow period, the high temperature and slow water flow velocity promote the phytoplankton growth and associated microbial activities, thus releasing more autochthonous DOM into the water column. Indeed, temperature is related to the phytoplankton biomass (Chl *a*) and indicators for autochthonous DOM (i.e. %C3 and %C5) that scored negatively on the first axis of the RDA (Figure 5-5a). Such seasonal variability of autochthonous DOM is also observed in many other estuaries (Ejarque et al., 2017; Hounshell et al., 2022; Singh et al., 2019).

While the variation of DOM in the Seine Estuary is clearly captured by dividing the dataset based on the water discharge, the spatial distribution of DOM seems to be much more complex (Figure 5-3). Spatially, the relative fluorescence of the terrestrial/aromatic components (i.e. %C1, %C2, and %C4) is much higher than for the rest of the components over the whole estuary during the high discharge periods occurring in winter (December to March generally), as reflected in the high HIX values, with no particular trends between upstream and downstream (Figure 5-3).

However, these terrestrial/aromatic indicators do not show similar trends as the absorbance proxies used for assessing the aromaticity ($SUVA_{254}$) and molecular weight (SR) of DOM along the estuary (Figure 5-3). Furthermore, two autochthonous components (%C3 and %C5) and fluorescence proxies (FI, BIX and $\gamma/\alpha$) do not show similar spatial variations. When samples are grouped based on hydrological conditions (high-flow and low-flow conditions), there are also distinct spatial variations observed for these indicators (Supplementary Figure 5-5). Such a heterogeneity could be due to distinct sources and transformation processes of estuarine DOM (Asmala et al., 2018). Hence, it remains difficult to identify the spatial trends in DOM composition along the estuary by simply visualizing the variations of the DOM optical properties or by grouping samples based on hydrological conditions (Figure 5-3 and Supplementary Figure 5-5).

### 5.3.2 DOM heterogeneity captured by unsupervised machine learning

To discover the hidden patterns of the complex estuarine DOM data and explore the potential (temporal and spatial) variability, we perform an unsupervised machine learning approach with unlabeled data. Specifically, we applied the K-means clustering to identify clusters of similar samples based on DOM optical compositions.

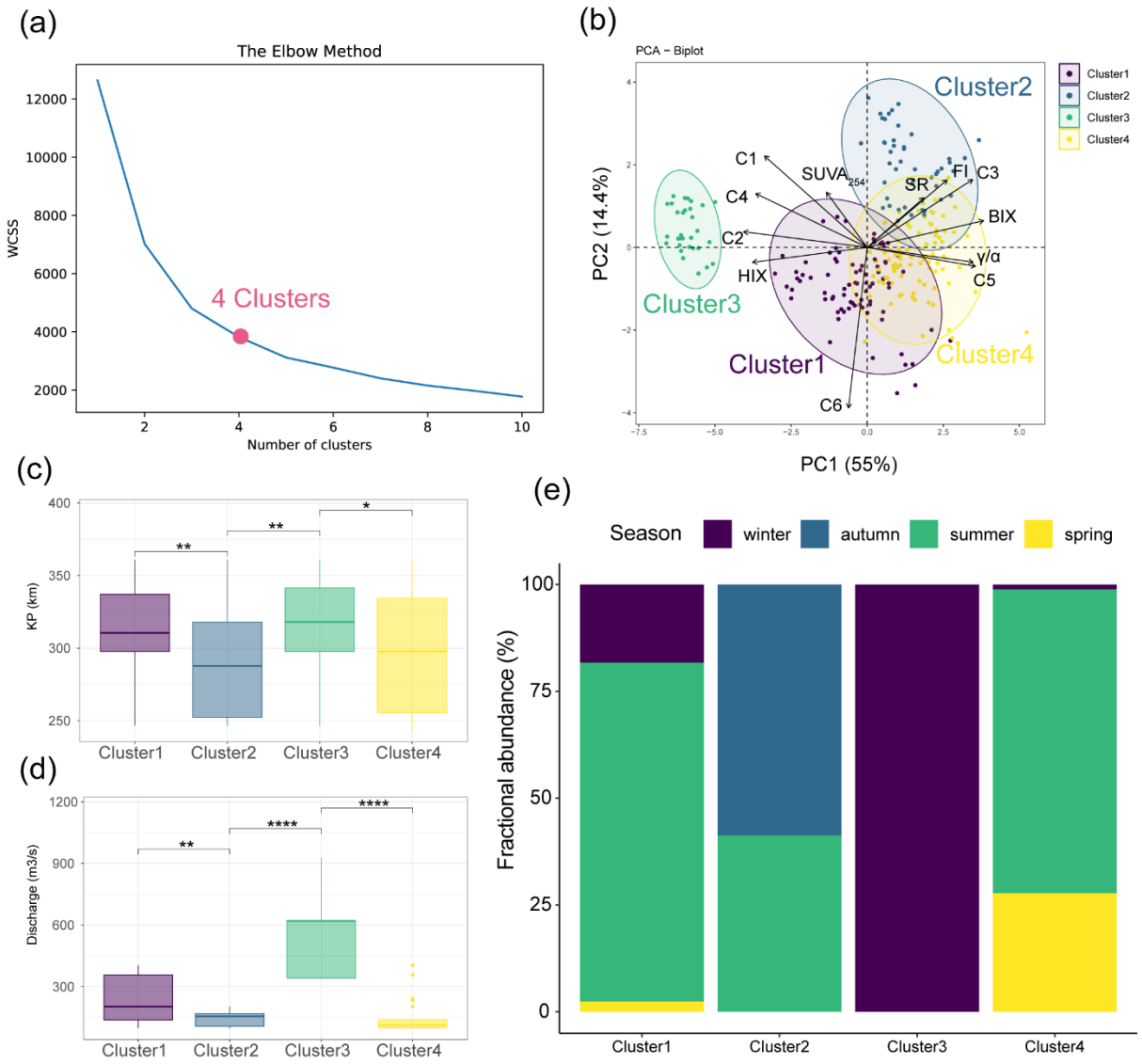



**Figure 5-6.** (a) Determination of optimal number of clusters (K) in K-means clustering based on the elbow method. (b) PCA analysis of DOM optical parameters. Samples ($n$=249) within different clusters were highlighted with 95% concentration ellipses. Adonis analysis (999 permutations) was performed to assess how many variations of DOM optical proxies are explained by the grouping (clusters). (c-d) Box plots showing the distribution of (c) KP (Kilometric Point; defined as the distance in kilometers from the city of Paris) and (d) mean monthly water discharge for the 4 clusters determined by K-means clustering. Statistical testing in (c-d) was performed with a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ****$p < 0.0001$). (e) Proportion of different seasons within each cluster.

After plotting WCSS (Within-Cluster Sum of Square) with the varying K value (from 1 to 10), we chose 4 clusters (K=4) in our dataset as this is the point of inflection on the curve (Figure 5-6a). Samples of these 4 clusters are clearly separated apart from each other in the corresponding

PCA (Figure 5-6b). The first axis (PC1) explains 55% of the variance, with negative loadings from aromatic/terrestrial indicators (%C1, %C2, %C4, HIX and $SUVA_{254}$), and positive loadings from the optical parameters associated with autochthonous contribution (C3, C5, BIX, FI, and γ/α) and molecular weight (SR). This axis thus separates water samples with different allochthonous *vs.* autochthonous contributions. More specifically, all samples in Cluster 3 and most data in Cluster 1 are located on the left side of PC1. This suggests that Cluster 3 and Cluster 1 mainly group together samples with higher aromaticity. Indeed, we observe significantly higher values of aromatic/terrestrial proxies (%C1, %C2, %C4, HIX and $SUVA_{254}$) especially in Cluster 3 and to some extent in Cluster 1, as compared to Cluster 2 and Cluster 4 ($p<0.05$, Wilcoxon test; Figure 5-7). Most samples in Cluster 2 and Cluster 4 are located on the right side of PC1, which implies that these two clusters (especially for cluster 2) contain water samples with more autochthonous material with lower molecular weight. This is in line with the significantly higher values of autochthonous indicators (%C3, %C5, BIX, FI, and γ/α) in Cluster 2 and Cluster 4 compared to other clusters ($p<0.05$, Wilcoxon test; Figure 5-7). Hence, Cluster 3 and Cluster 2 likely represent the highest allochthonous and highest autochthonous contributions, respectively, while Cluster 1 and Cluster 4 represent intermediate levels of allochthonous and autochthonous contributions, respectively.

Even though all the clusters contain samples spanning across the estuary, clusters 1 and 3 contain a higher number of samples from lower section of the estuary, whereas clusters 2 and 4 integrate more samples from upper section of the estuary (Figure 5-6c). This is reflected in the average KP (Kilometric Point; distance in kilometers from the city of Paris), which is significantly higher in clusters 1 (310.1 km) and 3 (314.1 km) than in clusters 2 (290 km) and 4 (297 km) ($p<0.05$, Wilcoxon test; Figure 5-6c).

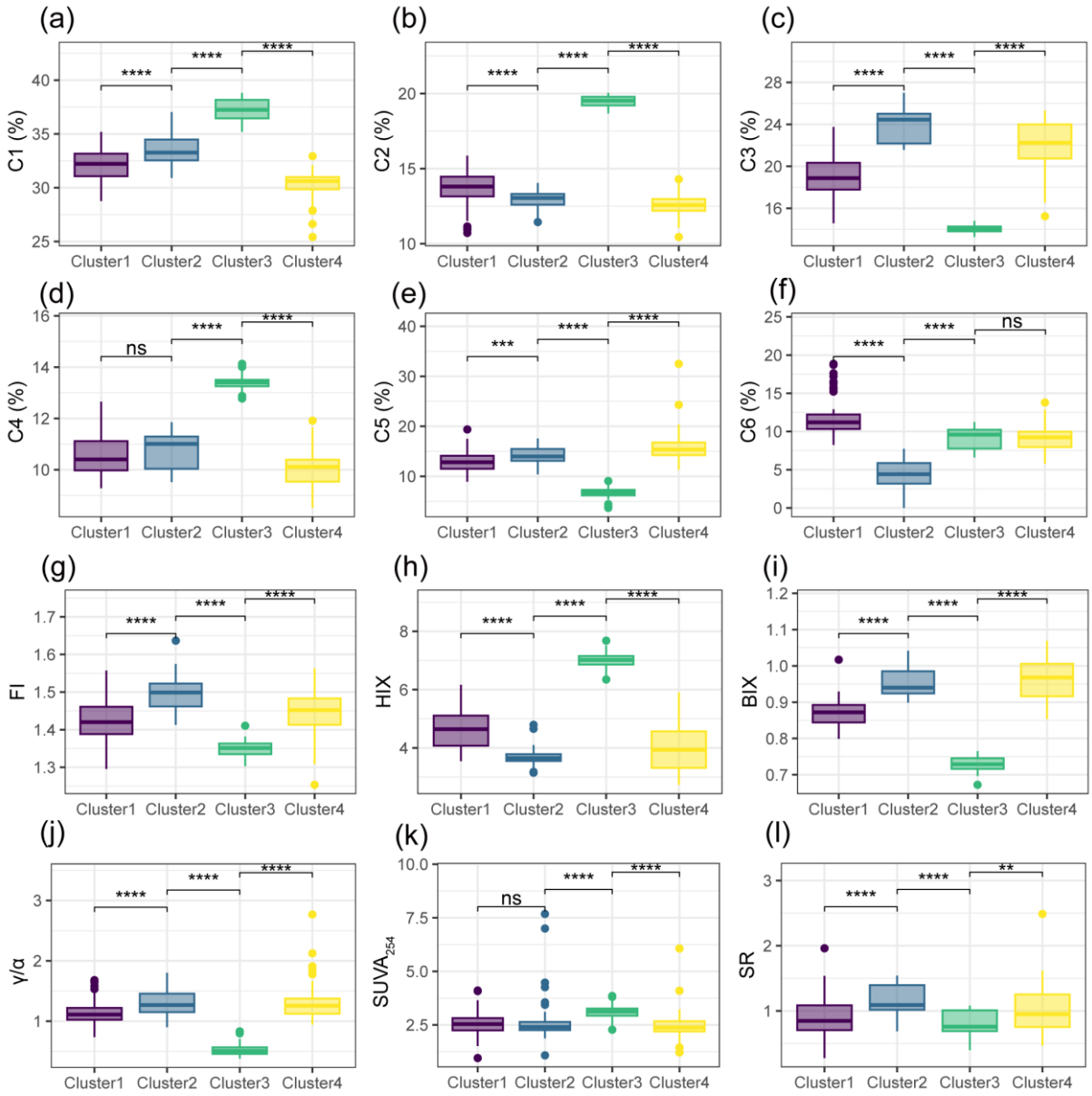



**Figure 5-7.** Box plots showing the distribution of DOM optical parameters within each cluster determined by K-means clustering. Statistical testing was performed using a Wilcoxon test (**$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

The different clusters also represent distinct hydrological conditions. Specifically, the average water discharge in Cluster 3 (572 $m^3/s$) and Cluster 1 (233.3 $m^3/s$) is significantly higher than that in Cluster 2 (145.7 $m^3/s$) and Cluster 4 (142.7 $m^3/s$) ($p<0.05$, Wilcoxon test; Figure 5-6d and Supplementary Table 5-2), which implies that cluster 3 represents high flow conditions,

whereas cluster 1 represents a mixture of low and high-water discharge conditions. As for clusters 2 and 4, they both represent low-flow conditions. Additionally, the clusters gather samples from distinct seasons (Figure 5-6e and Supplementary Table 5-2), i.e. mainly summer and winter for cluster 1, autumn and summer for cluster 2, only winter for cluster 3, and mainly summer and spring for cluster 4.

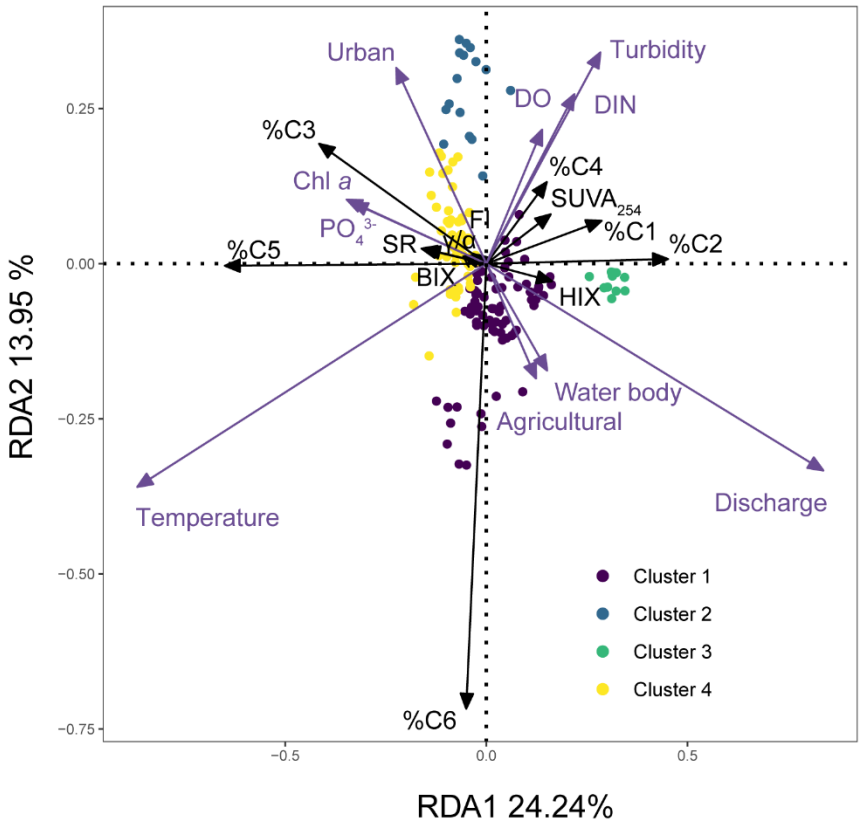



**Figure 5-8**. RDA analysis between available environmental variables and DOM optical parameters. Samples are colored according to clusters determined by K-means clustering. Physical parameters (turbidity, temperature, dissolved oxygen - DO), inorganic nutrients, and Chlorophyll *a* (Chl *a*) are measured by Serre-Fredj et al. (2023). Dissolved Inorganic Nitrogen (DIN) = $NO_3^-$-N + $NH_4^+$-N + $NO_2^-$-N.

Further insights into these clusters can be obtained by grouping samples using these 4 clusters in the RDA triplot (Figure 5-8). This highlights the relationship between phytoplankton biomass (Chl a) and autochthonous DOM parameters, especially %C3, which scored negatively on the first axis of the RDA, where Cluster 4 is located (Figure 5-8). Thus, Cluster 4 captures water

samples characterized by high contributions from biological/microbial origin, which may be linked to primary productivity. In addition, temperature is related with the transformation and (photo)degradation product (%C6) that scored negatively on the second axis of the RDA, where Cluster 1 is mainly located (Figure 5-8). This further suggests that Cluster 1 captures the waters associated with transformation and (photo)degradation processes.

The unsupervised machine learning approach is thus able to classify the DOM data into distinct clusters, each representing waters with unique DOM characteristics as well as hydrological and seasonal conditions. The interpretation of these clusters is summarized in Table 5-3.

**Table 5-3.** Interpretation of 4 clusters

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Discharge condition | low flow and high flow | low flow | high flow | low flow |
| Season | summer and winter | summer and autumn | winter | summer and spring |
| DOM characteristics | mainly allochthonous contribution as well as transformation and (photo)degradation products | highest autochthonous contribution | highest allochthonous contribution | mainly autochthonous contribution (linked to phytoplankton-related processes) |

We further explore the spatial dynamics of DOM optical properties for each cluster in the Seine Estuary, observing diverse behaviors of the optical parameters of DOM within these clusters from upstream to downstream (Figure 5-9). Notably, the spatial trend of %C3 (biological/microbial component) within Cluster 4 shows an initial increase followed by a decrease (Figure 5-9c). Given that Cluster 4 captures waters during periods of low flow in spring and summer (Table 5-3), the elevated %C3 values within this cluster could potentially be attributed to higher primary

productivity in this region during such low-flow conditions. Indeed, phytoplankton bloom occurs around this region (260<KP<340) particularly in low-flow periods, as shown in chapter 3 and chapter 4. Hence, the increased contributions from biological/microbial sources might be associated with phytoplankton-related processes. However, the other autochthonous component (%C5) within Cluster 4 shows a contrasting trend (Figure 5-9e), suggesting distinct sources and/or transformation processes for these components. In addition, within Cluster 1, %C5 and especially FI show an initial increase followed by a decreasing trend (Figure 5-9, e and g). This indicates higher contributions from protein-like component and microbial DOM in the aforementioned zone (260<KP<340) during both summer and winter, encompassing a combination of low and high-water discharge conditions. Furthermore, in this zone (260<KP<340), a significant decrease in %C2 and %C4 within Cluster 1 is observed (Figure 5-9, b and d), indicating lower contributions from terrestrial substances and/or DOM with bacterial origin. Given these pronounced spatial variations, this specific region (260<KP<340) can separate the estuary into three parts: Zone I (KP<260 km); Zone II (260 km<KP<340 km); Zone III (KP>340 km).

Overall, this unsupervised machine learning approach more efficiently traces DOM dynamics compared to traditional methods such as simply visualizing the data or grouping based on hydrological conditions. Notably, the unsupervised machine learning approach takes both hydrological conditions and seasonality into account, grouping together similar DOM samples, tracing pronounced spatial variabilities that cannot be captured by traditional methods. Estuarine zonation is further identified based on pronounced spatial variations of several parameters (%C3, %C5, %C2, and FI) especially in Cluster 1 and Cluster 4. However, identifying the primary DOM characteristics within each zone remains a challenge as distinct DOM parameters show different variabilities in each zone (Figure 5-9). In addition, the rationality of the defined zonation needs to be assessed.

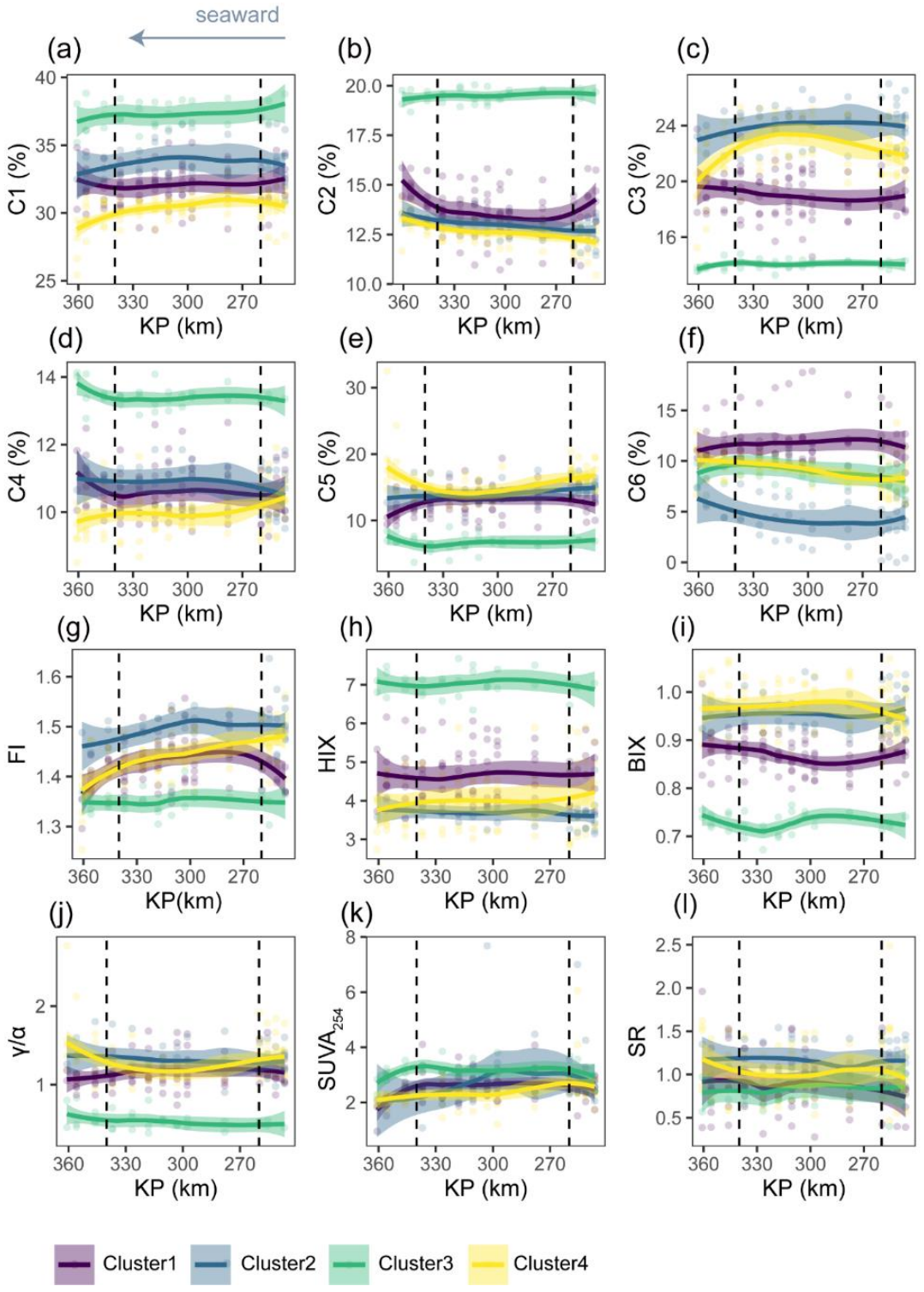



**Figure 5-9.** Spatial variations of DOM optical parameters for each of the clusters determined by K-means clustering. The trends showing spatial variations were according to locally estimated scatterplot smoothing (LOESS), with shaded area representing 95% confidence intervals. Samples (*n*=249) were grouped into 4 clusters determined by K-means clustering. Kilometric Point (KP) denotes the distance in kilometers from the city of Paris.

### 5.3.3 Rationality of the estuarine zonation evaluated by supervised machine learning

We combine samples from Cluster 1 and Cluster 3 to represent a high-flow scenario, which includes winter and flood periods in summer, as well as high-flow and a mixture of low and high-water discharge conditions (Table 5-3). In contrast, samples from Cluster 2 and Cluster 4 are combined to indicate a low-flow scenario, primarily involving spring, summer, and autumn, characterized by low flow (Table 5-3). Notably, this categorization takes into account both seasonality and hydrological conditions. We further evaluate the rationality of the defined zonation in these high-flow and low-flow scenarios.

To evaluate the rationality of the defined zonation in 5.3.2, we performed a supervised machine learning classification model to relate specific DOM optical compositions to each zone of the estuary in high-flow and low-flow scenarios. A cutting-edge ensemble machine learning classification method, named Light Gradient Boosted Machine (LightGBM) was evaluated in this study. The overall high scores for distinct parameters indicate the machine learning model could classify DOM optical properties as belonging to one of three zones with good performance (Supplementary Figure 5-7). We name the developed machine learning model as light Gradient Boosting Machine classification for DOM (GBM_DOM), further confirming the rationality of the defined estuarine zonation.

### 5.3.3 Explainable artificial intelligence and biogeochemical interpretations

We then used an explainable artificial intelligence approach, named SHAP (Shapley Additive exPlanations) analysis (Lundberg et al., 2020), to interpret the black box machine learning model (GBM_DOM) and evaluate the effect of input variables (DOM optical properties) on the model prediction (outcome in Zone I, Zone II, or Zone III). This would mean that such approach

enables us to identify the dominating DOM characteristics within each zone. The weight/importance of DOM parameters for GBM_DOM in high-flow and low-flow scenarios is sorted by SHAP values (Figure 5-10), with higher SHAP values indicating stronger positive feature importance (Lundberg et al., 2020).

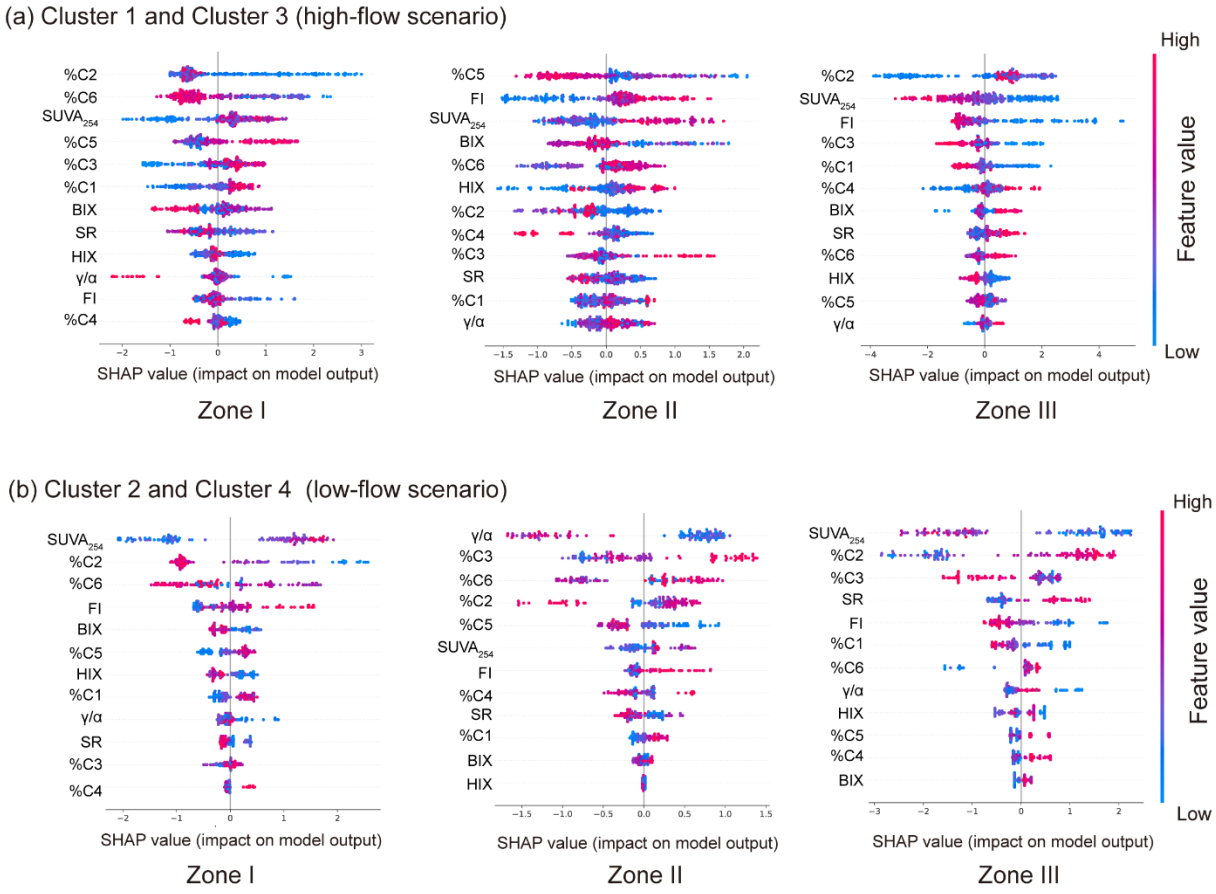



**Figure 5-10.** The ranking of feature importance for each zone for (a) Cluster 1 and Cluster 3 (high-flow scenario) and (b) Cluster 2 and Cluster 4 (low-flow scenario) based on LightGBM and SHAP library, with each bullet indicating a training example. The colorbar denotes the value of DOM optical parameters from low (blue) to high (pink).

The SHAP summary plots go into further detail about how each DOM parameter influences the model outcome in each zone. In these plots (Figure 5-10), the input variables (DOM optical properties) with higher impact on the model performance are shown at higher positions. The

colored bullets reflect the SHAP value of each sample in the training set, and their hue indicates the observed data value, ranging from low (blue) to high (pink). The bullets on the right side of the SHAP summary plot represent positive SHAP values (positive effect on the model output), whereas the bullets on the left side of the plot reflect negative SHAP values (negative effect on the model output).

Hence, variables positioned higher with pink bullets that have high SHAP values indicates a strong positive effect. In other words, these variables represent the main DOM characteristics within this zone for a specific scenario. For example, during the high-flow scenario, %C2 is at the top position in Zone III, with pink bullets having high SHAP values (Figure 5-10a), indicating the main DOM characteristics in this zone at high-flow scenario is %C2.

The ranking of the primary DOM characteristics and their interpretations in each zone for both high-flow and low-flow scenarios is then summarized in Table 5-4 and Table 5-5, respectively, taking into account the top 3 important variables with a positive effect.

**Table 5-4**. Ranking of main DOM characteristics in distinct zones in high-flow and low-flow scenario evaluated by GBM_DOM and SHAP library

| | Zone I | Zone II | Zone III |
|---|---|---|---|
| High-flow scenario | $SUVA_{254}$> %C5> %C3 | FI> $SUVA_{254}$> %C6 | %C2> %C4> BIX |
| Low-flow scenario | $SUVA_{254}$> FI> %C5 | %C3> %C6> %C2 | %C2> SR> %C6 |

During both high-flow and low-flow scenarios, Zone I is mainly contributed by $SUVA_{254}$ (Figure 5-10 and Table 5-4). This indicates that CDOM in the upper estuary (Zone I) is characterized by higher aromatic content (high $SUVA_{254}$) regardless of the seasonality. Interestingly, autochthonous contribution is also prominent in this zone, as higher FI, %C3, and %C5 positively impact the model output for this zone (Figure 5-10 and Table 5-4, 5-5). Such

information is less clearly captured through traditional approaches (i.e., simple visualization; Figure 5-3) or by unsupervised machine learning (Figure 5-9), which might be masked by complex hydrological conditions and/or the mixing of land use types within this highly dynamic area, along with the large number of parameters that need to be considered. By using machine learning and explainable artificial intelligence, we can effectively uncover hidden DOM signatures, identifying the spatial specificity of distinct DOM characteristics and potential controlling factors. In the Neuse River Basin, agricultural and urban land use were closely linked to higher proportions of high molecular weight and autochthonous DOM, respectively (Bhattacharya and Osburn, 2020). Hence, in the upper Seine Estuary (Zone I), simultaneous observation of high aromatic and high proportions of autochthonous material could be explained by significant contributions from both soil erosion and anthropogenic inputs (i.e. domestic and/or urban effluents). The high contributions from %C3 and especially %C5 may be linked to the discharge from urban wastewater treatment plants, which increases biological activity in the aquatic environment, leading to an increase in the proportion of proteinaceous and autochthonous compounds. Indeed, this zone is characterized by relatively higher agricultural and urban land use (Figure 5-1; site 10-14). In addition to the influence of the Seine watershed itself, the mixed agricultural and urban land use types in this area introduces extra complexity in DOM characterization. Monitoring and characterizing the variations in absorbance ($SUVA_{254}$) and florescence (FI, %C3, and %C5) indicators within this zone can thus provide insights into the dynamics of DOM with distinct sources.

**Table 5-5**. Interpretation of DOM characteristics in distinct zones in high-flow and low-flow scenario

| | Zone I | Zone II | Zone III |
|---|---|---|---|
| High-flow scenario (Cluster 1 + Cluster 3) | Dominated by aromatic material; high contribution from protein-like substances, followed by DOM with biological/microbial origin | Dominated by DOM with microbial origin; high contribution from aromatic material, followed by transformation and (photo)degradation products | Dominated by aromatic compounds that are subject to photodegradation; High contribution from DOM with bacterial origin or terrestrial origin in agricultural area, followed by DOM with biological/microbial origin |
| Low-flow scenario (Cluster 2+ Cluster 4) | Dominated by aromatic material; high contribution from DOM with microbial origin, followed by protein-like substances | Dominated by DOM with biological/microbial origin; high contribution from transformation and (photo)degradation products, followed by aromatic compounds that are subject to photodegradation | Dominated by aromatic compounds that are subject to photodegradation; High contribution from low-molecular-weight DOM, followed by transformation and (photo)degradation products |

The DOM properties in zone II (mid-estuary) in the high-flow scenario are mainly influenced by FI, whereas in the low-flow scenario are predominantly influenced by %C3 (Figure 5-10 and Table 5-4). This suggests that DOM in this zone is generally dominated by DOM with biological/microbial origin. As shown in chapter 3 and chapter 4, a phytoplankton bloom was observed in this zone as indicated by high Chl *a* concentration. During phytoplankton bloom demise, viral infection could be significantly intensive, further releasing phytoplankton-derived metabolites (autochthonous DOM) into the water column (Kuhlisch et al., 2021). This process was termed the "viral shunt", which links primary production with the DOM cycling (Fuhrman, 1999;

Wilhelm and Suttle, 1999). In addition to the viral shunt, autochthonous DOM can also be produced directly by phytoplankton and microbial degradation. Therefore, the release and subsequent transformation of phytoplankton-derived DOM could explain the dominant contributions of autochthonous DOM in Zone II, which potentially documents the footprint of the phytoplankton-related processes. Additionally, it is likely that agricultural land use contributes to the export of soil-derived DOM, leading to an increase in aromatic content of DOM in adjacent aquatic environments (Chen et al., 2021). Hence, the elevated signatures of aromatic CDOM (high $SUVA_{254}$; Figure 5-10 and Table 5-4, 5-5) during the high-flow scenario can also be attributed to the high portions of agricultural land use in this zone (Figure 5-1). In addition, the abundant terrestrial/aromatic DOM in this region may further undergo photobleaching (Ishii and Boyer, 2012). Indeed, significant photochemical degradation processes are evidenced by positive contributions of photodegradation products (%C6) in Zone II (Figure 5-10 and Table 5-4, 5-5). Overall, the use of the machine learning and explainable artificial intelligence suggests that transformation processes (both microbial and photochemical processing) and land use types control the DOM characteristics.

In the lower section of the Seine Estuary (zone III), the DOM properties are identical to typical marine environments, with low molecular weight CDOM (high SR) and high autochthonous contribution (high BIX) (Figure 5-10 and Table 5-4, 5-5). The elevated water discharge has the potential to flush nutrient-rich waters into this zone, which could potentially trigger phytoplankton blooms in this area, as demonstrated in chapter 4. Such processes could also result in the release of substantial amounts of autochthonous DOM. In addition, DOM characteristics in this zone are mainly influenced by %C2 both in high-flow and low flow scenarios (Figure 5-10 and Table 5-4). Interestingly, this component has an opposite influence on the model prediction compared to the other terrestrial component (%C1). %C1 shows negative contributions for this zone in both

scenarios (Figure 5-10). This suggests that %C1 and %C2, both previously categorized as terrestrial aromatic components, actually show different behaviors in Zone III. One possible explanation for such a distinction is that these components are influenced by varying levels of photodegradation. Previous studies have shown that only C2 can be produced through photochemical processes, whereas C1 cannot (Ishii and Boyer, 2012). The positive contribution of %C2 in this zone thus implies the possible accumulation of photochemically produced material in this area. Distinct behavior of terrestrial fluorescent components in this area highlights the complex sources and transformation processes of estuarine DOM, which could be assessed by machine learning and explainable artificial intelligence.

Overall, a generalized estuarine zonation in this study typically consists of 3 zones, including the upper (Zone I), mid (Zone II), and lower (Zone III) estuary, with each zone showing specific DOM characteristics and biogeochemical processes in high-flow and low-flow scenarios (Figure 5-11). Our model suggests that DOM in the Seine Estuary is dominated by aromatic material and autochthonous contribution in the upper estuary (KP<260; Figure 5-11). In the mid-estuary (260<KP<340), the main contribution to DOM comes from autochthonous sources as well as aromatic material, suggesting enhanced transformation (microbial and photochemical) processes. Subsequently, a transition to photochemically produced material, low molecular weight, and varying portions of autochthonous DOM in the lower estuary (KP>340) is observed, indicating the significant influence of marine water mass on DOM properties and other processes such as photodegradation, flocculation and precipitation. Our results demonstrate that the estuarine DOM originates from distinct sources and undergoes varying levels of in-estuary processing within specific zones. Machine learning is shown to be a powerful approach to disentangle the DOM complexity.

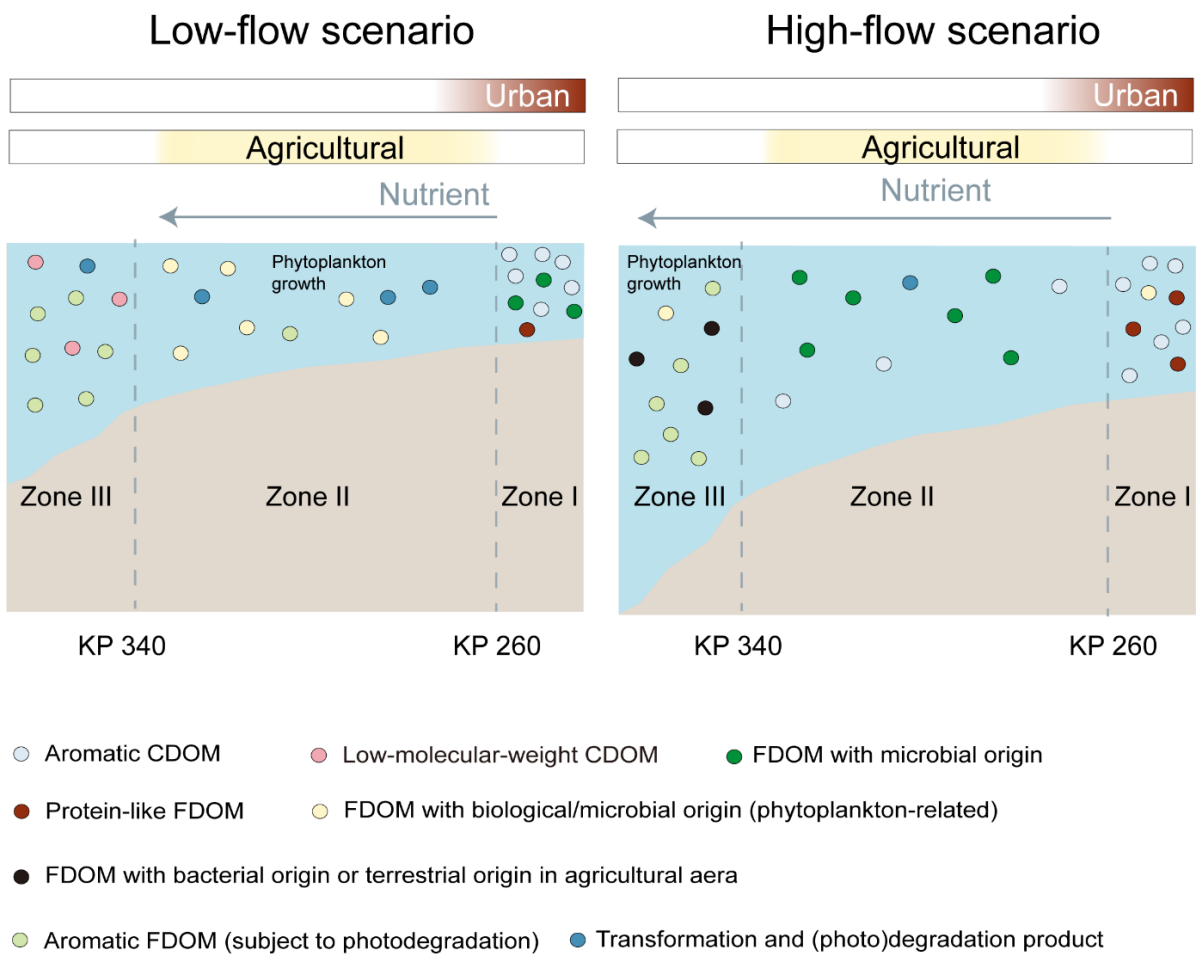



**Figure 5-11.** Schematic diagrams showing the zonation of the Seine Estuary in terms of DOM dynamics in low-flow and high-flow scenarios. Kilometric Point (KP) denotes the distance in kilometers from the city of Paris.

## 5.4. Conclusions and environmental implications

Combining unsupervised and supervised machine learning can provide novel insights into disentangling the DOM composition related to both seasonal and spatial variations in estuaries. Specifically, by applying unsupervised machine learning, groups of samples characterized by comparable DOM optical parameters, can be identified, providing an initial understanding of the DOM dynamics and zonation. Thereafter, supervised machine learning can validate and evaluate the rationality (i.e. according to accuracy in the classification results) of the generalized zonation

defined from unsupervised learning. The application of explainable artificial intelligence can further help to identify the dominating DOM parameters within each zone. The approach proposed in this study can be easily applied to other systems beyond estuaries, such as rivers, lakes, or coastal oceans. The established workflow can significantly contribute to environmental management and decision-making processes, which may lead to more sustainable and effective policies.

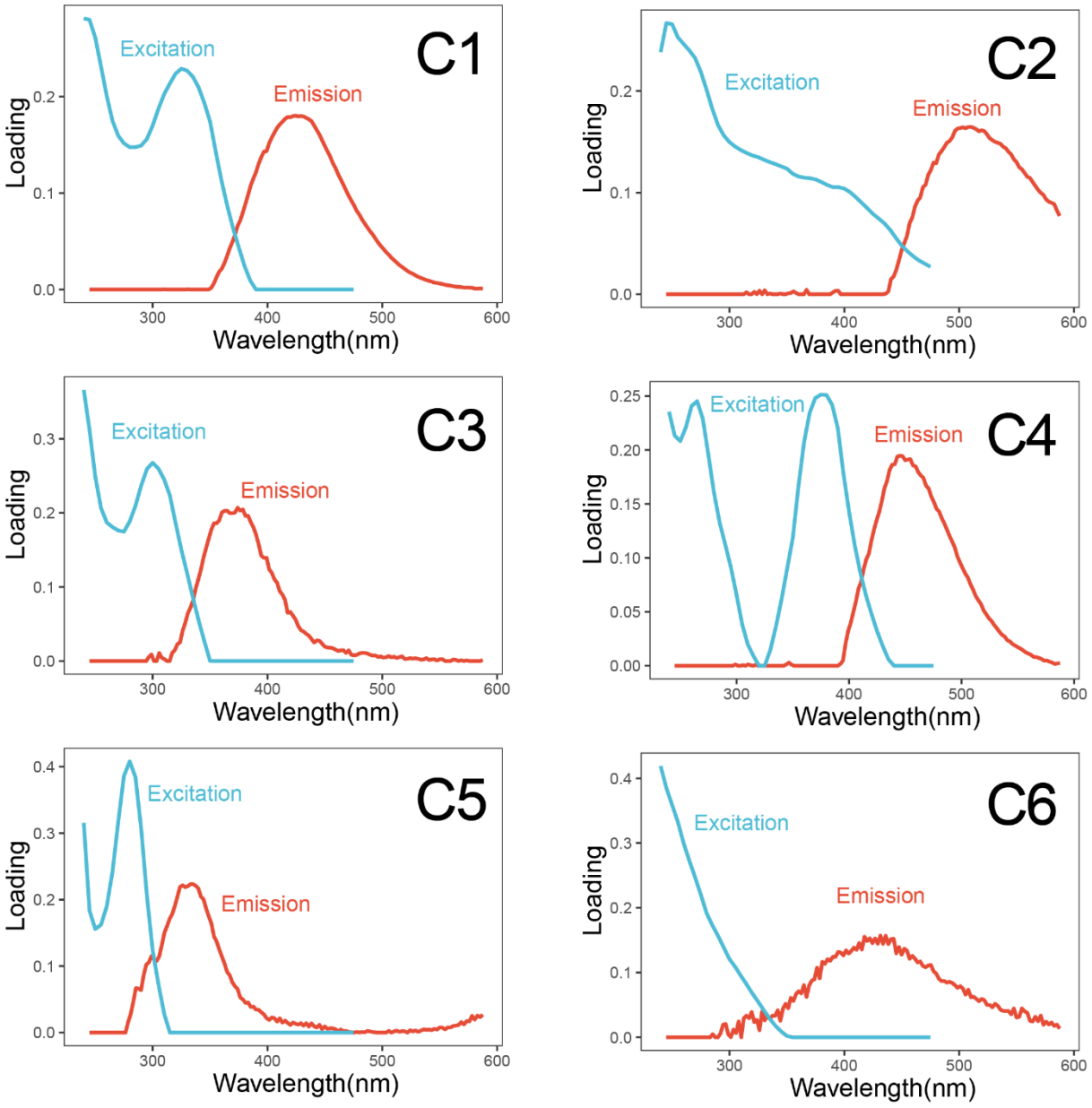
## 5.5. Annexes

**Supplementary Table 5-1.** RDA results

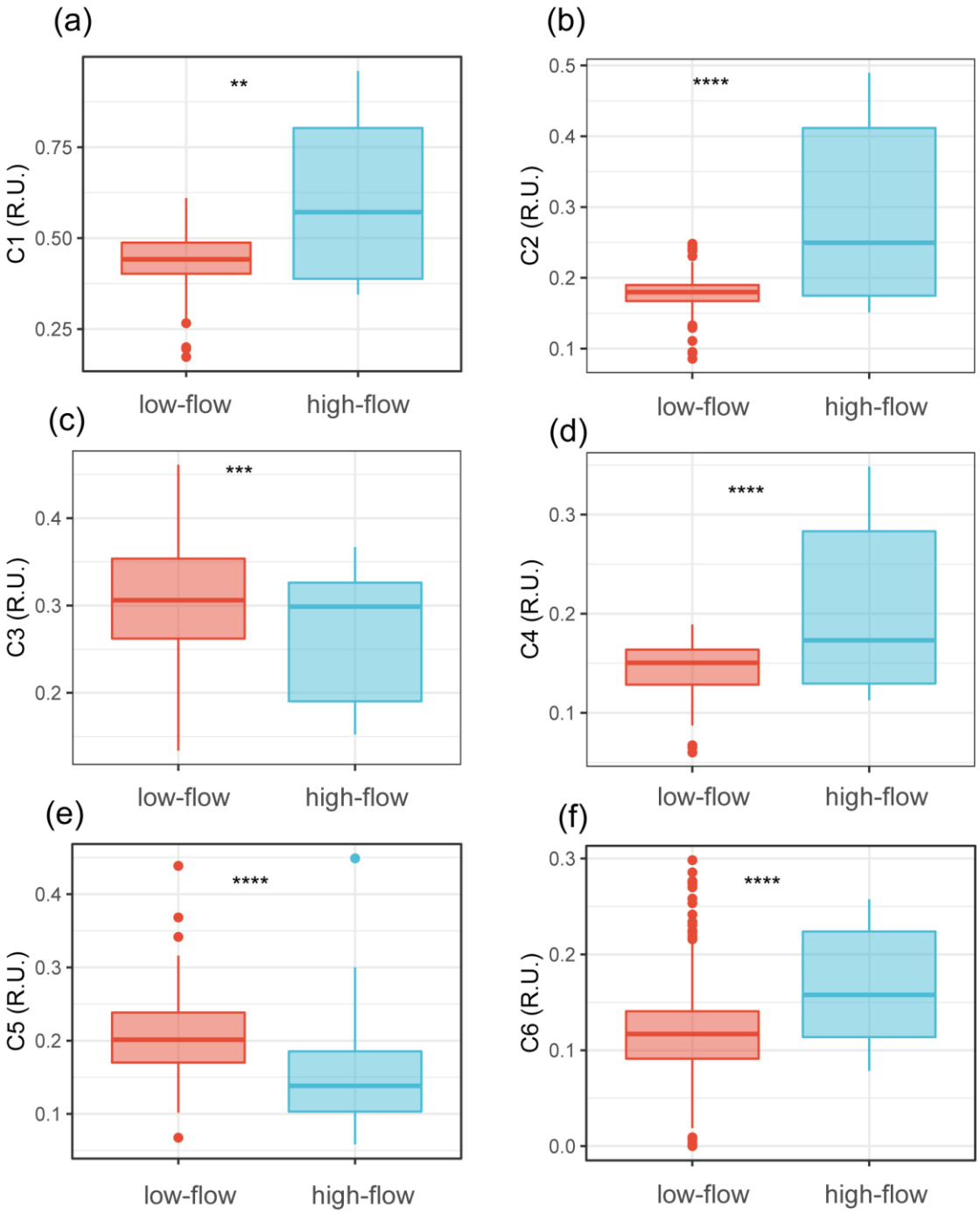
| Variables | Axis 1 | Axis 2 | Individual Importance (%) |
|---|---|---|---|
| Temperature | -0.87 | -0.36 | 13.8 *** |
| Discharge | 0.84 | -0.33 | 13.15*** |
| Turbidity | 0.28 | 0.34 | 2.55** |
| $PO_4^{3-}$ | -0.33 | 0.10 | 1.97* |
| Chl *a* | -0.35 | 0.10 | 1.48* |
| Urban | -0.22 | 0.32 | 1.43* |
| DIN | 0.22 | 0.27 | 1.36* |
| DO | 0.14 | 0.22 | ns |
| Agricultural | 0.12 | -0.19 | ns |
| Water body | 0.15 | -0.17 | ns |

**Supplementary Table 5-2.** Description of the 4 clusters

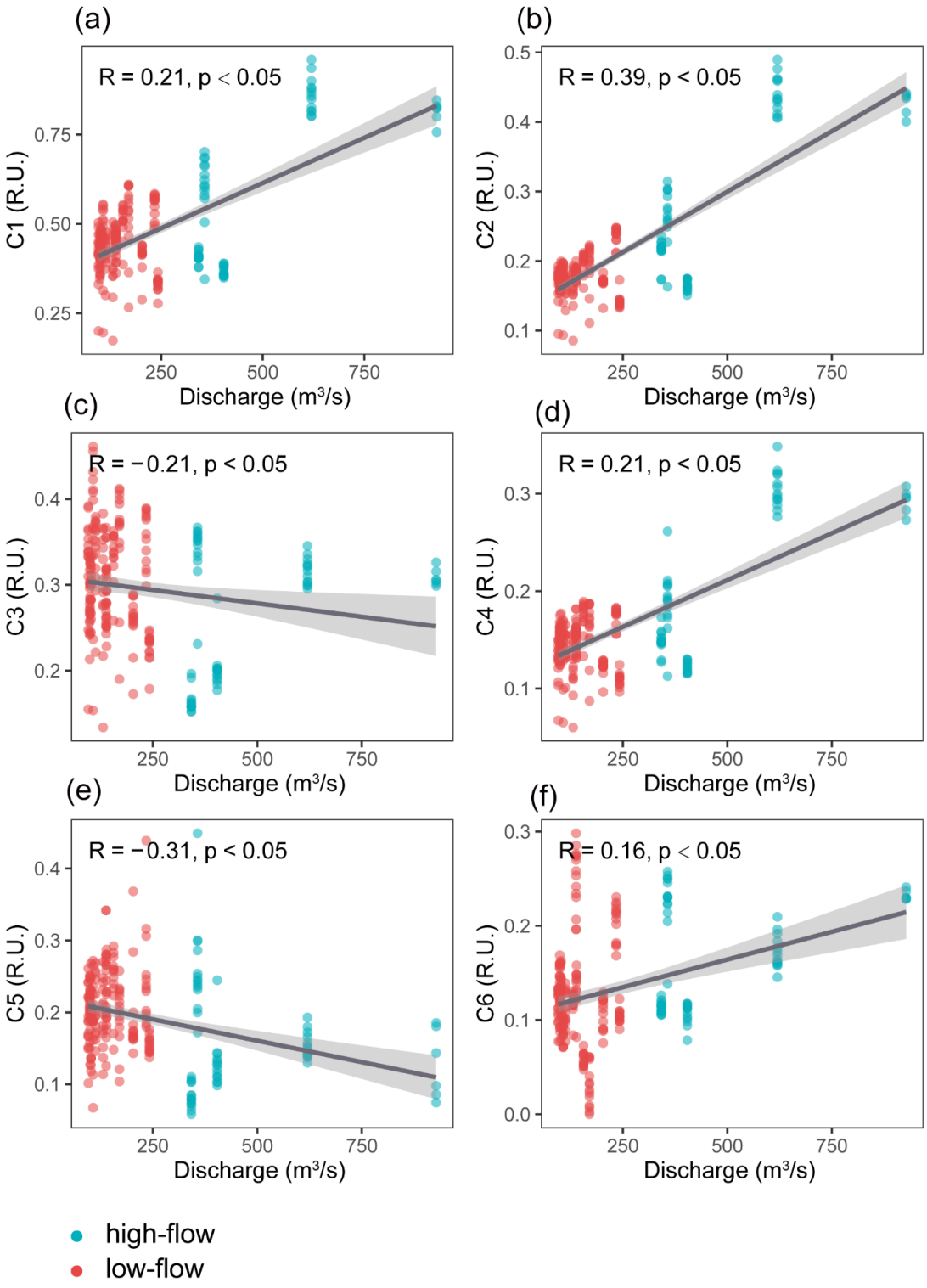
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Number of samples (total) | 82 | 46 | 31 | 90 |
| Number of samples (spring) | 2 | 0 | 0 | 25 |
| Number of samples (summer) | 65 | 19 | 0 | 64 |
| Number of samples (autumn) | 0 | 27 | 0 | 0 |
| Number of samples (winter) | 15 | 0 | 31 | 1 |
| | | | | |
| Min KP (km) | 246.6 | 246.6 | 246.6 | 243 |
| Max KP (km) | 360.8 | 360.8 | 360.8 | 360.8 |
| Mean KP (km) | 310.1 | 290 | 314.1 | 297 |
| | | | | |
| Min Discharge ($m^3/s$) | 99 | 95.6 | 342 | 95.6 |
| Max Discharge ($m^3/s$) | 404 | 203 | 928 | 404 |
| Mean Discharge ($m^3/s$) | 233.3 | 145.7 | 572 | 142.7 |



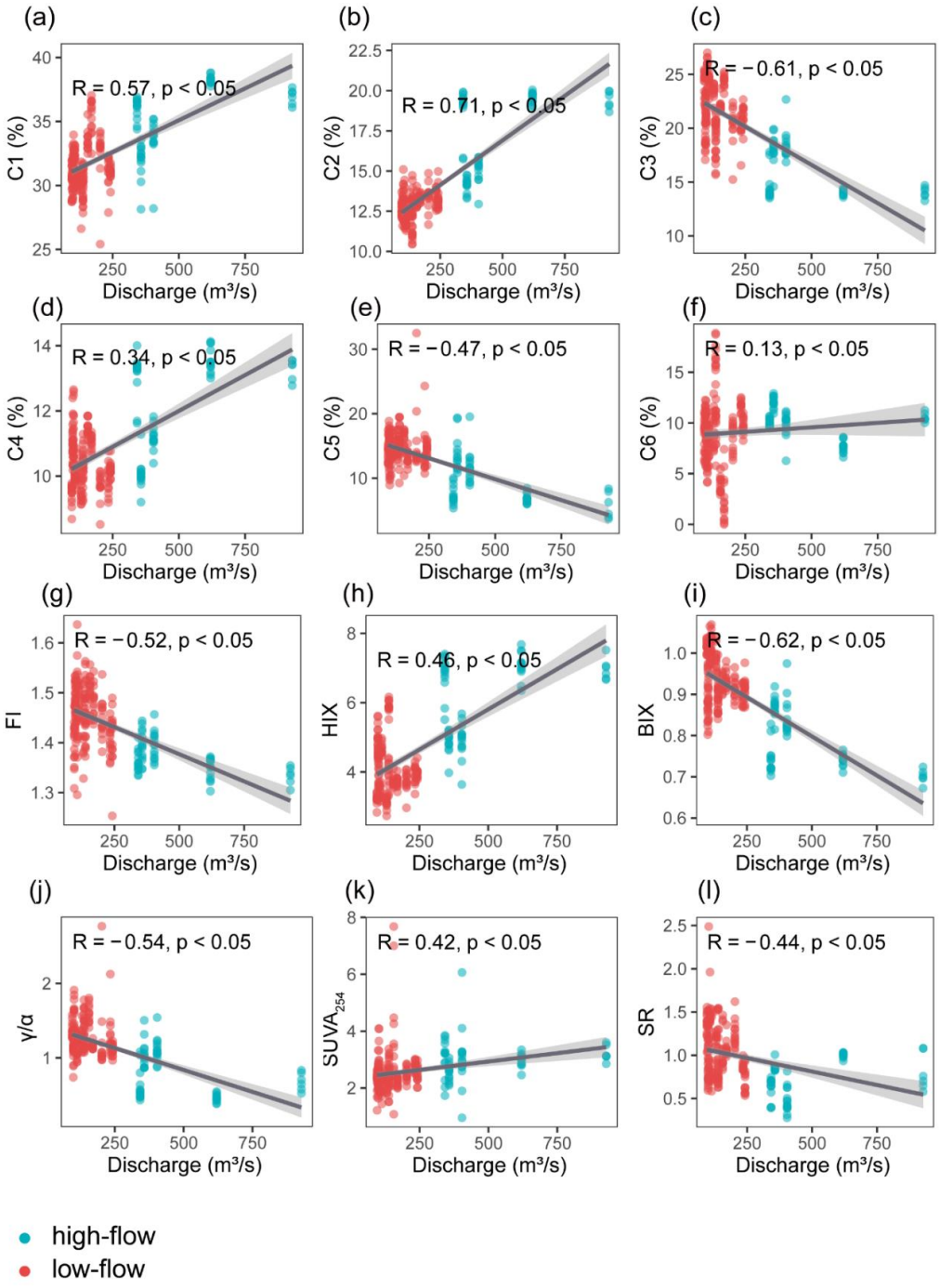

**Supplementary Figure 5-1**. Spectral characteristics of the six components determined by PARAFAC.



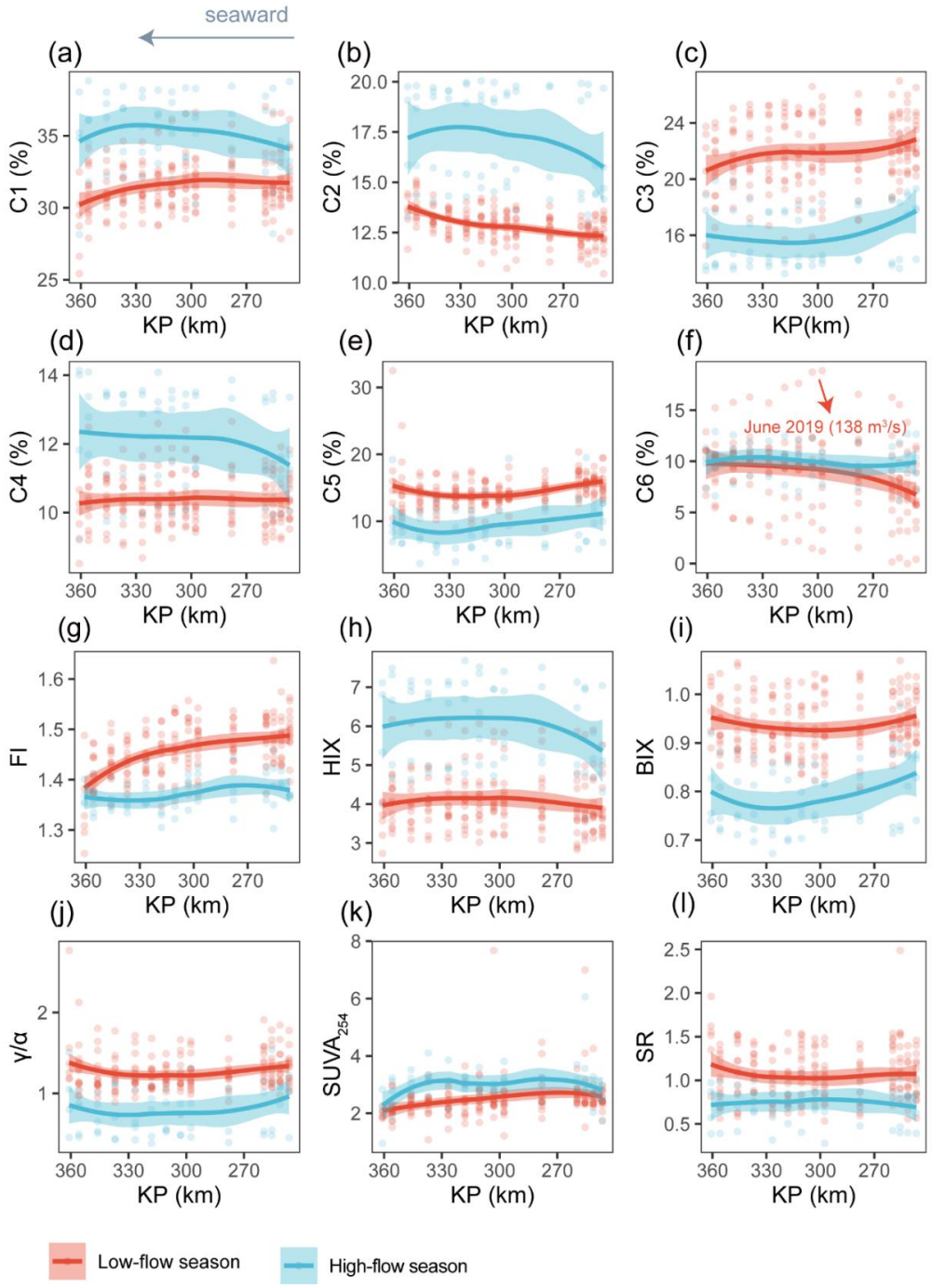

**Supplementary Figure 5-2**. Box plots showing the distribution of the fluorescence intensity (in Raman Units – R.U.) of the six PARAFAC components between high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) periods. Statistical testing was performed using a Wilcoxon test (**$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$).



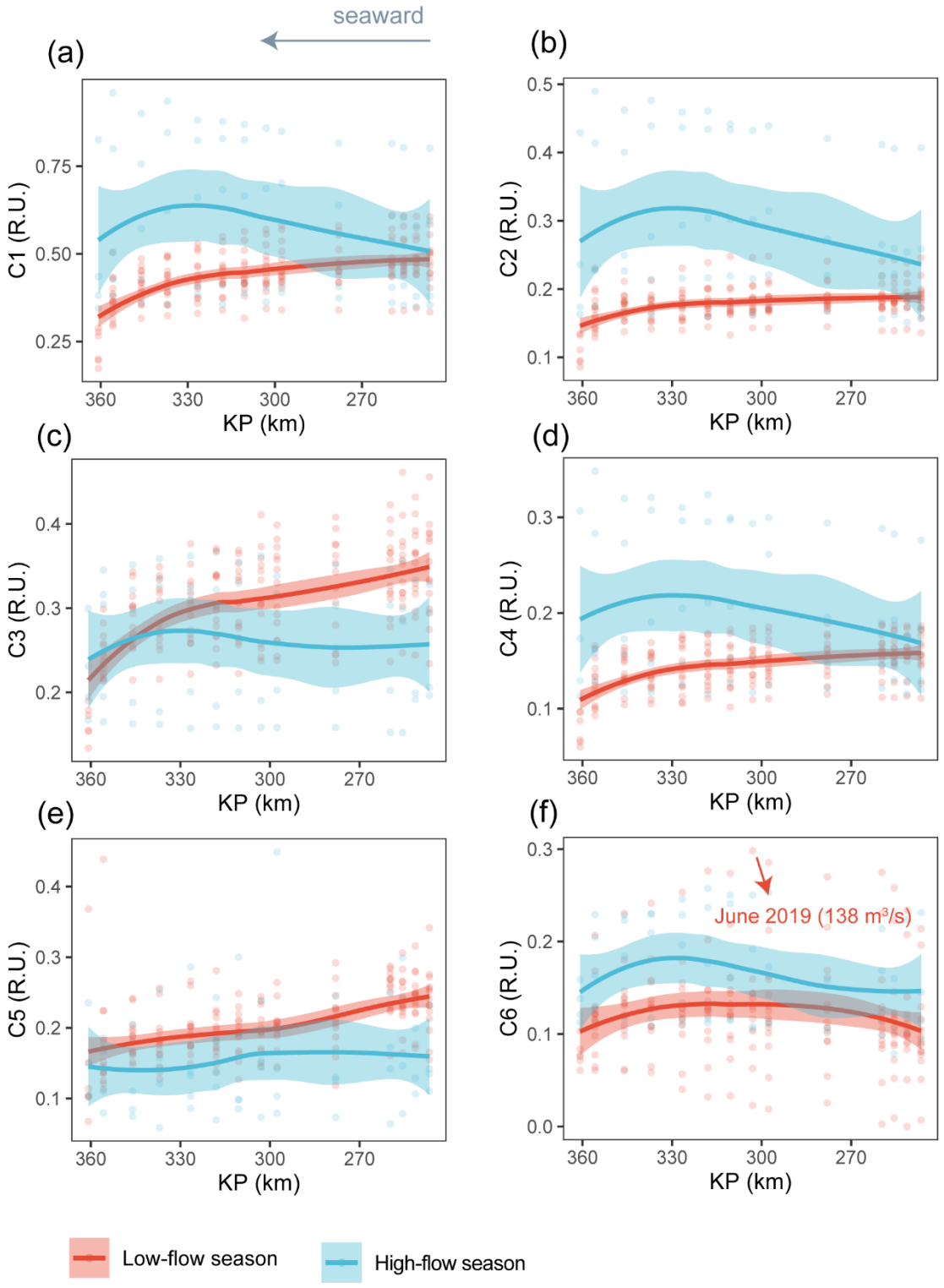

**Supplementary Figure 5-3**. Mean monthly water discharge plotted against the fluorescence intensity (in Raman Units – R.U.) of the six PARAFAC components, with shaded region representing 95% confidence intervals (Spearman's correlation). Samples (*n*=249) were colored by high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) periods.



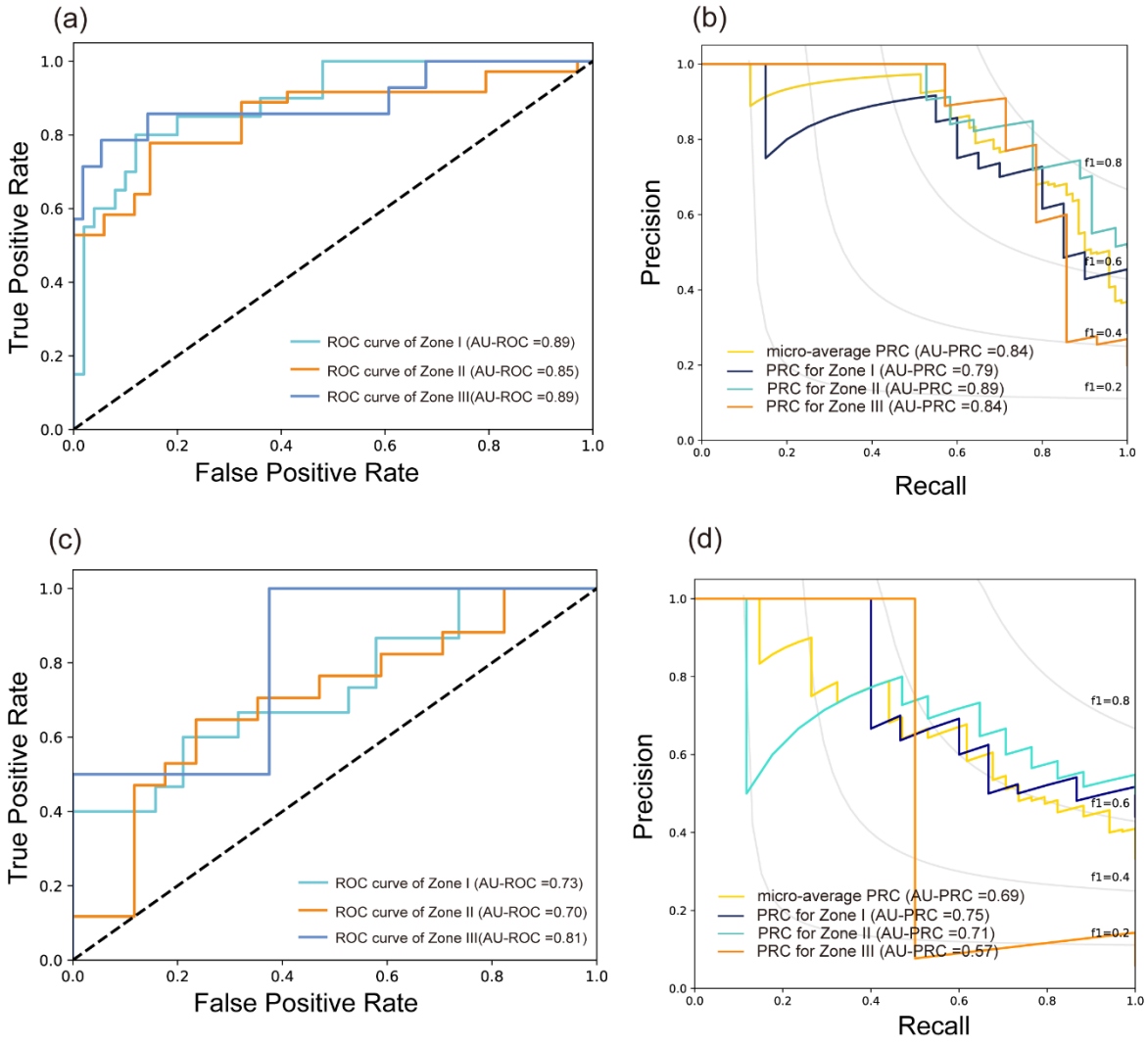

**Supplementary Figure 5-4**. Mean monthly water discharge plotted against distinct DOM optical parameters, including the relative percentage of the six PARAFAC components: (a) C1, (b) C2, (c) C3, (d) C4, (e) C5, (f) C6; the fluorescence indices (g) fluorescence index – FI, (h) humification index – HIX, (i) biological index – BIX, (j) fluorescence intensity ratio γ/α; and the absorbance indices (k) specific UV absorbance - $SUVA_{254}$, (l) spectral slope ratio – SR, with shaded region representing 95% confidence intervals (Spearman's correlation). Samples (*n*=249) were colored by hydrological conditions, including high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) periods.




**Supplementary Figure 5-5**. Spatio-temporal variations of DOM optical parameters, including the relative percentage of the six PARAFAC components: (a) C1, (b) C2, (c) C3, (d) C4, (e) C5, (f) C6; the fluorescence indices (g) fluorescence index – FI, (h) humification index – HIX, (i) biological index – BIX, (j) fluorescence intensity ratio γ/α; and the absorbance indices (k) specific UV absorbance - $SUVA_{254}$, (l) spectral slope ratio – SR. The trends showing spatial variations were according to locally estimated scatterplot smoothing (LOESS), with shaded area representing 95% confidence intervals. Samples (*n*=249) were grouped into high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) periods. Kilometric Point (KP) denotes the distance in kilometers from the city of Paris.




**Supplementary Figure 5-6**. Spatio-temporal variations of fluorescence intensities of PARAFAC components (in Raman Unit -R.U.). The trends showing spatial variations were according to locally estimated scatterplot smoothing (LOESS), with shaded area representing 95% confidence intervals. Samples (*n*=249) were grouped into high-flow (>250 $m^3/s$ - blue) and low-flow (<250 $m^3/s$ - red) periods. Kilometric Point (KP) denotes the distance in kilometers from the city of Paris.




**Supplementary Figure 5-7**. (a-b) Receiver Operating Characteristic curve (ROC curve) (a) and the Precision-Recall Curve (PRC) (b) evaluated by LightGBM for the high-flow scenario (Cluster 1 + Cluster 3). (c-d) ROC curve (c) and PRC (d) evaluated by LightGBM for the low-flow scenario (Cluster 2 + Cluster 4).

# Chapter 6:

Synthesis and perspectives

This PhD thesis aims to assess the estuarine functioning by simultaneously investigate the dynamics of Particulate Organic Matter (POM) and Dissolved Organic Matter (DOM). To this aim, the spatio-temporal variations of POM and DOM characteristics were determined using water samples (*n*=383) collected along the land-sea continuum of a human-impacted estuary (Seine Estuary, France) during 24 sampling campaigns (June 2019 to November 2022).

## 6.1. Development of a novel riverine runoff proxy

This thesis starts by investigating the POM dynamics in the Seine River basin. To date, reliable proxies for quantifying the relative contribution of terrestrial organic matter in aquatic systems are still lacking. Current available terrestrial proxies, such as the $\delta^{13}C$ of organic carbon (Lamb et al., 2006), Branched and Isoprenoid Tetraethers (BIT) index (Hopmans et al., 2004) and long-chain diols (Lattaud et al., 2017) have their own uncertainties and limitations. Additional molecular proxies for riverine runoff are thus needed, which may cross-validate available ones.

In Chapter 3, the POM dynamics was investigated at the bulk and molecular levels, through elemental and isotopic analyses as well as lipid biomarkers (i.e. branched glycerol dialkyl glycerol tetraethers, brGDGTs; branched glycerol monoalkyl glycerol tetraethers, brGMGTs). Both types of compounds can be produced *in situ* in water column and/or sediments in aquatic settings. Both of their distributions are strongly correlated to salinity, whereas only brGDGT distributions are significantly influenced by nitrogen nutrient loadings. Salinity correlates positively with two brGMGT homologues (H1020a and H1020b), and negatively with the other two brGMGTs (H1020c and H1034b), which leads to the development of a novel molecular proxy (Riverine IndeX, RIX), with higher RIX indicating more riverine contribution.

In addition, a specific zone in the Seine estuary (260<KP<340; KP: kilometric point, the distance in kilometers from the city of Paris) is identified, which is characterized by strong

phytoplankton productivity and heterotrophic bacteria activity, particularly during the low-flow period. This further shows the potential for identifying estuarine zonation based on POM dynamics.

## 6.2. Dynamics of different types of POM and their relationships with land use and hydroclimate conditions

Based on the results presented in chapter 3, the aforementioned zone (260<KP<340) is further hypothesized to relate with land use changes and hydroclimate conditions. To test this, the spatio-temporal variations of distinct types of POM (i.e. anthropogenic POM, phytoplankton-derived POM, and plant-derived POM) are investigated using bulk geochemical analysis as well as complementary lipid biomarkers (sterols, stanols, fatty acids, and *n*-alkanes) in chapter 4.

It is demonstrated how the dispersion and dynamics of various types of POM are linked to hydroclimate conditions and land use patterns. Specifically, anthropogenic POM (indicated by a sewage proxy) has a positive correlation with water discharge and urban land use. The sewage indicator gradually decreases along the estuary, suggesting a dilution of sewage contamination during the mixing of water masses. In addition to the spatial variability, anthropogenic POM also shows seasonal variations, with greater sewage contamination at high flows.

Furthermore, phytoplankton blooms (indicated by a diatom biomarker and a proxy for phytoplankton biomass - Chlorophyll *a*) are observed in the aforementioned zone (260<KP<340), which represents an agriculturally impacted region. Intense agricultural activities may release large amounts of nutrients that can be assimilated by phytoplankton at low flows, triggering phytoplankton blooms. During high-flow season, nutrients from agricultural activities can be transported further downstream (Xia et al., 2020), fueling phytoplankton growth in the coastal waters.

The plant-derived POM is mostly contributed by aquatic plants at high flows and a mix of terrestrial and aquatic plants at low flows. In addition, microbial degradation processes and potential priming effect may occur especially at that agriculturally impacted region (260<KP<340).

As a result, this region (260<KP<340), where phytoplankton blooms occur during low flows, could divide the estuary into three distinct zones, each showing unique characteristics under high-flow and low-flow scenarios: Zone I (KP<260), Zone II (260<KP<340), and Zone III (KP>340). Samples from these zones are distinguished and separated well by Principal Component Analysis (PCA) based on POM parameters, land use characteristics, and hydroclimate conditions.

## 6.3. Disentangling DOM composition by machine learning

Chapter 3 and Chapter 4 present a zonation of the Seine Estuary regarding POM dynamics. To explore DOM dynamics and its related estuarine zonation/functioning, DOM properties are further investigated by UV–Visible absorbance and Excitation-Emission Matrix fluorescence spectroscopy in Chapter 5. The DOM parameters are firstly visualized in a contour plot and grouped by hydrological conditions. However, it remains difficult to capture the spatial trends in DOM composition along the Seine estuary by these approaches.

The potential (temporal and spatial) variability of DOM is then explored by using unsupervised machine learning. This approach effectively captures DOM variability, identifying three estuarine zones (Zone I (KP<260), Zone II (260<KP<340), and Zone III (KP>340)) based on pronounced spatial variations in several DOM optical parameters (relative abundances of PARAFAC components C2 (terrestrial origin), C3 (microbial/biological origin), C5 (protein-like substances), and fluorescence index FI), particularly within two clusters. Thus, the same three functional zones of the Seine Estuary are identified by analyzing DOM and POM data separately.

Thereafter, the rationality of the defined zonation and main DOM characteristics within each zone are assessed by supervised machine learning as well as explainable artificial intelligence. This led to the development of a novel model (light Gradient Boosting Machine classification for DOM, GBM_DOM). This model successfully disentangles the DOM composition and captures main DOM characteristics in different zones of the Seine Estuary. Our model shows that aromatic material dominates DOM characteristics, with an autochthonous contribution in the upper estuary (KP<260). The predominant contribution to DOM in the mid-estuary (260<KP<340) originates from autochthonous sources as well as aromatic material, implying significant transformation (microbial and photochemical) processes. Following that, a change to photochemically produced material, low molecular weight, and autochthonous DOM is observed in the lower estuary (KP>340), suggesting a considerable influence of marine water masses and other processes such as photodegradation.

## 6.4. Estuarine functioning in terms of POM and DOM dynamics

Based on the results presented in chapters 3 to 5, a synthesized diagram showing POM and DOM dynamics, as well as associated estuarine zonation is shown in Figure 6-1.

Zone I (KP<260; Figure 6-1) is characterized by intensive urban land use and significant contributions from terrestrial (riverine) POM, anthropogenic POM, aromatic CDOM, and protein-like FDOM in both high-flow and low-flow scenarios. This could be explained by considerable contributions from both soil erosion and anthropogenic inputs (i.e. residential and/or municipal effluents).

Zone II (260<KP<340; Figure 6-1) is characterized by high portions of agricultural land use. This zone is especially contributed by phytoplankton-derived POM, FDOM with biological/microbial origin (phytoplankton-related) and FDOM produced by transformation and

(photo) degradation, notably at the low-flow scenario. Intense agricultural activity in this zone may result in substantial amounts of organic fertilizers, manure, as well as urban and industrial wastewater being released into the waters. The residence time of the water masses would increase during the low-flow season, potentially extending nutrient retention. As a result, the nutrient can be extensively assimilated by phytoplankton, triggering subsequent phytoplankton blooms as reflected by accumulation of phytoplankton-derived POM. The enhanced autochthonous DOM contributions in this zone at low flows might be associated with the presence of phytoplankton-derived POM. During phytoplankton bloom demise, viral infection and microbial degradation could release phytoplankton-derived autochthonous DOM into the water column. During high-flow scenario, this zone is characterized by high levels of anthropogenic POM, terrestrial (riverine) POM, aquatic plant-derived POM, FDOM with microbial origin, and aromatic CDOM that are likely derived from the upstream region.

Zone III (KP>340; Figure 6-1) is representative of costal environments with less contributions from anthropogenic POM, terrestrial (riverine) POM, aromatic CDOM, and protein-like FDOM inputs. Instead, DOM characteristics in this zone are mainly influenced by photochemically produced material both in high-flow and low flow scenarios. During the high-flow scenario, the increased water discharge effectively flushes nutrient-rich waters into this zone, thus leading to phytoplankton blooms and accumulation of phytoplankton-derived POM and FDOM with biological/microbial origin.

Estuaries thus act as effective filters/buffers that dilute anthropogenic POM and protein-like FDOM that likely originate from upstream regions characterized by significant portions of urban land use. Such functioning is more pronounced during the high-flow scenario.

On the other hand, estuaries play a crucial role as biogeochemical reactors, especially in low-flow scenarios, stimulating the growth of phytoplankton and accumulating phytoplankton-

derived POM and FDOM within an agriculturally impacted region (Zone II; Figure 6-1). Transformation processes may also play a key role in this zone during low-flow scenario as reflected by substantial contributions from transformation and (photo) degradation product.

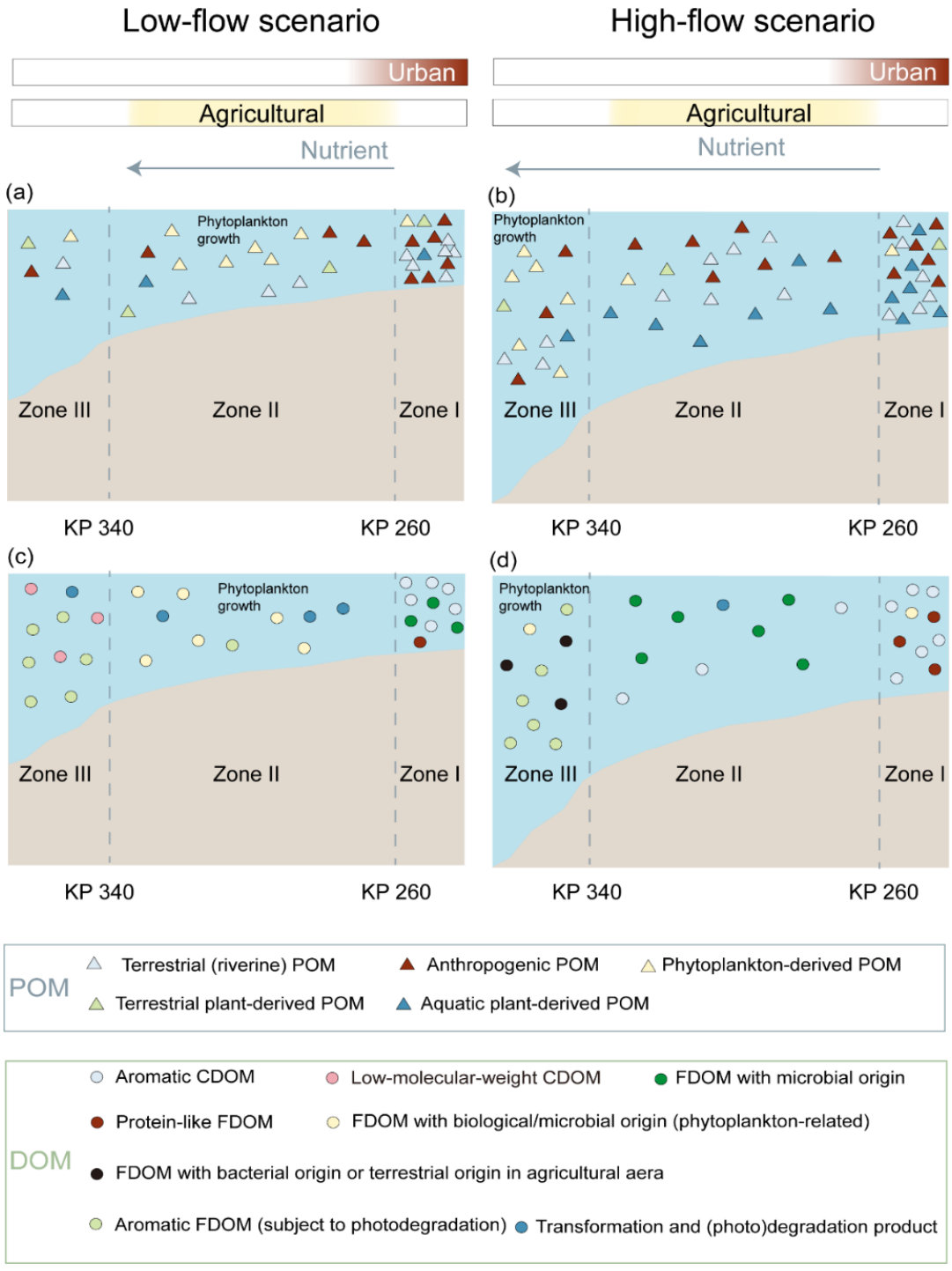



**Figure 6-1.** Schematic diagrams showing the dynamics of (a-b) POM and (c-d) DOM in the Seine Estuary in (a, c) low-flow and (b, d) high-flow scenarios.

## 6.5. Perspectives

This PhD thesis aimed at understanding the biogeochemical functioning of a human-impacted estuary, specifically focusing on how it regulates different types of DOM and POM.

First, it might be necessary to assess the general applicability of the novel proxy (RIX) proposed in this thesis to paleorecords and compare it with other available terrestrial proxies across the critical geological period, such as the Paleocene-Eocene thermal maximum.

It could also be interesting to explore molecular composition of DOM using Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS). This approach presents an opportunity to evaluate relationships between molecular composition of DOM and (microbial and photochemical) transformation processes. It is thus possible to explore which molecular formulas dominate each estuarine zone and how they relate to POM, as well as CDOM and FDOM.

A conceptual model of POM and DOM dynamics in the Seine estuary is proposed under varying flow conditions (Figure 6-1). Additionally, it could be interesting to use unsupervised and supervised machine learning to cross-interpret estuarine DOM and POM data. In this thesis, machine learning and explainable artificial intelligence show potential for assessing DOM dynamics and capturing main DOM characteristics within specific region. The approach proposed in Chapter 5 could also be applied to understand POM dynamics and to find dominating POM characteristics in each zone. By applying machine learning to larger datasets containing DOM and POM, we could uncover new insights into DOM and POM dynamics as well as related biogeochemical processes.

One of the main results of this PhD thesis is to propose a conceptual model to assess the functioning of estuarine ecosystems in terms of DOM and POM dynamics across different types of land use under high and low flow scenarios. However, a critical question remains: Are these

scenarios applicable to other estuaries? Although this thesis provides insights into the biogeochemical functioning of the Seine Estuary, it would be essential for exploring whether the similar patterns/ scenarios occur in different estuaries.

# References


Aiken, E., Bellue, S., Karlan, D., Udry, C., Blumenstock, J.E., 2022. Machine learning and phone data can improve targeting of humanitarian aid. Nature 603, 864–870. https://doi.org/10.1038/s41586-022-04484-9

Al Aukidy, M., Verlicchi, P., 2017. Contributions of combined sewer overflows and treated effluents to the bacterial load released into a coastal area. Science of The Total Environment 607–608, 483–496. https://doi.org/10.1016/j.scitotenv.2017.07.050

Alova, G., Trotter, P.A., Money, A., 2021. A machine-learning approach to predicting Africa's electricity mix based on planned power plants and their chances of success. Nat Energy 6, 158–166. https://doi.org/10.1038/s41560-020-00755-9

Amaral, V., Ortega, T., Romera-Castillo, C., Forja, J., 2021. Linkages between greenhouse gases (CO2, CH4, and N2O) and dissolved organic matter composition in a shallow estuary. Science of The Total Environment 788, 147863. https://doi.org/10.1016/j.scitotenv.2021.147863

Amaral, V., Romera-Castillo, C., García-Delgado, M., Gómez-Parra, A., Forja, J., 2020. Distribution of dissolved organic matter in estuaries of the southern Iberian Atlantic Basin: Sources, behavior and export to the coastal zone. Marine Chemistry 226, 103857. https://doi.org/10.1016/j.marchem.2020.103857

Andrew, A.A., Del Vecchio, R., Subramaniam, A., Blough, N.V., 2013. Chromophoric dissolved organic matter (CDOM) in the Equatorial Atlantic Ocean: Optical properties and their relation to CDOM structure and source. Marine Chemistry 148, 33–43. https://doi.org/10.1016/j.marchem.2012.11.001

Andrisoa, A., Stieglitz, T.C., Rodellas, V., Raimbault, P., 2019. Primary production in coastal lagoons supported by groundwater discharge and porewater fluxes inferred from nitrogen and carbon isotope signatures. Marine Chemistry 210, 48–60. https://doi.org/10.1016/j.marchem.2019.03.003

Antony, R., Willoughby, A.S., Grannas, A.M., Catanzano, V., Sleighter, R.L., Thamban, M., Hatcher, P.G., Nair, S., 2017. Molecular Insights on Dissolved Organic Matter

Transformation by Supraglacial Microbial Communities. Environ. Sci. Technol. 51, 4328–4337. https://doi.org/10.1021/acs.est.6b05780

Asmala, E., Autio, R., Kaartokallio, H., Pitkänen, L., Stedmon, C.A., Thomas, D.N., 2013. Bioavailability of riverine dissolved organic matter in three Baltic Sea estuaries and the effect of catchment land use. Biogeosciences 10, 6969–6986. https://doi.org/10.5194/bg-10-6969-2013

Asmala, E., Haraguchi, L., Markager, S., Massicotte, P., Riemann, B., Staehr, P.A., Carstensen, J., 2018. Eutrophication Leads to Accumulation of Recalcitrant Autochthonous Organic Matter in Coastal Environment. Global Biogeochemical Cycles 32, 1673–1687. https://doi.org/10.1029/2017GB005848

Aufdenkampe, A.K., Hedges, J.I., Richey, J.E., Krusche, A.V., Llerena, C.A., 2001. Sorptive fractionation of dissolved organic nitrogen and amino acids onto fine sediments within the Amazon Basin. Limnology and Oceanography 46, 1921–1935. https://doi.org/10.4319/lo.2001.46.8.1921

Avoine, J., Allen, G.P., Nichols, M., Salomon, J.C., Larsonneur, C., 1981. Suspended-sediment transport in the Seine estuary, France: Effect of man-made modifications on estuary—shelf sedimentology. Marine Geology, Estuary \3- Shelf Interrelationships 40, 119–137. https://doi.org/10.1016/0025-3227(81)90046-3

Bachi, G., Morelli, E., Gonnelli, M., Balestra, C., Casotti, R., Evangelista, V., Repeta, D.J., Santinelli, C., 2023. Fluorescent properties of marine phytoplankton exudates and lability to marine heterotrophic prokaryotes degradation. Limnology and Oceanography 68, 982–1000. https://doi.org/10.1002/lno.12325

Bagheri, M., Al-jabery, K., Wunsch, D., Burken, J.G., 2020. Examining plant uptake and translocation of emerging contaminants using machine learning: Implications to food security. Science of The Total Environment 698, 133999. https://doi.org/10.1016/j.scitotenv.2019.133999

Bagheri, M., Al-jabery, K., Wunsch, D.C., Burken, J.G., 2019. A deeper look at plant uptake of environmental contaminants using intelligent approaches. Science of The Total Environment 651, 561–569. https://doi.org/10.1016/j.scitotenv.2018.09.048

Bauer, M., Blodau, C., 2006. Mobilization of arsenic by dissolved organic matter from iron oxides, soils and sediments. Science of The Total Environment 354, 179–190. https://doi.org/10.1016/j.scitotenv.2005.01.027

Baxter, A.J., Hopmans, E.C., Russell, J.M., Sinninghe Damsté, J.S., 2019. Bacterial GMGTs in East African lake sediments: Their potential as palaeotemperature indicators. Geochimica et Cosmochimica Acta 259, 155–169. https://doi.org/10.1016/j.gca.2019.05.039

Baxter, A.J., Peterse, F., Verschuren, D., Sinninghe Damsté, J.S., 2021. Anoxic in situ production of bacterial GMGTs in the water column and surficial bottom sediments of a meromictic tropical crater lake: Implications for lake paleothermometry. Geochimica et Cosmochimica Acta 306, 171–188. https://doi.org/10.1016/j.gca.2021.05.015

Begum, M.S., Park, J.-H., Yang, L., Shin, K.H., Hur, J., 2023. Optical and molecular indices of dissolved organic matter for estimating biodegradability and resulting carbon dioxide production in inland waters: A review. Water Research 228, 119362. https://doi.org/10.1016/j.watres.2022.119362

Bergamaschi, B.A., Krabbenhoft, D.P., Aiken, G.R., Patino, E., Rumbold, D.G., Orem, W.H., 2012. Tidally Driven Export of Dissolved Organic Carbon, Total Mercury, and Methylmercury from a Mangrove-Dominated Estuary. Environ. Sci. Technol. 46, 1371–1378. https://doi.org/10.1021/es2029137

Bertassoli, D.J., Häggi, C., Chiessi, C.M., Schefuß, E., Hefter, J., Akabane, T.K., Sawakuchi, A.O., 2022. Controls on the distributions of GDGTs and n-alkane isotopic compositions in sediments of the Amazon River Basin. Chemical Geology 594, 120777. https://doi.org/10.1016/j.chemgeo.2022.120777

Bhattacharya, R., Osburn, C.L., 2020. Spatial patterns in dissolved organic matter composition controlled by watershed characteristics in a coastal river network: The Neuse River Basin, USA. Water Research 169, 115248. https://doi.org/10.1016/j.watres.2019.115248

Bianchi, T.S., 2011. The role of terrestrially derived organic carbon in the coastal ocean: A changing paradigm and the priming effect. Proceedings of the National Academy of Sciences 108, 19473–19481. https://doi.org/10.1073/pnas.1017982108

Bianchi, T.S., 2007. Biogeochemistry of Estuaries. Oxford University Press, USA.

Bianchi, T.S., Canuel, E.A., 2011. Chemical Biomarkers in Aquatic Ecosystems. Princeton University Press.

Bianchi, T.S., Thornton, D.C.O., Yvon-Lewis, S.A., King, G.M., Eglinton, T.I., Shields, M.R., Ward, N.D., Curtis, J., 2015. Positive priming of terrestrially derived dissolved organic matter in a freshwater microcosm system. Geophysical Research Letters 42, 5460–5467. https://doi.org/10.1002/2015GL064765

Bibi, R., Kang, H.Y., Kim, D., Jang, J., Kundu, G.K., Kim, Y.K., Kang, C.-K., 2020. Dominance of Autochthonous Phytoplankton-Derived Particulate Organic Matter in a Low-Turbidity Temperate Estuarine Embayment, Gwangyang Bay, Korea. Frontiers in Marine Science 7.

Bieroza, M., Baker, A., Bridgeman, J., 2012. Exploratory analysis of excitation–emission matrix fluorescence spectra with self-organizing maps—A tutorial. Education for Chemical Engineers 7, e22–e31. https://doi.org/10.1016/j.ece.2011.10.002

Bijl, P.K., Frieling, J., Cramwinckel, M.J., Boschman, C., Sluijs, A., Peterse, F., 2021. Maastrichtian–Rupelian paleoclimates in the southwest Pacific – a critical re-evaluation of biomarker paleothermometry and dinoflagellate cyst paleoecology at Ocean Drilling Program Site 1172. Climate of the Past 17, 2393–2425. https://doi.org/10.5194/cp-17-2393-2021

Billen, G., Garnier, J., Le Noë, J., Viennot, P., Gallois, N., Puech, T., Schott, C., Anglade, J., Mary, B., Beaudoin, N., Léonard, J., Mignolet, C., Théry, S., Thieu, V., Silvestre, M., Passy, P., 2021. The Seine Watershed Water-Agro-Food System: Long-Term Trajectories of C, N and P Metabolism, in: Flipo, N., Labadie, P., Lestel, L. (Eds.), The Seine River Basin, The Handbook of Environmental Chemistry. Springer International Publishing, Cham, pp. 91–115. https://doi.org/10.1007/698_2019_393

Bittar, T.B., Berger, S.A., Birsa, L.M., Walters, T.L., Thompson, M.E., Spencer, R.G.M., Mann, E.L., Stubbins, A., Frischer, M.E., Brandes, J.A., 2016. Seasonal dynamics of dissolved, particulate and microbial components of a tidal saltmarsh-dominated estuary under contrasting levels of freshwater discharge. Estuarine, Coastal and Shelf Science 182, 72–85. https://doi.org/10.1016/j.ecss.2016.08.046

Blewett, J., Elling, F.J., Naafs, B.D.A., Kattein, L., Evans, T.W., Lauretano, V., Gallego-Sala, A.V., Pancost, R.D., Pearson, A., 2022. Metabolic and ecological controls on the stable carbon isotopic composition of archaeal (isoGDGT and BDGT) and bacterial (brGDGT) lipids in wetlands and lignites. Geochimica et Cosmochimica Acta 320, 1–25. https://doi.org/10.1016/j.gca.2021.12.023

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 872, 162104. https://doi.org/10.1016/j.scitotenv.2023.162104

Boyle, E.S., Guerriero, N., Thiallet, A., Vecchio, R.D., Blough, N.V., 2009. Optical Properties of Humic Substances and CDOM: Relation to Structure. Environ. Sci. Technol. 43, 2262–2268. https://doi.org/10.1021/es803264g

Brankovits, D., Pohlman, J.W., Niemann, H., Leigh, M.B., Leewis, M.C., Becker, K.W., Iliffe, T.M., Alvarez, F., Lehmann, M.F., Phillips, B., 2017. Methane- and dissolved organic carbon-fueled microbial loop supports a tropical subterranean estuary ecosystem. Nat Commun 8, 1835. https://doi.org/10.1038/s41467-017-01776-x

Bray, E.E., Evans, E.D., 1961. Distribution of n-paraffins as a clue to recognition of source beds. Geochimica et Cosmochimica Acta 22, 2–15. https://doi.org/10.1016/0016-7037(61)90069-2

Buchan, A., LeCleir, G.R., Gulvik, C.A., González, J.M., 2014. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol 12, 686–698. https://doi.org/10.1038/nrmicro3326

Budge, S.M., Parrish, C.C., Mckenzie, C.H., 2001. Fatty acid composition of phytoplankton, settling particulate matter and sediments at a sheltered bivalve aquaculture site. Marine Chemistry 76, 285–303. https://doi.org/10.1016/S0304-4203(01)00068-8

Butturini, A., Guarch, A., Romaní, A.M., Freixa, A., Amalfitano, S., Fazi, S., Ejarque, E., 2016. Hydrological conditions control in situ DOM retention and release along a Mediterranean river. Water Research 99, 33–45. https://doi.org/10.1016/j.watres.2016.04.036

Cai, W.-J., 2011. Estuarine and Coastal Ocean Carbon Paradox: $CO_2$ Sinks or Sites of Terrestrial Carbon Incineration? Annu. Rev. Mar. Sci. 3, 123–145. https://doi.org/10.1146/annurev-marine-120709-142723

Canuel, E.A., Cammer, S.S., McIntosh, H.A., Pondell, C.R., 2012. Climate Change Impacts on the Organic Carbon Cycle at the Land-Ocean Interface. Annual Review of Earth and Planetary Sciences 40, 685–711. https://doi.org/10.1146/annurev-earth-042711-105511

Canuel, E.A., Hardison, A.K., 2016. Sources, Ages, and Alteration of Organic Matter in Estuaries. Annu. Rev. Mar. Sci. 8, 409–434. https://doi.org/10.1146/annurev-marine-122414-034058

Cao, J., Lian, E., Yang, S., Ge, H., Jin, X., He, J., Jia, G., 2022. The distribution of intact polar lipid-derived branched tetraethers along a freshwater-seawater pH gradient in coastal East China Sea. Chemical Geology 596, 120808. https://doi.org/10.1016/j.chemgeo.2022.120808

Carreira, R.S., Wagener, A.L.R., Readman, J.W., 2004. Sterols as markers of sewage contamination in a tropical urban estuary (Guanabara Bay, Brazil): space–time variations. Estuarine, Coastal and Shelf Science 60, 587–598. https://doi.org/10.1016/j.ecss.2004.02.014

Carvajalino-Fernández, M.A., Sævik, P.N., Johnsen, I.A., Albretsen, J., Keeley, N.B., 2020. Simulating particle organic matter dispersal beneath Atlantic salmon fish farms using different resuspension approaches. Marine Pollution Bulletin 161, 111685. https://doi.org/10.1016/j.marpolbul.2020.111685

Casciotti, K.L., 2016. Nitrogen and Oxygen Isotopic Studies of the Marine Nitrogen Cycle. Annual Review of Marine Science 8, 379–407. https://doi.org/10.1146/annurev-marine-010213-135052

Castillo, C.R., Sarmento, H., Alvarez-Salgado, X.A., Gasol, J.M., Marraséa, C., 2010. Production of chromophoric dissolved organic matter by marine phytoplankton. Limnology and Oceanography 55, 446–454.

Catalán, N., Pastor, A., Borrego, C.M., Casas-Ruiz, J.P., Hawkes, J.A., Gutiérrez, C., von Schiller, D., Marcé, R., 2021. The relevance of environment vs. composition on dissolved organic matter degradation in freshwaters. Limnology and Oceanography 66, 306–320. https://doi.org/10.1002/lno.11606

Celis-Hernandez, O., Cundy, A.B., Croudace, I.W., Ward, R.D., Busquets, R., Wilkinson, J.L., 2021. Assessing the role of the “estuarine filter” for emerging contaminants: pharmaceuticals, perfluoroalkyl compounds and plasticisers in sediment cores from two contrasting systems in the southern U.K. Water Research 189, 116610. https://doi.org/10.1016/j.watres.2020.116610

Chai, L., Huang, M., Fan, H., Wang, J., Jiang, D., Zhang, M., Huang, Y., 2019. Urbanization altered regional soil organic matter quantity and quality: Insight from excitation emission matrix (EEM) and parallel factor analysis (PARAFAC). Chemosphere 220, 249–258.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 321–357. https://doi.org/10.1613/jair.953

Chen, Q., Chen, F., Gonsior, M., Li, Y., Wang, Yu, He, C., Cai, R., Xu, J., Wang, Yimeng, Xu, D., Sun, J., Zhang, T., Shi, Q., Jiao, N., Zheng, Q., 2021. Correspondence between DOM molecules and microbial community in a subtropical coastal estuary on a spatiotemporal scale. Environment International 154, 106558. https://doi.org/10.1016/j.envint.2021.106558

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. Science of The Total Environment 796, 148968. https://doi.org/10.1016/j.scitotenv.2021.148968

Chen, Y., Zheng, F., Yang, H., Yang, W., Wu, R., Liu, X., Liang, H., Chen, H., Pei, H., Zhang, C., Pancost, R.D., Zeng, Z., 2022. The production of diverse brGDGTs by an Acidobacterium providing a physiological basis for paleoclimate proxies. Geochimica et Cosmochimica Acta. https://doi.org/10.1016/j.gca.2022.08.033

Cheng, Z., Yu, F., Ruan, X., Cheng, P., Chen, N., Tao, S., Zong, Y., Yang, H., Huang, Z., 2021. GDGTs as indicators for organic-matter sources in a small subtropical river-estuary system. Organic Geochemistry 153, 104180. https://doi.org/10.1016/j.orggeochem.2021.104180

Chin, Y.-Ping., Aiken, George., O'Loughlin, Edward., 1994. Molecular Weight, Polydispersity, and Spectroscopic Properties of Aquatic Humic Substances. Environ. Sci. Technol. 28, 1853–1858. https://doi.org/10.1021/es00060a015

Chu, M., Sachs, J.P., Zhang, H., Ding, Y., Jin, G., Zhao, M., 2020. Spatiotemporal variations of organic matter sources in two mangrove-fringed estuaries in Hainan, China. Organic Geochemistry 147, 104066. https://doi.org/10.1016/j.orggeochem.2020.104066

Chupakova, A.A., Chupakov, A.V., Neverova, N.V., Shirokova, L.S., Pokrovsky, O.S., 2018. Photodegradation of river dissolved organic matter and trace metals in the largest European Arctic estuary. Science of The Total Environment 622–623, 1343–1352. https://doi.org/10.1016/j.scitotenv.2017.12.030

Claustre, H., Marty, J., Cassiani, L., Dagaut, J., 1989. Fatty acid dynamics in phytoplankton and microzooplankton communities during a spring bloom in the coastal Ligurian Sea: Ecological implications. Mar. Microb. Food Webs. 3, 51–66.

Coble, P.G., 2007. Marine Optical Biogeochemistry: The Chemistry of Ocean Color. Chem. Rev. 107, 402–418. https://doi.org/10.1021/cr050350+

Coble, P.G., Green, S.A., Blough, N.V., Gagosian, R.B., 1990. Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. Nature 348, 432–435. https://doi.org/10.1038/348432a0

Coffinet, S., Huguet, A., Bergonzini, L., Pedentchouk, N., Williamson, D., Anquetil, C., Gałka, M., Kołaczek, P., Karpińska-Kołaczek, M., Majule, A., Laggoun-Défarge, F., Wagner, T., Derenne, S., 2018. Impact of climate change on the ecology of the Kyambangunguru crater marsh in southwestern Tanzania during the Late Holocene. Quaternary Science Reviews 196, 100–117. https://doi.org/10.1016/j.quascirev.2018.07.038

Cordeiro, L.G.S.M., Carreira, R.S., Wagener, A.L.R., 2008. Geochemistry of fecal sterols in a contaminated estuary in southeastern Brazil. Organic Geochemistry, Advances in Organic Geochemistry 2007 39, 1097–1103. https://doi.org/10.1016/j.orggeochem.2008.02.022

Cornu, J.Y., Gutierrez, M., Randriamamonjy, S., Gaudin, P., Ouedraogo, F., Sourzac, M., Parlanti, E., Lebeau, T., Janot, N., 2022. Contrasting effects of siderophores pyoverdine and desferrioxamine B on the mobility of iron, aluminum, and copper in Cu-contaminated soils. Geoderma 420, 115897. https://doi.org/10.1016/j.geoderma.2022.115897

Cory, R.M., Boyer, E.W., McKnight, D.M., 2011. Spectral methods to advance understanding of dissolved organic carbon dynamics in forested catchments, in: Forest Hydrology and Biogeochemistry: Synthesis of Past Research and Future Directions. Springer, pp. 117–135.

Couturier, M., Nozais, C., Chaillou, G., 2016. Microtidal subterranean estuaries as a source of fresh terrestrial dissolved organic matter to the coastal ocean. Marine Chemistry 186, 46–57. https://doi.org/10.1016/j.marchem.2016.08.001

Crampton-Flood, E.D., van der Weijst, C.M., van der Molen, G., Bouquet, M., Yedema, Y., Donders, T.H., Sangiorgi, F., Sluijs, A., Damsté, J.S.S., Peterse, F., 2021. Identifying marine and freshwater overprints on soil-derived branched GDGT temperature signals in Pliocene Mississippi and Amazon River fan sediments. Organic Geochemistry 154, 104200.

Cramwinckel, M.J., Huber, M., Kocken, I.J., Agnini, C., Bijl, P.K., Bohaty, S.M., Frieling, J., Goldner, A., Hilgen, F.J., Kip, E.L., Peterse, F., van der Ploeg, R., Röhl, U., Schouten, S., Sluijs, A., 2018. Synchronous tropical and polar temperature evolution in the Eocene. Nature 559, 382–386. https://doi.org/10.1038/s41586-018-0272-2

Cranwell, P.A., 1984. Lipid geochemistry of sediments from Upton Broad, a small productive lake. Organic Geochemistry 7, 25–37. https://doi.org/10.1016/0146-6380(84)90134-7

Cranwell, P.A., 1981. Diagenesis of free and bound lipids in terrestrial detritus deposited in a lacustrine sediment. Organic Geochemistry 3, 79–89. https://doi.org/10.1016/0146-6380(81)90002-4

Cranwell, P.A., Eglinton, G., Robinson, N., 1987. Lipids of aquatic organisms as potential contributors to lacustrine sediments—II. Organic Geochemistry 11, 513–527. https://doi.org/10.1016/0146-6380(87)90007-6

Crone, S., Vives-Flórez, M., Kvich, L., Saunders, A.M., Malone, M., Nicolaisen, M.H., Martínez-García, E., Rojas-Acosta, C., Catalina Gomez-Puerto, M., Calum, H., Whiteley, M., Kolter, R., Bjarnsholt, T., 2020. The environmental occurrence of Pseudomonas aeruginosa. APMIS 128, 220–231. https://doi.org/10.1111/apm.13010

Cuss, C.W., Guéguen, C., 2016. Analysis of dissolved organic matter fluorescence using self-organizing maps: mini-review and tutorial. Anal. Methods 8, 716–725. https://doi.org/10.1039/C5AY02549D

Dai, G., Zhu, E., Liu, Z., Wang, Y., Zhu, S., Wang, S., Ma, T., Jia, J., Wang, X., Hou, S., Fu, P., Peterse, F., Feng, X., 2019. Compositional Characteristics of Fluvial Particulate Organic Matter Exported From the World's Largest Alpine Wetland. Journal of Geophysical Research: Biogeosciences 124, 2709–2727. https://doi.org/10.1029/2019JG005231

Dai, J., Sun, M.-Y., 2007. Organic matter sources and their use by bacteria in the sediments of the Altamaha estuary during high and low discharge periods. Organic Geochemistry 38, 1–15. https://doi.org/10.1016/j.orggeochem.2006.10.002

Dai, M., Su, J., Zhao, Y., Hofmann, E.E., Cao, Z., Cai, W.-J., Gan, J., Lacroix, F., Laruelle, G.G., Meng, F., Müller, J.D., Regnier, P.A.G., Wang, G., Wang, Z., 2022. Carbon Fluxes in the Coastal Ocean: Synthesis, Boundary Processes, and Future Trends. Annual Review of Earth and Planetary Sciences 50, 593–626. https://doi.org/10.1146/annurev-earth-032320-090746

Darnaude, A.M., 2005. Fish ecology and terrestrial carbon use in coastal areas: implications for marine fish production. Journal of Animal Ecology 74, 864–876. https://doi.org/10.1111/j.1365-2656.2005.00978.x

Dartnell, L.R., Roberts, T.A., Moore, G., Ward, J.M., Muller, J.-P., 2013. Fluorescence characterization of clinically-important bacteria. PLoS One 8, e75270. https://doi.org/10.1371/journal.pone.0075270

David, V., Tortajada, S., Savoye, N., Breret, M., Lachaussée, N., Philippine, O., Robin, F.-X., Dupuy, C., 2020. Impact of human activities on the spatio-seasonal dynamics of plankton diversity in drained marshes and consequences on eutrophication. Water Research 170, 115287. https://doi.org/10.1016/j.watres.2019.115287

Day Jr., J.W., Yáñez-Arancibia, A., Kemp, W.M., Crump, B.C., 2012. Introduction to Estuarine Ecology, in: Estuarine Ecology. John Wiley & Sons, Ltd, pp. 1–18. https://doi.org/10.1002/9781118412787.ch1

De Jonge, C., Hopmans, E.C., Stadnitskaia, A., Rijpstra, W.I.C., Hofland, R., Tegelaar, E., Sinninghe Damsté, J.S., 2013. Identification of novel penta- and hexamethylated branched glycerol dialkyl glycerol tetraethers in peat using HPLC–MS2, GC–MS and GC–SMB-MS. Organic Geochemistry 54, 78–82. https://doi.org/10.1016/j.orggeochem.2012.10.004

De Jonge, C., Hopmans, E.C., Zell, C.I., Kim, J.-H., Schouten, S., Sinninghe Damsté, J.S., 2014. Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction. Geochimica et Cosmochimica Acta 141, 97–112. https://doi.org/10.1016/j.gca.2014.06.013

De Jonge, C., Stadnitskaia, A., Hopmans, E.C., Cherkashov, G., Fedotov, A., Streletskaya, I.D., Vasiliev, A.A., Sinninghe Damsté, J.S., 2015. Drastic changes in the distribution of branched tetraether lipids in suspended matter and sediments from the Yenisei River and Kara Sea (Siberia): Implications for the use of brGDGT-based proxies in coastal marine sediments. Geochimica et Cosmochimica Acta 165, 200–225. https://doi.org/10.1016/j.gca.2015.05.044

Dearing Crampton-Flood, E., Peterse, F., Sinninghe Damsté, J.S., 2019. Production of branched tetraethers in the marine realm: Svalbard fjord sediments revisited. Organic Geochemistry 138, 103907. https://doi.org/10.1016/j.orggeochem.2019.103907

DeFrancesco, C., Guéguen, C., 2021. Long-term trends in dissolved organic matter composition and its relation to sea ice in the Canada Basin, Arctic Ocean (2007–2017). Journal of Geophysical Research: Oceans 126, e2020JC016578.

Denk, T.R.A., Mohn, J., Decock, C., Lewicka-Szczebak, D., Harris, E., Butterbach-Bahl, K., Kiese, R., Wolf, B., 2017. The nitrogen cycle: A review of isotope effects and isotope modeling approaches. Soil Biology and Biochemistry 105, 121–137. https://doi.org/10.1016/j.soilbio.2016.11.015

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

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., 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

Ding, S., Schwab, V.F., Ueberschaar, N., Roth, V.-N., Lange, M., Xu, Y., Gleixner, G., Pohnert, G., 2016. Identification of novel 7-methyl and cyclopentanyl branched glycerol dialkyl glycerol tetraethers in lake sediments. Organic Geochemistry 102, 52–58. https://doi.org/10.1016/j.orggeochem.2016.09.009

Dittmar, T., Koch, B., Hertkorn, N., Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnology and Oceanography: Methods 6, 230–235.

Dixon, J.L., Osburn, C.L., Paerl, H.W., Peierls, B.L., 2014. Seasonal changes in estuarine dissolved organic matter due to variable flushing time and wind-driven mixing events. Estuarine, Coastal and Shelf Science 151, 210–220. https://doi.org/10.1016/j.ecss.2014.10.013

Du, Y., Zhang, Y., Chen, F., Chang, Y., Liu, Z., 2016. Photochemical reactivities of dissolved organic matter (DOM) in a sub-alpine lake revealed by EEM-PARAFAC: An insight into

the fate of allochthonous DOM in alpine lakes affected by climate change. Science of The Total Environment 568, 216–225. https://doi.org/10.1016/j.scitotenv.2016.06.036

Edzwald, J.K., Tobiason, J.E., 1999. Enhanced coagulation: US requirements and a broader view. Water Science and Technology 40, 63–70. https://doi.org/10.1016/S0273-1223(99)00641-1

Ejarque, E., Freixa, A., Vazquez, E., Guarch, A., Amalfitano, S., Fazi, S., Romaní, A.M., Butturini, A., 2017. Quality and reactivity of dissolved organic matter in a Mediterranean river across hydrological and spatial gradients. Science of The Total Environment 599–600, 1802–1812. https://doi.org/10.1016/j.scitotenv.2017.05.113

Elling, F.J., Kattein, L., David A. Naafs, B., Lauretano, V., Pearson, A., 2023. Heterotrophic origin and diverse sources of branched glycerol monoalkyl glycerol tetraethers (brGMGTs) in peats and lignites. Organic Geochemistry 104558. https://doi.org/10.1016/j.orggeochem.2023.104558

Etcheber, H., Taillez, A., Abril, G., Garnier, J., Servais, P., Moatar, F., Commarieu, M.-V., 2007. Particulate organic carbon in the estuarine turbidity maxima of the Gironde, Loire and Seine estuaries: origin and lability. Hydrobiologia 588, 245–259. https://doi.org/10.1007/s10750-007-0667-9

Fabre, C., Sauvage, S., Probst, J.-L., Sánchez-Pérez, J.M., 2020. Global-scale daily riverine DOC fluxes from lands to the oceans with a generic model. Global and Planetary Change 194, 103294. https://doi.org/10.1016/j.gloplacha.2020.103294

Fairbridge, R.W., 1980. The estuary: its definition and geodynamic cycle. Chemistry and biochemistry of estuaries 1–35.

Fedik, N., Zubatyuk, R., Kulichenko, M., Lubbers, N., Smith, J.S., Nebgen, B., Messerly, R., Li, Y.W., Boldyrev, A.I., Barros, K., Isayev, O., Tretiak, S., 2022. Extending machine learning beyond interatomic potentials for predicting molecular properties. Nat Rev Chem 6, 653–672. https://doi.org/10.1038/s41570-022-00416-3

Fellman, J.B., Hood, E., Edwards, R.T., D'Amore, D.V., 2009. Changes in the concentration, biodegradability, and fluorescent properties of dissolved organic matter during stormflows in coastal temperate watersheds. Journal of Geophysical Research: Biogeosciences 114.

Ficken, K.J., Li, B., Swain, D.L., Eglinton, G., 2000. An n-alkane proxy for the sedimentary input of submerged/floating freshwater aquatic macrophytes. Organic Geochemistry 31, 745–749. https://doi.org/10.1016/S0146-6380(00)00081-4

Fleming, S.W., Watson, J.R., Ellenson, A., Cannon, A.J., Vesselinov, V.C., 2021. Machine learning in Earth and environmental science requires education and research policy reforms. Nat. Geosci. 14, 878–880. https://doi.org/10.1038/s41561-021-00865-3

Flipo, N., Gallois, N., Labarthe, B., Baratelli, F., Viennot, P., Schuite, J., Rivière, A., Bonnet, R., Boé, J., 2020. Pluri-annual water budget on the Seine basin: past, current and future trends. The seine river basin 90, 59–89.

Flipo, N., Lestel, L., Labadie, P., Meybeck, M., Garnier, J., 2021. Trajectories of the Seine River Basin, in: Flipo, N., Labadie, P., Lestel, L. (Eds.), The Seine River Basin, The Handbook of Environmental Chemistry. Springer International Publishing, Cham, pp. 1–28. https://doi.org/10.1007/698_2019_437

Fox, B.G., Thorn, R.M.S., Anesio, A.M., Reynolds, D.M., 2017. The in situ bacterial production of fluorescent organic matter; an investigation at a species level. Water Research 125, 350–359. https://doi.org/10.1016/j.watres.2017.08.040

Freymond, C.V., Peterse, F., Fischer, L.V., Filip, F., Giosan, L., Eglinton, T.I., 2017. Branched GDGT signals in fluvial sediments of the Danube River basin: Method comparison and longitudinal evolution. Organic Geochemistry 103, 88–96. https://doi.org/10.1016/j.orggeochem.2016.11.002

Fuhrman, J.A., 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399, 541–548. https://doi.org/10.1038/21119

García-Martín, E.E., Sanders, R., Evans, C.D., Kitidis, V., Lapworth, D.J., Rees, A.P., Spears, B.M., Tye, A., Williamson, J.L., Balfour, C., Best, M., Bowes, M., Breimann, S., Brown, I.J., Burden, A., Callaghan, N., Felgate, S.L., Fishwick, J., Fraser, M., Gibb, S.W., Gilbert, P.J., Godsell, N., Gomez-Castillo, A.P., Hargreaves, G., Jones, O., Kennedy, P., Lichtschlag, A., Martin, A., May, R., Mawji, E., Mounteney, I., Nightingale, P.D., Olszewska, J.P., Painter, S.C., Pearce, C.R., Pereira, M.G., Peel, K., Pickard, A., Stephens, J.A., Stinchcombe, M., Williams, P., Woodward, E.M.S., Yarrow, D., Mayor, D.J., 2021. Contrasting Estuarine Processing of Dissolved Organic Matter Derived From Natural and Human-Impacted

Landscapes. Global Biogeochemical Cycles 35, e2021GB007023. https://doi.org/10.1029/2021GB007023

Garnier, J., Marescaux, A., Guillon, S., Vilmin, L., Rocher, V., Billen, G., Thieu, V., Silvestre, M., Passy, P., Raimonet, M., Groleau, A., Théry, S., Tallec, G., Flipo, N., 2021. Ecological Functioning of the Seine River: From Long-Term Modelling Approaches to High-Frequency Data Analysis, in: Flipo, Nicolas, Labadie, P., Lestel, L. (Eds.), The Seine River Basin, The Handbook of Environmental Chemistry. Springer International Publishing, Cham, pp. 189–216. https://doi.org/10.1007/698_2019_379

Gattuso, J.-P., Frankignoulle, M., Wollast, R., 1998. Carbon and Carbonate Metabolism in Coastal Aquatic Ecosystems. Annual Review of Ecology and Systematics 29, 405–434. https://doi.org/10.1146/annurev.ecolsys.29.1.405

Gladu, P.K., Patterson, G.W., Wikfors, G.H., Chitwood, D.J., Lusby, W.R., 1990. The occurrence of brassicasterol and epibrassicasterol in the chromophycota. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 97, 491–494. https://doi.org/10.1016/0305-0491(90)90149-N

Gliozzi, A., Paoli, G., De Rosa, M., Gambacorta, A., 1983. Effect of isoprenoid cyclization on the transition temperature of lipids in thermophilic archaebacteria. Biochimica et Biophysica Acta (BBA) - Biomembranes 735, 234–242. https://doi.org/10.1016/0005-2736(83)90298-5

Goñi, M.A., Juranek, L.W., Sipler, R.E., Welch, K.A., 2021. Particulate Organic Matter Distributions in the Water Column of the Chukchi Sea During Late Summer. Journal of Geophysical Research: Oceans 126, e2021JC017664. https://doi.org/10.1029/2021JC017664

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

Grasso, F., Verney, R., Le Hir, P., Thouvenin, B., Schulz, E., Kervella, Y., Khojasteh Pour Fard, I., Lemoine, J.-P., Dumas, F., Garnier, V., 2018. Suspended Sediment Dynamics in the Macrotidal Seine Estuary (France): 1. Numerical Modeling of Turbidity Maximum

Dynamics. Journal of Geophysical Research: Oceans 123, 558–577. https://doi.org/10.1002/2017JC013185

Grimalt, J.O., Fernandez, P., Bayona, J.M., Albaiges, J., 1990. Assessment of fecal sterols and ketones as indicators of urban sewage inputs to coastal waters. Environ. Sci. Technol. 24, 357–363. https://doi.org/10.1021/es00073a011

Grunert, B.K., Tzortziou, M., Neale, P., Menendez, A., Hernes, P., 2021. DOM degradation by light and microbes along the Yukon River-coastal ocean continuum. Sci Rep 11, 10236. https://doi.org/10.1038/s41598-021-89327-9

Guéguen, C., Cuss, C.W., Cassels, C.J., Carmack, E.C., 2014. Absorption and fluorescence of dissolved organic matter in the waters of the Canadian Arctic Archipelago, Baffin Bay, and the Labrador Sea. Journal of Geophysical Research: Oceans 119, 2034–2047. https://doi.org/10.1002/2013JC009173

Guenet, B., Danger, M., Abbadie, L., Lacroix, G., 2010. Priming effect: bridging the gap between terrestrial and aquatic ecology. Ecology 91, 2850–2861. https://doi.org/10.1890/09-1968.1

Guézennec, L., Lafite, R., Dupont, J.-P., Meyer, R., Boust, D., 1999. Hydrodynamics of suspended particulate matter in the tidal freshwater zone of a macrotidal estuary (the Seine Estuary, France). Estuaries 22, 717–727. https://doi.org/10.2307/1353058

Guillocheau, F., Robin, C., Allemand, P., Bourquin, S., Brault, N., Dromart, G., Friedenberg, R., Garcia, J.-P., Gaulier, J.-M., Gaumet, F., Grosdoy, B., Hanot, F., Le Strat, P., Mettraux, M., Nalpas, T., Prijac, C., Rigollet, C., Serrano, O., Grandjean, G., 2000. Meso-Cenozoic geodynamic evolution of the Paris Basin: 3D stratigraphic constraints. Geodinamica Acta 13, 189–245. https://doi.org/10.1016/S0985-3111(00)00118-2

Guo, W., Jia, G., Ye, F., Xiao, H., Zhang, Z., 2019. Lipid biomarkers in suspended particulate matter and surface sediments in the Pearl River Estuary, a subtropical estuary in southern China. Science of The Total Environment 646, 416–426. https://doi.org/10.1016/j.scitotenv.2018.07.159

Guo, W., Yang, L., Zhai, W., Chen, W., Osburn, C.L., Huang, X., Li, Y., 2014. Runoff-mediated seasonal oscillation in the dynamics of dissolved organic matter in different branches of a large bifurcated estuary—The Changjiang Estuary. Journal of Geophysical Research: Biogeosciences 119, 776–793. https://doi.org/10.1002/2013JG002540

Halamka, T.A., Raberg, J.H., McFarlin, J.M., Younkin, A.D., Mulligan, C., Liu, X.-L., Kopf, S.H., 2022. Production of diverse brGDGTs by Acidobacterium Solibacter usitatus in response to temperature, pH, and O2 provides a culturing perspective on brGDGT proxies and biosynthesis. Geobiology n/a. https://doi.org/10.1111/gbi.12525

Halvorson, H.M., Francoeur, S.N., Findlay, R.H., Kuehn, K.A., 2019. Algal-Mediated Priming Effects on the Ecological Stoichiometry of Leaf Litter Decomposition: A Meta-Analysis. Frontiers in Earth Science 7.

Hambly, A.C., Arvin, E., Pedersen, L.-F., Pedersen, P.B., Seredyńska-Sobecka, B., Stedmon, C.A., 2015. Characterising organic matter in recirculating aquaculture systems with fluorescence EEM spectroscopy. Water Research 83, 112–120. https://doi.org/10.1016/j.watres.2015.06.037

Hamza, W., 2021. Dust Storms and Its Benefits to the Marine Life of the Arabian Gulf, in: Jawad, L.A. (Ed.), The Arabian Seas: Biodiversity, Environmental Challenges and Conservation Measures. Springer International Publishing, Cham, pp. 141–160. https://doi.org/10.1007/978-3-030-51506-5_7

Hansell, D., Carlson, C., Repeta, D., Schlitzer, R., 2009. Dissolved Organic Matter in the Ocean: A Controversy Stimulates New Insights. Oceanog. 22, 202–211. https://doi.org/10.5670/oceanog.2009.109

Hansell, D.A., Carlson, C.A., Schlitzer, R., 2012. Net removal of major marine dissolved organic carbon fractions in the subsurface ocean. Global Biogeochemical Cycles 26. https://doi.org/10.1029/2011GB004069

Hansell, D.A., Orellana, M.V., 2021. Dissolved Organic Matter in the Global Ocean: A Primer. Gels 7, 128. https://doi.org/10.3390/gels7030128

Hao, Y., Ma, H., Wang, Q., Ge, L., Yang, Y., Zhu, C., 2021. Refractory DOM in industrial wastewater: Formation and selective oxidation of AOPs. Chemical Engineering Journal 406, 126857. https://doi.org/10.1016/j.cej.2020.126857

Harjung, A., Schweichhart, J., Rasch, G., Griebler, C., 2023. Large-scale study on groundwater dissolved organic matter reveals a strong heterogeneity and a complex microbial footprint. Science of The Total Environment 854, 158542. https://doi.org/10.1016/j.scitotenv.2022.158542

Harning, D.J., Curtin, L., Geirsdóttir, Á., D'Andrea, W.J., Miller, G.H., Sepúlveda, J., 2020. Lipid Biomarkers Quantify Holocene Summer Temperature and Ice Cap Sensitivity in Icelandic Lakes. Geophysical Research Letters 47, e2019GL085728. https://doi.org/10.1029/2019GL085728

Hart, G.L.W., Mueller, T., Toher, C., Curtarolo, S., 2021. Machine learning for alloys. Nat Rev Mater 6, 730–755. https://doi.org/10.1038/s41578-021-00340-w

Hartigan, J.A., Wong, M.A., 1979. A K-Means Clustering Algorithm. Journal of the Royal Statistical Society: Series C (Applied Statistics) 28, 100–108. https://doi.org/10.2307/2346830

Haywood, B.J., White, J.R., Cook, R.L., 2018. Investigation of an early season river flood pulse: Carbon cycling in a subtropical estuary. Science of The Total Environment 635, 867–877. https://doi.org/10.1016/j.scitotenv.2018.03.379

He, D., Mead, R.N., Belicka, L., Pisani, O., Jaffé, R., 2014. Assessing source contributions to particulate organic matter in a subtropical estuary: A biomarker approach. Organic Geochemistry 75, 129–139. https://doi.org/10.1016/j.orggeochem.2014.06.012

He, D., Zhang, K., Tang, J., Cui, X., Sun, Y., 2018. Using fecal sterols to assess dynamics of sewage input in sediments along a human-impacted river-estuary system in eastern China. Science of The Total Environment 636, 787–797. https://doi.org/10.1016/j.scitotenv.2018.04.314

He, W., Chen, M., Schlautman, M.A., Hur, J., 2016. Dynamic exchanges between DOM and POM pools in coastal and inland aquatic ecosystems: A review. Science of The Total Environment 551–552, 415–428. https://doi.org/10.1016/j.scitotenv.2016.02.031

He, X.-S., Fan, Q.-D., 2016. Investigating the effect of landfill leachates on the characteristics of dissolved organic matter in groundwater using excitation–emission matrix fluorescence spectra coupled with fluorescence regional integration and self-organizing map. Environ Sci Pollut Res 23, 21229–21237. https://doi.org/10.1007/s11356-016-7308-7

Hedges, J.I., Keil, R.G., Benner, R., 1997. What happens to terrestrial organic matter in the ocean? Organic Geochemistry 27, 195–212. https://doi.org/10.1016/S0146-6380(97)00066-1

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

Hopmans, E.C., Schouten, S., Sinninghe Damsté, J.S., 2016. The effect of improved chromatography on GDGT-based palaeoproxies. Organic Geochemistry 93, 1–6. https://doi.org/10.1016/j.orggeochem.2015.12.006

Hopmans, E.C., Weijers, J.W.H., Schefuß, E., Herfort, L., Sinninghe Damsté, J.S., Schouten, S., 2004. A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. Earth and Planetary Science Letters 224, 107–116. https://doi.org/10.1016/j.epsl.2004.05.012

Hounshell, A.G., Fegley, S.R., Hall, N.S., Osburn, C.L., Paerl, H.W., 2022. Riverine Discharge and Phytoplankton Biomass Control Dissolved and Particulate Organic Matter Dynamics over Spatial and Temporal Scales in the Neuse River Estuary, North Carolina. Estuaries and Coasts 45, 96–113. https://doi.org/10.1007/s12237-021-00955-w

Hounshell, A.G., Peierls, B.L., Osburn, C.L., Paerl, H.W., 2017. Stimulation of Phytoplankton Production by Anthropogenic Dissolved Organic Nitrogen in a Coastal Plain Estuary. Environ. Sci. Technol. 51, 13104–13112. https://doi.org/10.1021/acs.est.7b03538

Housh, M., Ostfeld, A., 2015. An integrated logit model for contamination event detection in water distribution systems. Water Research 75, 210–223. https://doi.org/10.1016/j.watres.2015.02.016

Hu, J., Meyers, P.A., Chen, G., Peng, P., Yang, Q., 2012. Archaeal and bacterial glycerol dialkyl glycerol tetraethers in sediments from the Eastern Lau Spreading Center, South Pacific Ocean. Organic Geochemistry 43, 162–167. https://doi.org/10.1016/j.orggeochem.2011.10.012

Hu, T., Luo, M., Xu, Y., Gong, S., Chen, D., 2021. Production of labile protein-like dissolved organic carbon associated with anaerobic methane oxidization in the haima cold seeps, south china sea. Frontiers in Marine Science 8, 797084.

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. Environ. Sci. Technol. 51, 6936–6944. https://doi.org/10.1021/acs.est.7b01210

Hu, X., Zhou, Y., Zhou, L., Zhang, Y., Wu, L., Xu, H., Zhu, G., Jang, K.-S., Spencer, R.G.M., Jeppesen, E., Brookes, J.D., Wu, F., 2022. Urban and agricultural land use regulates the

molecular composition and bio-lability of fluvial dissolved organic matter in human-impacted southeastern China. carbon res 1, 19. https://doi.org/10.1007/s44246-022-00020-6

Hu, Y., Lu, Y., Edmonds, J.W., Liu, C., Wang, S., Das, O., Liu, J., Zheng, C., 2016. Hydrological and land use control of watershed exports of dissolved organic matter in a large arid river basin in northwestern China. Journal of Geophysical Research: Biogeosciences 121, 466–478. https://doi.org/10.1002/2015JG003082

Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., 2021. Machine learning in natural and engineered water systems. Water Research 205, 117666. https://doi.org/10.1016/j.watres.2021.117666

Huguet, A., Coffinet, S., Roussel, A., Gayraud, F., Anquetil, C., Bergonzini, L., Bonanomi, G., Williamson, D., Majule, A., Derenne, S., 2019. Evaluation of 3-hydroxy fatty acids as a pH and temperature proxy in soils from temperate and tropical altitudinal gradients. Organic Geochemistry 129, 1–13. https://doi.org/10.1016/j.orggeochem.2019.01.002

Huguet, A., Meador, T.B., Laggoun-Défarge, F., Könneke, M., Wu, W., Derenne, S., Hinrichs, K.-U., 2017. Production rates of bacterial tetraether lipids and fatty acids in peatland under varying oxygen concentrations. Geochimica et Cosmochimica Acta 203, 103–116. https://doi.org/10.1016/j.gca.2017.01.012

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

Huguet, A., Vacher, L., Saubusse, S., Etcheber, H., Abril, G., Relexans, S., Ibalot, F., Parlanti, E., 2010. New insights into the size distribution of fluorescent dissolved organic matter in estuarine waters. Organic Geochemistry 41, 595–610. https://doi.org/10.1016/j.orggeochem.2010.02.006

Huguet, C., de Lange, G.J., Gustafsson, Ö., Middelburg, J.J., Sinninghe Damsté, J.S., Schouten, S., 2008. Selective preservation of soil organic matter in oxidized marine sediments (Madeira Abyssal Plain). Geochimica et Cosmochimica Acta 72, 6061–6068. https://doi.org/10.1016/j.gca.2008.09.021

Huguet, C., Hopmans, E.C., Febo-Ayala, W., Thompson, D.H., Sinninghe Damsté, J.S., Schouten, S., 2006. An improved method to determine the absolute abundance of glycerol

dibiphytanyl glycerol tetraether lipids. Organic Geochemistry 37, 1036–1041. https://doi.org/10.1016/j.orggeochem.2006.05.008

Ishii, S.K.L., Boyer, T.H., 2012. Behavior of Reoccurring PARAFAC Components in Fluorescent Dissolved Organic Matter in Natural and Engineered Systems: A Critical Review. Environ. Sci. Technol. 46, 2006–2017. https://doi.org/10.1021/es2043504

Jaffé, R., Cawley, K.M., Yamashita, Y., 2014. Applications of Excitation Emission Matrix Fluorescence with Parallel Factor Analysis (EEM-PARAFAC) in Assessing Environmental Dynamics of Natural Dissolved Organic Matter (DOM) in Aquatic Environments: A Review, in: Advances in the Physicochemical Characterization of Dissolved Organic Matter: Impact on Natural and Engineered Systems, ACS Symposium Series. American Chemical Society, pp. 27–73. https://doi.org/10.1021/bk-2014-1160.ch003

Jeong, Y.-J., Park, H.-J., Baek, N., Seo, B.-S., Lee, K.-S., Kwak, J.-H., Choi, S.-K., Lee, S.-M., Yoon, K.-S., Lim, S.-S., Choi, W.-J., 2023. Assessment of sources variability of riverine particulate organic matter with land use and rainfall changes using a three-indicator (δ13C, δ15N, and C/N) Bayesian mixing model. Environmental Research 216, 114653. https://doi.org/10.1016/j.envres.2022.114653

Jiang, T., Skyllberg, U., Björn, E., Green, N.W., Tang, J., Wang, D., Gao, J., Li, C., 2017. Characteristics of dissolved organic matter (DOM) and relationship with dissolved mercury in Xiaoqing River-Laizhou Bay estuary, Bohai Sea, China. Environmental Pollution 223, 19–30. https://doi.org/10.1016/j.envpol.2016.12.006

Jorner, K., Tomberg, A., Bauer, C., Sköld, C., Norrby, P.-O., 2021. Organic reactivity from mechanism to machine learning. Nat Rev Chem 5, 240–255. https://doi.org/10.1038/s41570-021-00260-x

Ju, A., Wang, H., Wang, L., Weng, Y., 2023. Application of machine learning algorithms for prediction of ultraviolet absorption spectra of chromophoric dissolved organic matter (CDOM) in seawater. Frontiers in Marine Science 10.

Kalle, K., 1966. The problem of the Gelbstoff in the sea. Oceanagr. Mar. Biol. Ann. Rev. 4, 91–104.

Kalle, K., 1949. Fluoreszenz und gelbstoff in Bottnishen und Finnischen Meerbusen. Dtsch. Hydrogr. Z. 2, 117–124.

Kalle, K., 1938. Zum Problem der Meereswasserfarbe. Ann. Hydrol. Mar. Mitt. 66, 1–13.

Kang, M., He, J., Jia, G., 2023. Evaluation of heterocyte glycolipids with a hexose sugar moiety for tracing terrestrial organic matter in the South China Sea. Chemical Geology 635, 121604. https://doi.org/10.1016/j.chemgeo.2023.121604

Kawamura, K., Ishiwatari, R., Ogura, K., 1987. Early diagenesis of organic matter in the water column and sediments: Microbial degradation and resynthesis of lipids in Lake Haruna. Organic Geochemistry 11, 251–264. https://doi.org/10.1016/0146-6380(87)90036-2

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Keil, R.G., Mayer, L.M., Quay, P.D., Richey, J.E., Hedges, J.I., 1997. Loss of organic matter from riverine particles in deltas. Geochimica et Cosmochimica Acta 61, 1507–1511. https://doi.org/10.1016/S0016-7037(97)00044-6

Kim, J., Kim, Y., Park, S.E., Kim, T.-H., Kim, B.-G., Kang, D.-J., Rho, T., 2022. Impact of aquaculture on distribution of dissolved organic matter in coastal Jeju Island, Korea, based on absorption and fluorescence spectroscopy. Environ Sci Pollut Res 29, 553–563. https://doi.org/10.1007/s11356-021-15553-3

Kim, J., Wang, X., Kang, C., Yu, J., Li, P., 2021. Forecasting air pollutant concentration using a novel spatiotemporal deep learning model based on clustering, feature selection and empirical wavelet transform. Science of The Total Environment 801, 149654. https://doi.org/10.1016/j.scitotenv.2021.149654

Kim, J.-H., Ludwig, W., Buscail, R., Dorhout, D., Sinninghe Damsté, J.S., 2015. Tracing tetraether lipids from source to sink in the Rhône River system (NW Mediterranean). Frontiers in Earth Science 3.

Kirkels, F.M.S.A., Ponton, C., Galy, V., West, A.J., Feakins, S.J., Peterse, F., 2020. From Andes to Amazon: Assessing Branched Tetraether Lipids as Tracers for Soil Organic Carbon in the Madre de Dios River System. Journal of Geophysical Research: Biogeosciences 125, e2019JG005270. https://doi.org/10.1029/2019JG005270

Kirkels, F.M.S.A., Usman, M.O., Peterse, F., 2022a. Distinct sources of bacterial branched GMGTs in the Godavari River basin (India) and Bay of Bengal sediments. Organic Geochemistry 167, 104405. https://doi.org/10.1016/j.orggeochem.2022.104405

Kirkels, F.M.S.A., Zwart, H.M., Usman, M.O., Hou, S., Ponton, C., Giosan, L., Eglinton, T.I., Peterse, F., 2022b. From soil to sea: sources and transport of organic carbon traced by tetraether lipids in the monsoonal Godavari River, India. Biogeosciences 19, 3979–4010. https://doi.org/10.5194/bg-19-3979-2022

Kolb, P., Zorndt, A., Burchard, H., Gräwe, U., Kösters, F., 2022. Modelling the impact of anthropogenic measures on saltwater intrusion in the Weser estuary. Ocean Science 18, 1725–1739. https://doi.org/10.5194/os-18-1725-2022

Koprivnjak, J.-F., Pfromm, P.H., Ingall, E., Vetter, T.A., Schmitt-Kopplin, P., Hertkorn, N., Frommberger, M., Knicker, H., Perdue, E.M., 2009. Chemical and spectroscopic characterization of marine dissolved organic matter isolated using coupled reverse osmosis–electrodialysis. Geochimica et Cosmochimica Acta 73, 4215–4231. https://doi.org/10.1016/j.gca.2009.04.010

Kou, Q., Zhu, L., Ju, J., Wang, J., Xu, T., Li, C., Ma, Q., 2022. Influence of salinity on glycerol dialkyl glycerol tetraether-based indicators in Tibetan Plateau lakes: Implications for paleotemperature and paleosalinity reconstructions. Palaeogeography, Palaeoclimatology, Palaeoecology 601, 111127. https://doi.org/10.1016/j.palaeo.2022.111127

Kowalczuk, P., Cooper, W.J., Whitehead, R.F., Durako, M.J., Sheldon, W., 2003. Characterization of CDOM in an organic-rich river and surrounding coastal ocean in the South Atlantic Bight. Aquat. Sci. 65, 384–401. https://doi.org/10.1007/s00027-003-0678-1

Kowalczuk, P., Durako, M.J., Young, H., Kahn, A.E., Cooper, W.J., Gonsior, M., 2009. Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Interannual variability. Marine Chemistry 113, 182–196. https://doi.org/10.1016/j.marchem.2009.01.015

Kowalczuk, P., Tilstone, G.H., Zabłocka, M., Röttgers, R., Thomas, R., 2013. Composition of dissolved organic matter along an Atlantic Meridional Transect from fluorescence spectroscopy and Parallel Factor Analysis. Marine Chemistry 157, 170–184. https://doi.org/10.1016/j.marchem.2013.10.004

Koza, J.R., Bennett, F.H., Andre, D., Keane, M.A., 1996. Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming, in: Gero, J.S., Sudweeks, F. (Eds.), Artificial Intelligence in Design '96. Springer Netherlands, Dordrecht, pp. 151–170. https://doi.org/10.1007/978-94-009-0279-4_9

Kuhlisch, C., Schleyer, G., Shahaf, N., Vincent, F., Schatz, D., Vardi, A., 2021. Viral infection of algal blooms leaves a unique metabolic footprint on the dissolved organic matter in the ocean. Science Advances 7, eabf4680. https://doi.org/10.1126/sciadv.abf4680

Lagarda, M.J., García-Llatas, G., Farré, R., 2006. Analysis of phytosterols in foods. Journal of Pharmaceutical and Biomedical Analysis, Nutraceuticals Analysis 41, 1486–1496. https://doi.org/10.1016/j.jpba.2006.02.052

Lai, J., Zou, Y., Zhang, J., Peres-Neto, P.R., 2022. Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. Methods in Ecology and Evolution 13, 782–788. https://doi.org/10.1111/2041-210X.13800

Lakowicz, J.R. (Ed.), 2006. Introduction to Fluorescence, in: Principles of Fluorescence Spectroscopy. Springer US, Boston, MA, pp. 1–26. https://doi.org/10.1007/978-0-387-46312-4_1

Lamb, A.L., Wilson, G.P., Leng, M.J., 2006. A review of coastal palaeoclimate and relative sea-level reconstructions using δ13C and C/N ratios in organic material. Earth-Science Reviews, ISOtopes in PALaeoenvironmental reconstruction (ISOPAL) 75, 29–57. https://doi.org/10.1016/j.earscirev.2005.10.003

Lambert, T., Teodoru, C.R., Nyoni, F.C., Bouillon, S., Darchambeau, F., Massicotte, P., Borges, A.V., 2016. Along-stream transport and transformation of dissolved organic matter in a large tropical river. Biogeosciences 13, 2727–2741. https://doi.org/10.5194/bg-13-2727-2016

Lattaud, J., Kim, J.-H., De Jonge, C., Zell, C., Sinninghe Damsté, J.S., Schouten, S., 2017. The C32 alkane-1,15-diol as a tracer for riverine input in coastal seas. Geochimica et Cosmochimica Acta 202, 146–158. https://doi.org/10.1016/j.gca.2016.12.030

Leavitt, P.R., Brock, C.S., Ebel, C., Patoine, A., 2006. Landscape-scale effects of urban nitrogen on a chain of freshwater lakes in central North America. Limnology and Oceanography 51, 2262–2277. https://doi.org/10.4319/lo.2006.51.5.2262

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

Leeming, R., Ball, A., Ashbolt, N., Nichols, P., 1996. Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters. Water Research 30, 2893–2900. https://doi.org/10.1016/S0043-1354(96)00011-5

Leenheer, J.A., Croué, J.-P., 2003. Peer reviewed: characterizing aquatic dissolved organic matter. Environmental science & technology 37, 18A-26A.

Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research 18, 559–563.

Li, A., Bernal, S., Kohler, B., Thomas, S.A., Martí, E., Packman, A.I., 2021. Residence Time in Hyporheic Bioactive Layers Explains Nitrate Uptake in Streams. Water Resources Research 57, e2020WR027646. https://doi.org/10.1029/2020WR027646

Li, C., Sun, L., Jia, J., Cai, Y., Wang, X., 2016. Risk assessment of water pollution sources based on an integrated k-means clustering and set pair analysis method in the region of Shiyan, China. Science of The Total Environment 557–558, 307–316. https://doi.org/10.1016/j.scitotenv.2016.03.069

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, 131–154. https://doi.org/10.1080/10643389.2017.1309186

Li, S., Meng, L., Zhao, C., Gu, Y., Spencer, R.G.M., Álvarez–Salgado, X.A., Kellerman, A.M., McKenna, A.M., Huang, T., Yang, H., Huang, C., 2023. Spatiotemporal response of dissolved organic matter diversity to natural and anthropogenic forces along the whole mainstream of the Yangtze River. Water Research 234, 119812. https://doi.org/10.1016/j.watres.2023.119812

Liao, Z., Lu, J., Xie, K., Wang, Y., Yuan, Y., 2023. Prediction of Photochemical Properties of Dissolved Organic Matter Using Machine Learning. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.2c07545

Liu, D., Gao, H., Yu, H., Song, Y., 2022. Applying EEM-PARAFAC combined with moving-window 2DCOS and structural equation modeling to characterize binding properties of Cu (II) with DOM from different sources in an urbanized river. Water Research 227, 119317. https://doi.org/10.1016/j.watres.2022.119317

Liu, K., Xiao, X., Zhang, D., Ding, Y., Li, L., Zhao, M., 2021. Quantitative estimates of organic carbon contributions to the river-estuary-marine system in the Jiaozhou Bay, China. Ecological Indicators 129, 107929. https://doi.org/10.1016/j.ecolind.2021.107929

Liu, X.-L., Summons, R.E., Hinrichs, K.-U., 2012. Extending the known range of glycerol ether lipids in the environment: structural assignments based on tandem mass spectral fragmentation patterns. Rapid Communications in Mass Spectrometry 26, 2295–2302. https://doi.org/10.1002/rcm.6355

Liu, Y., Hu, Yucheng, Yu, C., Gao, Y., Liu, Z., Mostofa, K.M.G., Li, S., Hu, Yumei, Yu, G., 2023. Spatiotemporal optical properties of dissolved organic matter in a sluice-controlled coastal plain river with both salinity and trophic gradients. Journal of Environmental Sciences 129, 1–15. https://doi.org/10.1016/j.jes.2022.09.031

Liu, Y., Sun, J., Wang, X., Liu, X., Wu, X., Chen, Z., Gu, T., Wang, W., Yu, L., Guo, Y., 2021. Fluorescence characteristics of chromophoric dissolved organic matter in the Eastern Indian Ocean: a case study of three subregions. Frontiers in Marine Science 8, 742595.

Liu, Y., Ye, Q., Huang, W.-L., Feng, L., Wang, Y.-H., Xie, Z., Yong, S.-S., Zhang, S., Jiang, B., Zheng, Y., Wang, J.-J., 2020. Spectroscopic and molecular-level characteristics of dissolved organic matter in the Pearl River Estuary, South China. Science of The Total Environment 710, 136307. https://doi.org/10.1016/j.scitotenv.2019.136307

Lloyd, J.B.F., 1971. Synchronized Excitation of Fluorescence Emission Spectra. Nature Physical Science 231, 64–65. https://doi.org/10.1038/physci231064a0

Lønborg, C., Carreira, C., Jickells, T., Álvarez-Salgado, X.A., 2020. Impacts of Global Change on Ocean Dissolved Organic Carbon (DOC) Cycling. Frontiers in Marine Science 7.

Lopes dos Santos, R.A., Vane, C.H., 2020. Molecular and bulk geochemical proxies in sediments from the Conwy Estuary, UK. Organic Geochemistry 150, 104119. https://doi.org/10.1016/j.orggeochem.2020.104119

Lopes dos Santos, R.A., Vane, C.H., 2016. Signatures of tetraether lipids reveal anthropogenic overprinting of natural organic matter in sediments of the Thames Estuary, UK. Organic Geochemistry 93, 68–76. https://doi.org/10.1016/j.orggeochem.2016.01.003

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874 [cs, stat].

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

Luo, Y., Zhang, Y., Lang, M., Guo, X., Xia, T., Wang, T., Jia, H., Zhu, L., 2021. Identification of sources, characteristics and photochemical transformations of dissolved organic matter with EEM-PARAFAC in the Wei River of China. Frontiers of Environmental Science & Engineering 15, 1–10.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. University of California Press, pp. 281–298.

Mangalgiri, K.P., Timko, S.A., Gonsior, M., Blaney, L., 2017. PARAFAC Modeling of Irradiation- and Oxidation-Induced Changes in Fluorescent Dissolved Organic Matter Extracted from Poultry Litter. Environ. Sci. Technol. 51, 8036–8047. https://doi.org/10.1021/acs.est.6b06589

Mariotti, A., Lancelot, C., Billen, G., 1984. Natural isotopic composition of nitrogen as a tracer of origin for suspended organic matter in the Scheldt estuary. Geochimica et Cosmochimica Acta 48, 549–555. https://doi.org/10.1016/0016-7037(84)90283-7

Martens, J., Mueller, C.W., Joshi, P., Rosinger, C., Maisch, M., Kappler, A., Bonkowski, M., Schwamborn, G., Schirrmeister, L., Rethemeyer, J., 2023. Stabilization of mineral-associated organic carbon in Pleistocene permafrost. Nat Commun 14, 2120. https://doi.org/10.1038/s41467-023-37766-5

Martínez-Sosa, P., Tierney, J.E., 2019. Lacustrine brGDGT response to microcosm and mesocosm incubations. Organic Geochemistry 127, 12–22. https://doi.org/10.1016/j.orggeochem.2018.10.011

Martínez-Sosa, P., Tierney, J.E., Stefanescu, I.C., Dearing Crampton-Flood, E., Shuman, B.N., Routson, C., 2021. A global Bayesian temperature calibration for lacustrine brGDGTs. Geochimica et Cosmochimica Acta 305, 87–105. https://doi.org/10.1016/j.gca.2021.04.038

Martins, C.C., Seyffert, B.H., Braun, J.A.F., Fillmann, G., 2011. Input of organic matter in a large south american tropical estuary (Paranaguá Estuarine System, Brazil) indicated by

sedimentary sterols and multivariate statistical approach. J. Braz. Chem. Soc. 22, 1585–1594. https://doi.org/10.1590/S0103-50532011000800023

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

McCalley, D.V., Cooke, M., Nickless, G., 1981. Effect of sewage treatment on faecal sterols. Water Research 15, 1019–1025. https://doi.org/10.1016/0043-1354(81)90211-6

McCallister, S.L., Bauer, J.E., Ducklow, H.W., Canuel, E.A., 2006. Sources of estuarine dissolved and particulate organic matter: A multi-tracer approach. Organic Geochemistry 37, 454–468. https://doi.org/10.1016/j.orggeochem.2005.12.005

McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 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

Meng, F., Huang, G., Yang, X., Li, Z., Li, J., Cao, J., Wang, Z., Sun, L., 2013. Identifying the sources and fate of anthropogenically impacted dissolved organic matter (DOM) in urbanized rivers. Water Research 47, 5027–5039. https://doi.org/10.1016/j.watres.2013.05.043

Meyers, P.A., Ishiwatari, R., 1993. Lacustrine organic geochemistry—an overview of indicators of organic matter sources and diagenesis in lake sediments. Organic Geochemistry 20, 867–900. https://doi.org/10.1016/0146-6380(93)90100-P

Miano, T.M., Martin, J.P., Sposito, G., 1988. Flourescence Spectroscopy of Humic Substances. Soil Science Society of America Journal 52, 1016–1019. https://doi.org/10.2136/sssaj1988.03615995005200040021x

Michael Beman, J., Arrigo, K.R., Matson, P.A., 2005. Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. Nature 434, 211–214. https://doi.org/10.1038/nature03370

Michellod, D., Bien, T., Birgel, D., Violette, M., Kleiner, M., Fearn, S., Zeidler, C., Gruber-Vodicka, H.R., Dubilier, N., Liebeke, M., 2023. De novo phytosterol synthesis in animals. Science 380, 520–526. https://doi.org/10.1126/science.add7830

Mielnik, L., Kowalczuk, P., 2018. Optical characteristic of humic acids from lake sediments by excitation-emission matrix fluorescence with PARAFAC model. J Soils Sediments 18, 2851–2862. https://doi.org/10.1007/s11368-018-1947-x

Monroy, P., Hernández-García, E., Rossi, V., López, C., 2017. Modeling the dynamical sinking of biogenic particles in oceanic flow. Nonlinear Processes in Geophysics 24, 293–305. https://doi.org/10.5194/npg-24-293-2017

Moore, J.W., Semmens, B.X., 2008. Incorporating uncertainty and prior information into stable isotope mixing models. Ecology Letters 11, 470–480. https://doi.org/10.1111/j.1461-0248.2008.01163.x

Moreau, R.A., Whitaker, B.D., Hicks, K.B., 2002. Phytosterols, phytostanols, and their conjugates in foods: structural diversity, quantitative analysis, and health-promoting uses. Progress in Lipid Research 41, 457–500. https://doi.org/10.1016/S0163-7827(02)00006-1

Mori, C., Santos, I.R., Brumsack, H.-J., Schnetger, B., Dittmar, T., Seidel, M., 2019. Non-conservative Behavior of Dissolved Organic Matter and Trace Metals (Mn, Fe, Ba) Driven by Porewater Exchange in a Subtropical Mangrove-Estuary. Frontiers in Marine Science 6.

Morii, H., Eguchi, T., Nishihara, M., Kakinuma, K., König, H., Koga, Y., 1998. A novel ether core lipid with H-shaped C80-isoprenoid hydrocarbon chain from the hyperthermophilic methanogen Methanothermus fervidus. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism 1390, 339–345. https://doi.org/10.1016/S0005-2760(97)00183-5

Mouchel, J.-M., Lucas, F.S., Moulin, L., Wurtzer, S., Euzen, A., Haghe, J.-P., Rocher, V., Azimi, S., Servais, P., 2021. Bathing Activities and Microbiological River Water Quality in the Paris Area: A Long-Term Perspective, in: Flipo, N., Labadie, P., Lestel, L. (Eds.), The Seine River Basin, The Handbook of Environmental Chemistry. Springer International Publishing, Cham, pp. 323–353. https://doi.org/10.1007/698_2019_397

Mudge, S.M., Bebianno, M.J., 1997. Sewage contamination following an accidental spillage in the Ria Formosa, Portugal. Marine Pollution Bulletin 34, 163–170. https://doi.org/10.1016/S0025-326X(96)00082-3

Mudge, S.M., Lintern, D.G., 1999. Comparison of Sterol Biomarkers for Sewage with other Measures in Victoria Harbour, B.C., Canada. Estuarine, Coastal and Shelf Science 48, 27–38. https://doi.org/10.1006/ecss.1999.0406

Müller, P.J., 1977. CN ratios in Pacific deep-sea sediments: Effect of inorganic ammonium and organic nitrogen compounds sorbed by clays. Geochimica et Cosmochimica Acta 41, 765–776. https://doi.org/10.1016/0016-7037(77)90047-3

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., 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

Murphy, K.R., Timko, S.A., Gonsior, M., Powers, L.C., Wünsch, U.J., Stedmon, C.A., 2018. Photochemistry Illuminates Ubiquitous Organic Matter Fluorescence Spectra. Environ. Sci. Technol. 52, 11243–11250. https://doi.org/10.1021/acs.est.8b02648

Naafs, B.D.A., Inglis, G.N., Zheng, Y., Amesbury, M.J., Biester, H., Bindler, R., Blewett, J., Burrows, M.A., del Castillo Torres, D., Chambers, F.M., Cohen, A.D., Evershed, R.P., Feakins, S.J., Gałka, M., Gallego-Sala, A., Gandois, L., Gray, D.M., Hatcher, P.G., Honorio Coronado, E.N., Hughes, P.D.M., Huguet, A., Könönen, M., Laggoun-Défarge, F., Lähteenoja, O., Lamentowicz, M., Marchant, R., McClymont, E., Pontevedra-Pombal, X., Ponton, C., Pourmand, A., Rizzuti, A.M., Rochefort, L., Schellekens, J., De Vleeschouwer, F., Pancost, R.D., 2017. Introducing global peat-specific temperature and pH calibrations based on brGDGT bacterial lipids. Geochimica et Cosmochimica Acta 208, 285–301. https://doi.org/10.1016/j.gca.2017.01.038

Naafs, B.D.A., McCormick, D., Inglis, G.N., Pancost, R.D., 2018. Archaeal and bacterial H-GDGTs are abundant in peat and their relative abundance is positively correlated with temperature. Geochimica et Cosmochimica Acta 227, 156–170. https://doi.org/10.1016/j.gca.2018.02.025

Narvaez-Montoya, C., Mahlknecht, J., Torres-Martínez, J.A., Mora, A., Bertrand, G., 2023. Seawater intrusion pattern recognition supported by unsupervised learning: A systematic review and application. Science of The Total Environment 864, 160933. https://doi.org/10.1016/j.scitotenv.2022.160933

Nellemann, C., Corcoran, E., 2009. Blue carbon: the role of healthy oceans in binding carbon: a rapid response assessment. UNEP/Earthprint.

Nelson, N.B., Siegel, D.A., 2013. The Global Distribution and Dynamics of Chromophoric Dissolved Organic Matter. Annual Review of Marine Science 5, 447–476. https://doi.org/10.1146/annurev-marine-120710-100751

Omanović, D., Marcinek, S., Santinelli, C., 2023. TreatEEM—A Software Tool for the Interpretation of Fluorescence Excitation-Emission Matrices (EEMs) of Dissolved Organic Matter in Natural Waters. Water 15, 2214. https://doi.org/10.3390/w15122214

Osburn, C.L., Anderson, N.J., Stedmon, C.A., Giles, M.E., Whiteford, E.J., McGenity, T.J., Dumbrell, A.J., Underwood, G.J.C., 2017. Shifts in the Source and Composition of Dissolved Organic Matter in Southwest Greenland Lakes Along a Regional Hydro-climatic Gradient. Journal of Geophysical Research: Biogeosciences 122, 3431–3445. https://doi.org/10.1002/2017JG003999

Osburn, C.L., Atar, J.N., Boyd, T.J., Montgomery, M.T., 2019. Antecedent precipitation influences the bacterial processing of terrestrial dissolved organic matter in a North Carolina estuary. Estuarine, Coastal and Shelf Science 221, 119–131. https://doi.org/10.1016/j.ecss.2019.03.016

Osburn, C.L., Mikan, M.P., Etheridge, J.R., Burchell, M.R., Birgand, F., 2015. Seasonal variation in the quality of dissolved and particulate organic matter exchanged between a salt marsh and its adjacent estuary. Journal of Geophysical Research: Biogeosciences 120, 1430–1449. https://doi.org/10.1002/2014JG002897

Park, J., Lee, W.H., Kim, K.T., Park, C.Y., Lee, S., Heo, T.-Y., 2022. Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. Science of The Total Environment 832, 155070. https://doi.org/10.1016/j.scitotenv.2022.155070

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

Parr, T.B., Cronan, C.S., Ohno, T., Findlay, S.E.G., Smith, S.M.C., Simon, K.S., 2015. Urbanization changes the composition and bioavailability of dissolved organic matter in headwater streams. Limnology and Oceanography 60, 885–900. https://doi.org/10.1002/lno.10060

Patriarca, C., Sedano-Núñez, V.T., Garcia, S.L., Bergquist, J., Bertilsson, S., Sjöberg, P.J., Tranvik, L.J., Hawkes, J.A., 2021. Character and environmental lability of cyanobacteria-derived dissolved organic matter. Limnology and Oceanography 66, 496–509.

Pearson, A., Ingalls, A.E., 2013. Assessing the Use of Archaeal Lipids as Marine Environmental Proxies. Annual Review of Earth and Planetary Sciences 41, 359–384. https://doi.org/10.1146/annurev-earth-050212-123947

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Peer, S., Vybornova, A., Saracevic, Z., Krampe, J., Zessner, M., Zoboli, O., 2022. Enhanced statistical evaluation of fluorescence properties to identify dissolved organic matter dynamics during river high-flow events. Science of The Total Environment 851, 158016. https://doi.org/10.1016/j.scitotenv.2022.158016

Peters, D.P.C., Havstad, K.M., Cushing, J., Tweedie, C., Fuentes, O., Villanueva-Rosales, N., 2014. Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. Ecosphere 5, art67. https://doi.org/10.1890/ES13-00359.1

Peterse, F., Kim, J.-H., Schouten, S., Kristensen, D.K., Koç, N., Sinninghe Damsté, J.S., 2009. Constraints on the application of the MBT/CBT palaeothermometer at high latitude environments (Svalbard, Norway). Organic Geochemistry 40, 692–699. https://doi.org/10.1016/j.orggeochem.2009.03.004

Pirnay, J.-P., Matthijs, S., Colak, H., Chablain, P., Bilocq, F., Van Eldere, J., De Vos, D., Zizi, M., Triest, L., Cornelis, P., 2005. Global Pseudomonas aeruginosa biodiversity as reflected in a Belgian river. Environmental Microbiology 7, 969–980. https://doi.org/10.1111/j.1462-2920.2005.00776.x

Pisani, O., Oros, D.R., Oyo-Ita, O.E., Ekpo, B.O., Jaffé, R., Simoneit, B.R.T., 2013. Biomarkers in surface sediments from the Cross River and estuary system, SE Nigeria: Assessment of organic matter sources of natural and anthropogenic origins. Applied Geochemistry 31, 239–250. https://doi.org/10.1016/j.apgeochem.2013.01.010

Podgorski, J., Berg, M., 2020. Global threat of arsenic in groundwater. Science 368, 845–850. https://doi.org/10.1126/science.aba1510

Porcal, P., Dillon, P.J., Molot, L.A., 2015. Temperature Dependence of Photodegradation of Dissolved Organic Matter to Dissolved Inorganic Carbon and Particulate Organic Carbon. PLOS ONE 10, e0128884. https://doi.org/10.1371/journal.pone.0128884

Porcal, P., Dillon, P.J., Molot, L.A., 2013. Seasonal changes in photochemical properties of dissolved organic matter in small boreal streams. Biogeosciences 10, 5533–5543. https://doi.org/10.5194/bg-10-5533-2013

Priyanka, M., Saravanakumar, M.P., 2022. New insights on aging mechanism of microplastics using PARAFAC analysis: Impact on 4-nitrophenol removal via Statistical Physics Interpretation. Science of The Total Environment 807, 150819. https://doi.org/10.1016/j.scitotenv.2021.150819

Qin, X., Yao, B., Jin, L., Zheng, X., Ma, J., Benedetti, M.F., Li, Y., Ren, Z., 2020. Characterizing soil dissolved organic matter in typical soils from China using fluorescence EEM–PARAFAC and UV–visible absorption. Aquatic Geochemistry 26, 71–88.

Quintana-Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., Morel, S., 2008. Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France. Journal of applied meteorology and climatology 47, 92–107.

Raberg, J.H., Miller, G.H., Geirsdóttir, Á., Sepúlveda, J., 2022. Near-universal trends in brGDGT lipid distributions in nature. Science Advances 8, eabm7625. https://doi.org/10.1126/sciadv.abm7625

Rada, J.P.A., Duarte, A.C., Pato, P., Cachada, A., Carreira, R.S., 2016. Sewage contamination of sediments from two Portuguese Atlantic coastal systems, revealed by fecal sterols. Marine Pollution Bulletin 103, 319–324. https://doi.org/10.1016/j.marpolbul.2016.01.010

Ralston, D.K., Geyer, W.R., 2019. Response to Channel Deepening of the Salinity Intrusion, Estuarine Circulation, and Stratification in an Urbanized Estuary. Journal of Geophysical Research: Oceans 124, 4784–4802. https://doi.org/10.1029/2019JC015006

Rampen, S.W., Abbas, B.A., Schouten, S., Sinninghe Damste, J.S., 2010. A comprehensive study of sterols in marine diatoms (Bacillariophyta): Implications for their use as tracers for diatom productivity. Limnology and Oceanography 55, 91–105. https://doi.org/10.4319/lo.2010.55.1.0091

Regier, P., Jaffé, R., 2016. Short-Term Dissolved Organic Carbon Dynamics Reflect Tidal, Water Management, and Precipitation Patterns in a Subtropical Estuary. Frontiers in Marine Science 3.

Rejano, F., Casquero-Vera, J.A., Lyamani, H., Andrews, E., Casans, A., Pérez-Ramírez, D., Alados-Arboledas, L., Titos, G., Olmo, F.J., 2023. Impact of urban aerosols on the cloud condensation activity using a clustering model. Science of The Total Environment 858, 159657. https://doi.org/10.1016/j.scitotenv.2022.159657

Repeta, D.J., 2015. Chapter 2 - Chemical Characterization and Cycling of Dissolved Organic Matter, in: Hansell, D.A., Carlson, C.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter (Second Edition). Academic Press, Boston, pp. 21–63. https://doi.org/10.1016/B978-0-12-405940-5.00002-9

Riopel, R., Caron, F., Siemann, S., 2014. Fluorescence Characterization of Natural Organic Matter at a Northern Ontario Wastewater Treatment Plant. Water Air Soil Pollut 225, 2126. https://doi.org/10.1007/s11270-014-2126-3

Romero, E., Garnier, J., Billen, G., Ramarson, A., Riou, P., Le Gendre, R., 2022. Assessing the water quality of the Seine land-to-sea continuum for three agro-food system scenarios. Frontiers in Marine Science 9.

Romero, E., Garnier, J., Billen, G., Ramarson, A., Riou, P., Le Gendre, R., 2019. Modeling the biogeochemical functioning of the Seine estuary and its coastal zone: Export, retention, and transformations. Limnology and Oceanography 64, 895–912. https://doi.org/10.1002/lno.11082

Romero, E., Le Gendre, R., Garnier, J., Billen, G., Fisson, C., Silvestre, M., Riou, P., 2016. Long-term water quality in the lower Seine: Lessons learned over 4 decades of monitoring. Environmental Science & Policy 58, 141–154. https://doi.org/10.1016/j.envsci.2016.01.016

Rommerskirchen, F., Eglinton, G., Dupont, L., Güntner, U., Wenzel, C., Rullkötter, J., 2003. A north to south transect of Holocene southeast Atlantic continental margin sediments: Relationship between aerosol transport and compound-specific δ13C land plant biomarker and pollen records. Geochemistry, Geophysics, Geosystems 4. https://doi.org/10.1029/2003GC000541

Rontani, J.-F., Belt, S.T., Amiraux, R., 2018. Biotic and abiotic degradation of the sea ice diatom biomarker IP25 and selected algal sterols in near-surface Arctic sediments. Organic Geochemistry 118, 73–88. https://doi.org/10.1016/j.orggeochem.2018.01.003

Russell, J.M., Hopmans, E.C., Loomis, S.E., Liang, J., Sinninghe Damsté, J.S., 2018. Distributions of 5- and 6-methyl branched glycerol dialkyl glycerol tetraethers (brGDGTs) in East African lake sediment: Effects of temperature, pH, and new lacustrine paleotemperature calibrations. Organic Geochemistry 117, 56–69. https://doi.org/10.1016/j.orggeochem.2017.12.003

Ryba, S.A., Burgess, R.M., 2002. Effects of sample preparation on the measurement of organic carbon, hydrogen, nitrogen, sulfur, and oxygen concentrations in marine sediments. Chemosphere 48, 139–147. https://doi.org/10.1016/S0045-6535(02)00027-9

Saeidnia, S., Manayi, A., Gohari, A.R., Abdollahi, M., 2014. The story of beta-sitosterol-a review. European journal of medicinal plants 4, 590.

Sanchez, N.P., Skeriotis, A.T., Miller, C.M., 2014. A PARAFAC-based long-term assessment of DOM in a multi-coagulant drinking water treatment scheme. Environmental science & technology 48, 1582–1591.

Santos, L., Santos, E.B.H., Dias, J.M., Cunha, A., Almeida, A., 2014. Photochemical and microbial alterations of DOM spectroscopic properties in the estuarine system Ria de Aveiro. Photochem Photobiol Sci 13, 1146–1159. https://doi.org/10.1039/c4pp00005f

Sarkar, S., Wilkes, H., Prasad, S., Brauer, A., Riedel, N., Stebich, M., Basavaiah, N., Sachse, D., 2014. Spatial heterogeneity in lipid biomarker distributions in the catchment and sediments of a crater lake in central India. Organic Geochemistry 66, 125–136. https://doi.org/10.1016/j.orggeochem.2013.11.009

Savoye, N., Aminot, A., Tréguer, P., Fontugne, M., Naulet, N., Kérouel, R., 2003. Dynamics of particulate organic matter δ15N and δ13C during spring phytoplankton blooms in a macrotidal ecosystem (Bay of Seine, France). Marine Ecology Progress Series 255, 27–41. https://doi.org/10.3354/meps255027

Savoye, N., David, V., Morisseau, F., Etcheber, H., Abril, G., Billy, I., Charlier, K., Oggian, G., Derriennic, H., Sautour, B., 2012. Origin and composition of particulate organic matter in a macrotidal turbid estuary: The Gironde Estuary, France. Estuarine, Coastal and Shelf Science 108, 16–28. https://doi.org/10.1016/j.ecss.2011.12.005

Schouten, S., Baas, M., Hopmans, E.C., Reysenbach, A.-L., Damsté, J.S.S., 2008. Tetraether membrane lipids of Candidatus "Aciduliprofundum boonei", a cultivated obligate thermoacidophilic euryarchaeote from deep-sea hydrothermal vents. Extremophiles 12, 119–124.

Schouten, S., Hopmans, E.C., Schefuß, E., Sinninghe Damsté, J.S., 2002. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? Earth and Planetary Science Letters 204, 265–274. https://doi.org/10.1016/S0012-821X(02)00979-2

Schouten, S., Hopmans, E.C., Sinninghe Damsté, J.S., 2013. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. Organic Geochemistry 54, 19–61. https://doi.org/10.1016/j.orggeochem.2012.09.006

Sciscenko, I., Arques, A., Micó, P., Mora, M., García-Ballesteros, S., 2022a. Emerging applications of EEM-PARAFAC for water treatment: a concise review. Chemical Engineering Journal Advances 10, 100286. https://doi.org/10.1016/j.ceja.2022.100286

Sciscenko, I., Mora, M., Micó, P., Escudero-Oñate, C., Oller, I., Arques, A., 2022b. EEM-PARAFAC as a convenient methodology to study fluorescent emerging pollutants degradation: (fluoro)quinolones oxidation in different water matrices. Science of The Total Environment 852, 158338. https://doi.org/10.1016/j.scitotenv.2022.158338

Sciscenko, I., Thị Mỹ Hắng, H., Escudero-Oñate, C., Oller, I., Arques, A., 2021. Fluorescence Spectroscopy and Chemometrics: A Simple and Easy Way for the Monitoring of Fluoroquinolone Mixture Degradation. ACS Omega 6, 4663–4671. https://doi.org/10.1021/acsomega.0c05370

Seritti, A., Morelli, E., Nannicini, L., Del Vecchio, R., 1994. Production of hydrophobic fluorescent organic matter by the marine diatom Phaeodactylum tricornutum. Chemosphere 28, 117–129. https://doi.org/10.1016/0045-6535(94)90205-4

Serre-Fredj, L., Chasselin, L., Jolly, O., Claquin, P., 2023. Complex drivers of primary production along an anthropised estuary (Seine estuary—France). Frontiers in Environmental Science 11.

Servais, P., Garcia-Armisen, T., George, I., Billen, G., 2007. Fecal bacteria in the rivers of the Seine drainage network (France): Sources, fate and modelling. Science of The Total

Environment, Human activity and material fluxes in a regional river basin: the Seine River watershed 375, 152–167. https://doi.org/10.1016/j.scitotenv.2006.12.010

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

Sikes, E.L., Uhle, M.E., Nodder, S.D., Howard, M.E., 2009. Sources of organic matter in a coastal marine environment: Evidence from n-alkanes and their δ13C distributions in the Hauraki Gulf, New Zealand. Marine Chemistry 113, 149–163. https://doi.org/10.1016/j.marchem.2008.12.003

Silva, T.R., Lopes, S.R.P., Spörl, G., Knoppers, B.A., Azevedo, D.A., 2012. Source characterization using molecular distribution and stable carbon isotopic composition of n-alkanes in sediment cores from the tropical Mundaú–Manguaba estuarine–lagoon system, Brazil. Organic Geochemistry, Advances in Organic Geochemistry 2011: Proceedings of the 25th International Meeting on Organic Geochemistry 53, 25–33. https://doi.org/10.1016/j.orggeochem.2012.05.009

Simjouw, J.-P., Minor, E.C., Mopper, K., 2005. Isolation and characterization of estuarine dissolved organic matter: Comparison of ultrafiltration and C18 solid-phase extraction techniques. Marine Chemistry 96, 219–235. https://doi.org/10.1016/j.marchem.2005.01.003

Singh, S., Dash, P., Sankar, M.S., Silwal, S., Lu, Y., Shang, P., Moorhead, R.J., 2019. Hydrological and Biogeochemical Controls of Seasonality in Dissolved Organic Matter Delivery to a Blackwater Estuary. Estuaries and Coasts 42, 439–454. https://doi.org/10.1007/s12237-018-0473-9

Sinninghe Damsté, J.S., 2016. Spatial heterogeneity of sources of branched tetraethers in shelf systems: The geochemistry of tetraethers in the Berau River delta (Kalimantan, Indonesia). Geochimica et Cosmochimica Acta 186, 13–31. https://doi.org/10.1016/j.gca.2016.04.033

Sinninghe Damsté, J.S., Rijpstra, W.I.C., Hopmans, E.C., Weijers, J.W.H., Foesel, B.U., Overmann, J., Dedysh, S.N., 2011. 13,16-Dimethyl octacosanedioic acid (iso-diabolic acid), a common membrane-spanning lipid of Acidobacteria subdivisions 1 and 3. Appl Environ Microbiol 77, 4147–4154. https://doi.org/10.1128/AEM.00466-11

Sinninghe Damsté, J.S., C. Hopmans, E., D. Pancost, R., Schouten, S., J. Geenevasen, J.A., 2000. Newly discovered non-isoprenoid glycerol dialkyl glycerol tetraether lipids in sediments. Chemical Communications 0, 1683–1684. https://doi.org/10.1039/B004517I

Sluijs, A., Frieling, J., Inglis, G.N., Nierop, K.G.J., Peterse, F., Sangiorgi, F., Schouten, S., 2020. Late Paleocene–early Eocene Arctic Ocean sea surface temperatures: reassessing biomarker paleothermometry at Lomonosov Ridge. Climate of the Past 16, 2381–2400. https://doi.org/10.5194/cp-16-2381-2020

Smith, R.W., Bianchi, T.S., Li, X., 2012. A re-evaluation of the use of branched GDGTs as terrestrial biomarkers: Implications for the BIT Index. Geochimica et Cosmochimica Acta 80, 14–29. https://doi.org/10.1016/j.gca.2011.11.025

Sobczak, W.V., Cloern, J.E., Jassby, A.D., Müller-Solger, A.B., 2002. Bioavailability of organic matter in a highly disturbed estuary: The role of detrital and algal resources. Proceedings of the National Academy of Sciences 99, 8101–8105. https://doi.org/10.1073/pnas.122614399

Søndergaard, M., Stedmon, C.A., Borch, N.H., 2003. Fate of terrigenous dissolved organic matter (DOM) in estuaries: Aggregation and bioavailability. Ophelia 57, 161–176. https://doi.org/10.1080/00785236.2003.10409512

Stedmon, C.A., Bro, R., 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. Limnology and Oceanography: Methods 6, 572–579. https://doi.org/10.4319/lom.2008.6.572

Stedmon, C.A., Markager, S., 2005. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. Limnology and Oceanography 50, 686–697. https://doi.org/10.4319/lo.2005.50.2.0686

Stedmon, C.A., Markager, S., Bro, R., 2003a. 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

Stedmon, C.A., Markager, S., Bro, R., 2003b. 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

Stedmon, C.A., Markager, S., Tranvik, L., Kronberg, L., Slätis, T., Martinsen, W., 2007. Photochemical production of ammonium and transformation of dissolved organic matter in

the Baltic Sea. Marine Chemistry 104, 227–240. https://doi.org/10.1016/j.marchem.2006.11.005

Stedmon, C.A., Nelson, N.B., 2015. Chapter 10 - The Optical Properties of DOM in the Ocean, in: Hansell, D.A., Carlson, C.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter (Second Edition). Academic Press, Boston, pp. 481–508. https://doi.org/10.1016/B978-0-12-405940-5.00010-8

Strickland, J.D.H., Parsons, T.R., 1972. A Practical Handbook of Seawater Analysis, 2nd edition. https://doi.org/10.25607/OBP-1791

Stubbins, A., Lapierre, J.-F., Berggren, M., Prairie, Y.T., Dittmar, T., del Giorgio, P.A., 2014. What's in an EEM? Molecular Signatures Associated with Dissolved Organic Fluorescence in Boreal Canada. Environ. Sci. Technol. 48, 10598–10606. https://doi.org/10.1021/es502086e

Swanson, K., Wu, E., Zhang, A., Alizadeh, A.A., Zou, J., 2023. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell 186, 1772–1791. https://doi.org/10.1016/j.cell.2023.01.035

Tahmasebi, P., Kamrava, S., Bai, T., Sahimi, M., 2020. Machine learning in geo- and environmental sciences: From small to large scale. Advances in Water Resources 142, 103619. https://doi.org/10.1016/j.advwatres.2020.103619

Tang, G., Wang, Q., 2022. Impact of environmental factors and tributary contributions on tidal dissolved organic matter dynamics. Chemosphere 308, 136384. https://doi.org/10.1016/j.chemosphere.2022.136384

Tang, X., Naafs, B.D.A., Pancost, R.D., Liu, Z., Fan, T., Zheng, Y., 2021. Exploring the Influences of Temperature on "H-Shaped" Glycerol Dialkyl Glycerol Tetraethers in a Stratigraphic Context: Evidence From Two Peat Cores Across the Late Quaternary. Frontiers in Earth Science 8.

Tao, H., Wu, T., Aldeghi, M., Wu, T.C., Aspuru-Guzik, A., Kumacheva, E., 2021. Nanoparticle synthesis assisted by machine learning. Nat Rev Mater 6, 701–716. https://doi.org/10.1038/s41578-021-00337-5

Tao, K., Xu, Y., Wang, Yinghui, Wang, Yuntao, He, D., 2021. Source, sink and preservation of organic matter from a machine learning approach of polar lipid tracers in sediments and

soils from the Yellow River and Bohai Sea, eastern China. Chemical Geology 582, 120441. https://doi.org/10.1016/j.chemgeo.2021.120441

Thibault, A., 2018. Dynamique de la matière organique dans la Seine : approche globale et moléculaire (phdthesis). Sorbonne Université.

Thibault, A., Derenne, S., Parlanti, E., Anquetil, C., Sourzac, M., Budzinski, H., Fuster, L., Laverman, A., Roose-Amsaleg, C., Viollier, E., Huguet, A., 2019. Dynamics of organic matter in the Seine Estuary (France): Bulk and structural approaches. Marine Chemistry 212, 108–119. https://doi.org/10.1016/j.marchem.2019.04.007

Thurman, E.M., Malcolm, R.L., 1981. Preparative isolation of aquatic humic substances. Environ. Sci. Technol. 15, 463–466. https://doi.org/10.1021/es00086a012

Tierney, J.E., Russell, J.M., 2009. Distributions of branched GDGTs in a tropical lake system: Implications for lacustrine application of the MBT/CBT paleoproxy. Organic Geochemistry 40, 1032–1036. https://doi.org/10.1016/j.orggeochem.2009.04.014

Touron, A., Berthe, T., Gargala, G., Fournier, M., Ratajczak, M., Servais, P., Petit, F., 2007. Assessment of faecal contamination and the relationship between pathogens and faecal bacterial indicators in an estuarine environment (Seine, France). Marine Pollution Bulletin 54, 1441–1450. https://doi.org/10.1016/j.marpolbul.2007.05.009

Tranvik, L.J., 1992. Allochthonous dissolved organic matter as an energy source for pelagic bacteria and the concept of the microbial loop, in: Salonen, K., Kairesalo, T., Jones, R.I. (Eds.), Dissolved Organic Matter in Lacustrine Ecosystems: Energy Source and System Regulator, Developments in Hydrobiology. Springer Netherlands, Dordrecht, pp. 107–114. https://doi.org/10.1007/978-94-011-2474-4_8

Uda, I., Sugai, A., Itoh, Y.H., Itoh, T., 2001. Variation in molecular species of polar lipids from Thermoplasma acidophilum depends on growth temperature. Lipids 36, 103–105. https://doi.org/10.1007/s11745-001-0914-2

Vane, C.H., Kim, A.W., McGowan, S., Leng, M.J., Heaton, T.H.E., Kendrick, C.P., Coombs, P., Yang, H., Swann, G.E.A., 2010. Sedimentary records of sewage pollution using faecal markers in contrasting peri-urban shallow lakes. Science of The Total Environment 409, 345–356. https://doi.org/10.1016/j.scitotenv.2010.09.033

Venkatesan, M.I., Mirsadeghi, F.H., 1992. Coprostanol as sewage tracer in McMurdo Sound, Antarctica. Marine Pollution Bulletin, Environmental Awareness in Antarctica: History,

Problems, and Future Solutions 25, 328–333. https://doi.org/10.1016/0025-326X(92)90691-X

Véquaud, P., Thibault, A., Derenne, S., Anquetil, C., Collin, S., Contreras, S., Nottingham, A.T., Sabatier, P., Werne, J.P., Huguet, A., 2022. FROG: A global machine-learning temperature calibration for branched GDGTs in soils and peats. Geochimica et Cosmochimica Acta 318, 468–494. https://doi.org/10.1016/j.gca.2021.12.007

Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyroux, J.-M., Blanchard, M., Baillon, M., 2010. Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite. Hydrology and Earth System Sciences 14, 459–478.

Vidal, L.O., Lambert, T., Cotovicz Jr., L.C., Bernardes, M.C., Sobrinho, R., Thompson, F., Garcia, G.D., Knoppers, B.A., Gatts, P.V., Régis, C.R., Abril, G., Rezende, C.E., 2023. Seasonal and diel modulation of DOM in a mangrove-dominated estuary. Science of The Total Environment 857, 159045. https://doi.org/10.1016/j.scitotenv.2022.159045

Vione, D., Minella, M., Maurino, V., Minero, C., 2014. Indirect Photochemistry in Sunlit Surface Waters: Photoinduced Production of Reactive Transient Species. Chemistry – A European Journal 20, 10590–10606. https://doi.org/10.1002/chem.201400413

Volkman, J., 2003. Sterols in microorganisms. Appl Microbiol Biotechnol 60, 495–506. https://doi.org/10.1007/s00253-002-1172-8

Volkman, J.K., 1986. A review of sterol markers for marine and terrigenous organic matter. Organic Geochemistry 9, 83–99. https://doi.org/10.1016/0146-6380(86)90089-6

Volkman, J.K., Tanoue, E., 2002. Chemical and Biological Studies of Particulate Organic Matter in the Ocean. Journal of Oceanography 58, 265–279. https://doi.org/10.1023/A:1015809708632

Wang, C., Zhang, C., Wang, Yameng, Jia, G., Wang, Yaping, Zhu, C., Yu, Q., Zou, X., 2022. Anthropogenic perturbations to the fate of terrestrial organic matter in a river-dominated marginal sea. Geochimica et Cosmochimica Acta 333, 242–262. https://doi.org/10.1016/j.gca.2022.07.012

Wang, H., An, Z., Lu, H., Zhao, Z., Liu, W., 2020. Calibrating bacterial tetraether distributions towards in situ soil temperature and application to a loess-paleosol sequence. Quaternary Science Reviews 231, 106172. https://doi.org/10.1016/j.quascirev.2020.106172

Wang, H., Liu, W., He, Y., Zhou, A., Zhao, H., Liu, H., Cao, Y., Hu, J., Meng, B., Jiang, J., Kolpakova, M., Krivonogov, S., Liu, Z., 2021. Salinity-controlled isomerization of lacustrine brGDGTs impacts the associated MBT5ME' terrestrial temperature index. Geochimica et Cosmochimica Acta 305, 33–48. https://doi.org/10.1016/j.gca.2021.05.004

Wang, X., Zhang, M., Liu, L., Wang, Z., Lin, K., 2022. Using EEM-PARAFAC to identify and trace the pollution sources of surface water with receptor models in Taihu Lake Basin, China. Journal of Environmental Management 321, 115925. https://doi.org/10.1016/j.jenvman.2022.115925

Wang, Y., Zhang, M., Zhang, D., Shen, Z., 2016. The influence of sediment particle size on the properties of adsorbed dissolved organic matter in the Yangtze Estuary and its interactions with As/Sb. Marine Pollution Bulletin 105, 351–358. https://doi.org/10.1016/j.marpolbul.2015.10.070

Ward, N.D., Bianchi, T.S., Sawakuchi, H.O., Gagne-Maynard, W., Cunha, A.C., Brito, D.C., Neu, V., de Matos Valerio, A., da Silva, R., Krusche, A.V., Richey, J.E., Keil, R.G., 2016. The reactivity of plant-derived organic matter and the potential importance of priming effects along the lower Amazon River. Journal of Geophysical Research: Biogeosciences 121, 1522–1539. https://doi.org/10.1002/2016JG003342

Weete, J.D., Abril, M., Blackwell, M., 2010. Phylogenetic Distribution of Fungal Sterols. PLOS ONE 5, e10899. https://doi.org/10.1371/journal.pone.0010899

Wei, L.-L., Zhao, Q.-L., Xue, S., Jia, T., Tang, F., You, P.-Y., 2009. Behavior and characteristics of DOM during a laboratory-scale horizontal subsurface flow wetland treatment: Effect of DOM derived from leaves and roots. Ecological Engineering 35, 1405–1414. https://doi.org/10.1016/j.ecoleng.2009.05.016

Wei, X., Garnier, J., Thieu, V., Passy, P., Le Gendre, R., Billen, G., Akopian, M., Laruelle, G.G., 2022. Nutrient transport and transformation in macrotidal estuaries of the French Atlantic coast: a modeling approach using the Carbon-Generic Estuarine Model. Biogeosciences 19, 931–955. https://doi.org/10.5194/bg-19-931-2022

Weigelhofer, G., Jirón, T.S., Yeh, T.-C., Steniczka, G., Pucher, M., 2020. Dissolved Organic Matter Quality and Biofilm Composition Affect Microbial Organic Matter Uptake in Stream Flumes. Water 12, 3246. https://doi.org/10.3390/w12113246

Weijers, J.W.H., Schouten, S., Hopmans, E.C., Geenevasen, J.A.J., David, O.R.P., Coleman, J.M., Pancost, R.D., Damsté, J.S.S., 2006. Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. Environmental Microbiology 8, 648–657. https://doi.org/10.1111/j.1462-2920.2005.00941.x

Weijers, J.W.H., Schouten, S., van den Donker, J.C., Hopmans, E.C., Sinninghe Damsté, J.S., 2007. Environmental controls on bacterial tetraether membrane lipid distribution in soils. Geochimica et Cosmochimica Acta 71, 703–713. https://doi.org/10.1016/j.gca.2006.10.003

Weijers, J.W.H., Wiesenberg, G.L.B., Bol, R., Hopmans, E.C., Pancost, R.D., 2010. Carbon isotopic composition of branched tetraether membrane lipids in soils suggest a rapid turnover and a heterotrophic life style of their source organism(s). Biogeosciences 7, 2959–2973. https://doi.org/10.5194/bg-7-2959-2010

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

Wells, M.J.M., Hooper, J., Mullins, G.A., Bell, K.Y., 2022. Development of a fluorescence EEM-PARAFAC model for potable water reuse monitoring: Implications for inter-component protein–fulvic–humic interactions. Science of The Total Environment 820, 153070. https://doi.org/10.1016/j.scitotenv.2022.153070

Wheeler, K.I., Levia, D.F., Hudson, J.E., 2017. Tracking senescence-induced patterns in leaf litter leachate using parallel factor analysis (PARAFAC) modeling and self-organizing maps. Journal of Geophysical Research: Biogeosciences 122, 2233–2250. https://doi.org/10.1002/2016JG003677

Wilhelm, S.W., Suttle, C.A., 1999. Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. BioScience 49, 781–788. https://doi.org/10.2307/1313569

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., Xenopoulos, M.A., 2009. Effects of agricultural land use on the composition of fluvial dissolved organic matter. Nature Geosci 2, 37–41. https://doi.org/10.1038/ngeo391

Wolf, R., Thrane, J.-E., Hessen, D.O., Andersen, T., 2018. Modelling ROS formation in boreal lakes from interactions between dissolved organic matter and absorbed solar photon flux. Water Research 132, 331–339. https://doi.org/10.1016/j.watres.2018.01.025

Wu, J., Yang, H., Pancost, R.D., Naafs, B.D.A., Qian, S., Dang, X., Sun, H., Pei, H., Wang, R., Zhao, S., Xie, S., 2021. Variations in dissolved O2 in a Chinese lake drive changes in microbial communities and impact sedimentary GDGT distributions. Chemical Geology 579, 120348. https://doi.org/10.1016/j.chemgeo.2021.120348

Wu, W., Ruan, J., Ding, S., Zhao, L., Xu, Y., Yang, H., Ding, W., Pei, Y., 2014. Source and distribution of glycerol dialkyl glycerol tetraethers along lower Yellow River-estuary–coast transect. Marine Chemistry 158, 17–26. https://doi.org/10.1016/j.marchem.2013.11.006

Xia, Y., Zhang, M., Tsang, D.C.W., Geng, N., Lu, D., Zhu, L., Igalavithana, A.D., Dissanayake, P.D., Rinklebe, J., Yang, X., Ok, Y.S., 2020. Recent advances in control technologies for non-point source pollution with nitrogen and phosphorous from agricultural runoff: current practices and future prospects. Applied Biological Chemistry 63, 8. https://doi.org/10.1186/s13765-020-0493-6

Xiao, X., Fahl, K., Müller, J., Stein, R., 2015. Sea-ice distribution in the modern Arctic Ocean: Biomarker records from trans-Arctic Ocean surface sediments. Geochimica et Cosmochimica Acta 155, 16–29. https://doi.org/10.1016/j.gca.2015.01.029

Xie, M., Chen, M., Wang, W.-X., 2018. Spatial and temporal variations of bulk and colloidal dissolved organic matter in a large anthropogenically perturbed estuary. Environmental Pollution 243, 1528–1538. https://doi.org/10.1016/j.envpol.2018.09.119

Xie, R., Qi, J., Shi, C., Zhang, P., Wu, R., Li, J., Waniek, J.J., 2023. Changes of dissolved organic matter following salinity invasion in different seasons in a nitrogen rich tidal reach. Science of The Total Environment 880, 163251. https://doi.org/10.1016/j.scitotenv.2023.163251

Xie, S., Liu, X.-L., Schubotz, F., Wakeham, S.G., Hinrichs, K.-U., 2014. Distribution of glycerol ether lipids in the oxygen minimum zone of the Eastern Tropical North Pacific Ocean. Organic Geochemistry 71, 60–71. https://doi.org/10.1016/j.orggeochem.2014.04.006

Xiong, J., Shen, J., 2022. Vertical Transport Timescale of Surface-Produced Particulate Material in the Chesapeake Bay. Journal of Geophysical Research: Oceans 127, e2021JC017592. https://doi.org/10.1029/2021JC017592

Xu, S., Zhang, Z., Jia, G., Yu, K., Lei, F., Zhu, X., 2020. Controlling factors and environmental significance of BIT and δ13C of sedimentary GDGTs from the Pearl River Estuary, China over recent decades. Estuarine, Coastal and Shelf Science 233, 106534. https://doi.org/10.1016/j.ecss.2019.106534

Yamashita, Y., Jaffé, R., 2008. Characterizing the Interactions between Trace Metals and Dissolved Organic Matter Using Excitation−Emission Matrix and Parallel Factor Analysis. Environ. Sci. Technol. 42, 7374–7379. https://doi.org/10.1021/es801357h

Yamashita, Y., Jaffé, R., Maie, N., Tanoue, E., 2008. Assessing the dynamics of dissolved organic matter (DOM) in coastal environments by excitation emission matrix fluorescence and parallel factor analysis (EEM-PARAFAC). Limnology and oceanography 53, 1900–1908.

Yamashita, Y., Panton, A., Mahaffey, C., Jaffé, R., 2011. Assessing the spatial and temporal variability of dissolved organic matter in Liverpool Bay using excitation–emission matrix fluorescence and parallel factor analysis. Ocean Dynamics 61, 569–579. https://doi.org/10.1007/s10236-010-0365-4

Yang, L., Guo, W., Hong, H., Wang, G., 2013. Non-conservative behaviors of chromophoric dissolved organic matter in a turbid estuary: Roles of multiple biogeochemical processes. Estuarine, Coastal and Shelf Science 133, 285–292. https://doi.org/10.1016/j.ecss.2013.09.007

Yang, L., Hur, J., Zhuang, W., 2015. Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: a review. Environmental Science and Pollution Research 22, 6500–6510.

Yedema, Y.W., Sangiorgi, F., Sluijs, A., Sinninghe Damsté, J.S., Peterse, F., 2023. The dispersal of fluvially discharged and marine, shelf-produced particulate organic matter in the northern Gulf of Mexico. Biogeosciences 20, 663–686. https://doi.org/10.5194/bg-20-663-2023

Yi, Y., Liu, T., Merder, J., He, C., Bao, H., Li, P., Li, S., Shi, Q., He, D., 2023. Unraveling the Linkages between Molecular Abundance and Stable Carbon Isotope Ratio in Dissolved

Organic Matter Using Machine Learning. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.3c00221

Zell, C., Kim, J.-H., Balsinha, M., Dorhout, D., Fernandes, C., Baas, M., Sinninghe Damsté, J.S., 2014. Transport of branched tetraether lipids from the Tagus River basin to the coastal ocean of the Portuguese margin: consequences for the interpretation of the MBT'/CBT paleothermometer. Biogeosciences 11, 5637–5655. https://doi.org/10.5194/bg-11-5637-2014

Zell, C., Kim, J.-H., Moreira-Turcq, P., Abril, G., Hopmans, E.C., Bonnet, M.-P., Sobrinho, R.L., Damsté, J.S.S., 2013. Disentangling the origins of branched tetraether lipids and crenarchaeol in the lower Amazon River: Implications for GDGT-based proxies. Limnology and Oceanography 58, 343–353. https://doi.org/10.4319/lo.2013.58.1.0343

Zhang, L., Xu, Y.J., Li, S., 2023. Source and quality of dissolved organic matter in streams are reflective to land use/land cover, climate seasonality and pCO2. Environmental Research 216, 114608. https://doi.org/10.1016/j.envres.2022.114608

Zhang, L., Zhu, X., Huang, X., Liu, C., Yang, Y., 2021. Effects of land use and nutrients on the characteristics of dissolved organic matter in the Nanchong Section of Jialing River, China in December 2019. Water Supply 22, 1863–1875. https://doi.org/10.2166/ws.2021.307

Zhang, X., Cao, F., Huang, Y., Tang, J., 2022. Variability of dissolved organic matter in two coastal wetlands along the Changjiang River Estuary: Responses to tidal cycles, seasons, and degradation processes. Science of The Total Environment 807, 150993. https://doi.org/10.1016/j.scitotenv.2021.150993

Zhang, Y., Liang, X., Wang, Z., Xu, L., 2015. A novel approach combining self-organizing map and parallel factor analysis for monitoring water quality of watersheds under non-point source pollution. Sci Rep 5, 16079. https://doi.org/10.1038/srep16079

Zhang, Z.-X., Li, J., Chen, Z., Sun, Z., Yang, H., Fu, M., Peng, X., 2020. The effect of methane seeps on the bacterial tetraether lipid distributions at the Okinawa Trough. Marine Chemistry 225, 103845. https://doi.org/10.1016/j.marchem.2020.103845

Zhao, C., Xu, X., Chen, H., Wang, F., Li, P., He, C., Shi, Q., Yi, Y., Li, X., Li, S., He, D., 2023. Exploring the Complexities of Dissolved Organic Matter Photochemistry from the Molecular Level by Using Machine Learning Approaches. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.3c00199

Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.-J., Zhang, H., 2021a. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. Environ. Sci. Technol. 55, 12741–12754. https://doi.org/10.1021/acs.est.1c01339

Zhong, S., Zhang, K., Wang, D., Zhang, H., 2021b. Shedding light on "Black Box" machine learning models for predicting the reactivity of HO radicals toward organic compounds. Chemical Engineering Journal 405, 126627. https://doi.org/10.1016/j.cej.2020.126627

Zhou, Y., He, D., He, C., Li, P., Fan, D., Wang, A., Zhang, K., Chen, B., Zhao, C., Wang, Y., Shi, Q., Sun, Y., 2021. Spatial changes in molecular composition of dissolved organic matter in the Yangtze River Estuary: Implications for the seaward transport of estuarine DOM. Science of The Total Environment 759, 143531. https://doi.org/10.1016/j.scitotenv.2020.143531

Zhou, Y., Li, Y., Yao, X., Ding, W., Zhang, Yibo, Jeppesen, E., Zhang, Yunlin, Podgorski, D.C., Chen, C., Ding, Y., Wu, H., Spencer, R.G.M., 2019. Response of chromophoric dissolved organic matter dynamics to tidal oscillations and anthropogenic disturbances in a large subtropical estuary. Science of The Total Environment 662, 769–778. https://doi.org/10.1016/j.scitotenv.2019.01.220

Zhou, Z., Guo, L., Shiller, A.M., Lohrenz, S.E., Asper, V.L., Osburn, C.L., 2013. Characterization of oil components from the Deepwater Horizon oil spill in the Gulf of Mexico using fluorescence EEM and PARAFAC techniques. Marine Chemistry 148, 10–21. https://doi.org/10.1016/j.marchem.2012.10.003

Zhu, C., Wagner, T., Pan, J.-M., Pancost, R.D., 2011. Multiple sources and extensive degradation of terrestrial sedimentary organic matter across an energetic, wide continental shelf. Geochemistry, Geophysics, Geosystems 12. https://doi.org/10.1029/2011GC003506

Zhu, W.-Z., Yang, G.-P., Zhang, H.-H., 2017. Photochemical behavior of dissolved and colloidal organic matter in estuarine and oceanic waters. Science of The Total Environment 607–608, 214–224. https://doi.org/10.1016/j.scitotenv.2017.06.163

Zhuang, W.-E., Chen, W., Cheng, Q., Yang, L., 2021. Assessing the priming effect of dissolved organic matter from typical sources using fluorescence EEMs-PARAFAC. Chemosphere 264, 128600. https://doi.org/10.1016/j.chemosphere.2020.128600

Zhuang, W.-E., Chen, W., Yang, L., 2023. Coupled effects of dam, hydrology, and estuarine filtering on dissolved organic carbon and optical properties in the reservoir-river-estuary continuum. Journal of Hydrology 617, 128893. https://doi.org/10.1016/j.jhydrol.2022.128893

Zhuang, W.-E., Chen, W., Yang, L., 2022. Effects of Photodegradation on the Optical Indices of Chromophoric Dissolved Organic Matter from Typical Sources. International Journal of Environmental Research and Public Health 19, 14268. https://doi.org/10.3390/ijerph192114268

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

# List of figures

# List of tables

**Dynamique de la matière organique dissoute et particulaire le long du continuum terre-mer de l'estuaire de la Seine (France)**

Résumé:

Les estuaires sont des zones critiques d'un point de vue écologique, économique et biogéochimique, qui jouent un rôle important dans la régulation de la Matière Organique Dissoute (MOD) et de la Matière Organique Particulaire (MOP). À ce jour, la MOD et la MOP ont le plus souvent été étudiées séparément dans les estuaires, ce qui freine la compréhension de leur dynamique globale et du fonctionnement biogéochimique estuarien associé. L'objectif de cette thèse était de déterminer les sources, les transformations et le devenir de la MOP et de la MOD dans l'estuaire de Seine (Nord-Ouest de la France). Dans ce but, les variations spatio-temporelles de leurs caractéristiques ont été étudiées à partir d'échantillons d'eau ($n$=383) collectés le long du continuum terre-mer de cet estuaire lors de 24 campagnes de prélèvements de juin 2019 à novembre 2022. Dans un premier temps, la dynamique de la MOP a été étudiée à l'échelle globale et moléculaire, via des analyses élémentaires et isotopiques et celles de biomarqueurs lipidiques. Un nouveau marqueur moléculaire (Riverine IndeX, RIX) basé sur des lipides membranaires d'origine bactérienne a ainsi été développé pour tracer les apports de MO provenant de la rivière en amont de l'estuaire. La MOP est un mélange de molécules d'origines variées (terrestre, algale/microbienne, anthropique), avec des dynamiques distinctes le long de l'estuaire de Seine. Cela montre que les transformations complexes auxquelles la MOP est soumise sont étroitement liées à sa composition. De plus, les propriétés de la MOD ont été étudiées en combinant les techniques optiques (absorbance UV-visible et matrice d'excitation-émission de fluorescence) à des algorithmes d'apprentissage automatique non supervisé et supervisé. Cela a conduit à l'élaboration d'un modèle démêlant la complexité de la MOD et révélant des caractéristiques spécifiques de cette dernière dans différentes zones de l'estuaire de Seine, avec des niveaux de masse moléculaire, d'aromaticité et de matière autochtone variables. De telles signatures de MOD ne sont pas identifiées efficacement en utilisant les approches traditionnelles. Enfin, ce travail de thèse a montré que les dynamiques de la MOD et de la MOP sont découplées le long du continuum terre-mer de Seine et que l'estuaire contrôle les processus de transport et transformation des constituants variés de la matière organique, qui sont liés, notamment, aux conditions hydroclimatiques et aux modes d'occupation des sols.

Mots-clés : Matière organique ; Estuaire de Seine ; biomarqueurs lipidiques ; spectroscopie de fluorescence ; dynamique ; algorithmes d'apprentissage automatique

**Dynamics of dissolved and particulate organic matter along the land-sea continuum of the Seine Estuary (France)**

Abstract:

Estuaries are critical zones from ecological, economical, and biogeochemical points of view and play an important role in regulating Dissolved Organic Matter (DOM) and Particulate Organic Matter (POM). To date, estuarine DOM and POM were mostly studied separately, hampering our understanding of their overall dynamics and associated estuarine biogeochemical functioning. The aim of this PhD thesis was to determine the sources, transformations and fate of estuarine POM and DOM in the Seine Estuary (North Western France). To this aim, the spatio-temporal variations of POM and DOM characteristics were determined using water samples ($n$=383) collected along the land-sea continuum of this estuary during 24 sampling campaigns from June 2019 to November 2022. First, the POM dynamics was investigated at the bulk and molecular scales, through elemental and isotopic analyses as well as lipid biomarkers. A novel molecular proxy (Riverine IndeX, RIX) based on bacterial membrane lipids was developed to trace the riverine POM inputs into estuaries. POM is a mixture of molecules from different sources (terrestrial, algal/microbial, anthropogenic), which showed distinct dynamics along the Seine estuary, highlighting that the complex transformations to which POM is subjected is closely dependent on its composition. Furthermore, DOM properties were investigated by optical techniques (UV–Visible absorbance and Excitation-Emission Matrix fluorescence spectroscopy) coupled with unsupervised and supervised machine learning. This led to the development of a model disentangling the complexity of DOM and capturing specific characteristics of the latter in different zones of the Seine Estuary, with varying levels of molecular weight, aromaticity, and autochthonous material. Such DOM signatures are not effectively identified using traditional approaches. Finally, this PhD work shows that DOM and POM dynamics are decoupled along the Seine land-sea continuum and that the estuary controls the transport and transformation of various constituents of organic matter, which are linked, in particular, to hydroclimatic conditions and land use.

Keywords: Organic matter; Seine Estuary; lipid biomarkers; fluorescence spectroscopy; dynamics; machine learning algorithms