Open water swimming : Monitoring and hydrodynamic modelling for sanitary conditions assessment

Natalia Angelotti de Ponte Rodrigues

École des Ponts ParisTech

THÈSE DE DOCTORAT
de l'École des Ponts ParisTech

Open water swimming: Monitoring and hydrodynamic modelling for sanitary conditions assessment

École doctorale Science Ingénierie et Environnement

Spécialité du doctorat : Sciences et Techniques de l'Environnement

Thèse préparée au sein du Laboratoire Eau Environnement et Systèmes Urbains, Ecole des Ponts ParisTech, et du Laboratoire d'Hydraulique Saint-Venant, EDF R&D et Ecole des Ponts ParisTech

Thèse soutenue le 20 décembre 2023, par
Natalia ANGELOTTI DE PONTE RODRIGUES

Composition du jury :

Céline CASENAVE Directrice de recherche, UMR MISTEA – INRAE	*Rapporteure et Présidente*
Jean-Baptiste BURNET Directeur de recherche, LIST (LU)	*Rapporteur*
Sarah DORNER Professeure, Polytechnique Montréal (CA)	*Examinatrice*
Paul KENNOUCHE Docteur, Ville de Paris	*Examinateur*
Brigitte VINÇON-LEITE Directrice de recherche, LEESU – ENPC	*Directrice de thèse*
Rémi CARMIGNIANI Chargé de recherche, LHSV – EDF R&D, ENPC	*Co-Directeur de thèse*

Abstract

Swimming in open water has become very popular, particularly in urban areas where summer heatwaves are increasingly frequent. To control sanitary risks, the monitoring of indicators of pathogenic microorganisms, namely faecal indicator bacteria (FIB), such as *E. coli* is required. Reference laboratory methods take at least 24h to provide results.

For opening a bathing area to the public, FIB concentrations must meet the regulatory limits. In watercourses, this condition can be estimated from measurements at an upstream point, defined to take into account the transfer time to the bathing area and the response time of the FIB measurement.

The transfer time is affected by flowrate and thermal stratification. Variations in current velocity can be computed using hydrodynamic models, which are also useful to estimate the FIB spatiotemporal distribution

In order to propose a system for monitoring and forecasting the microbiological quality of bathing areas, two research lines have been developed: (1) the implementation of a 3D hydrodynamic model to compute the velocity field and the water temperature and (2) the definition of a FIB proxy for faster microbiological water quality assessment.

The specific objectives of the PhD thesis are as follows:

i. To implement the 3D hydrodynamic model TELEMAC-3D, coupled with the WAQTEL thermal module;
ii. To design and install a monitoring system to obtain field data that can be used for input and validation of the hydrodynamic model;
iii. To investigate the influence of thermal stratification on the spatiotemporal distribution of faecal contamination;
iv. To identify a fast-measuring indicator of FIB.

The study site, La Villette basin, is located in Paris (France). It consists of a canal upstream (25m wide, 800m long), and a basin downstream (70m wide, 700m long, 3m deep). The average discharge is around $2.1 m^3.s^{-1}$.

The installed fully autonomous monitoring system consists of two monitoring points located: (1) 1000 m upstream of the bathing area and (2) downstream, close to the bathing area. At both sites, water temperature is measured at three depths, and electrical conductivity at mid-depth, with a 10 min time step.

From spring to mid-autumn, the data showed a daily thermal stratification of the water column, breaking up at night. Conductivity data allowed to estimate the transfer time between both monitoring points. The values varied from 6h30 to more than 24h.

To obtain real-time information on FIB level, the relationship between FIB and the fluorescence of dissolved organic matter (FDOM) was investigated. Fluorescence spectroscopy combined with PARAFAC modelling was used to characterize DOM and microbiological activity in five watercourses in the Paris metropolitan region. The results showed a higher presence of a protein-like component in samples with higher microbial contamination. A relationship was established between this component and FIB.

For a short-term (3 days) FIB level forecast, a relationship between FIB and hydro-meteorological data (discharge, rainfall and dry weather duration) was investigated. In La

Villette basin, a relationship between *E. coli* and the precipitation measured upstream, between one and three days before the day of interest, was found.

The *in-situ* measured values were used as model input and for validation. The FIB concentration was derived from FDOM data and from hydro-meteorological data using the established relationships.

The results of the 3D hydro-microbiological model were compared to field data. The observed alternation of thermal stratification and mixing was correctly reproduced. Its effect on *E. coli* spatial distribution was highlighted. The conductivity variation was well simulated. The transfer time between the two measuring points was correctly estimated and can be applied to other passive tracers.

This work highlights the relevance of a 3D hydrodynamic model for assessing contaminant transport and sanitary conditions in a river area of interest.

Keywords: *Urban waters, Early warning system, Faecal indicator bacteria (FIB), 3D hydrodynamic modelling, Thermal stratification, Fluorescent Dissolved Organic Matter (FDOM).*

Résumé

La baignade en eau libre connait un fort engouement, notamment en milieu urbain où les canicules estivales sont de plus en plus fréquentes.

Pour éviter les risques sanitaires, le suivi d'indicateurs de micro-organismes pathogènes, les bactéries indicatrices fécales (BIF) telles que *E. coli*, est nécessaire. Les résultats des méthodes de référence en laboratoire sont obtenus au mieux en 24h.

Pour qu'une baignade soit ouverte au public, les concentrations de BIF doivent être inférieures à des seuils réglementaires. En cours d'eau, ces conditions peuvent être estimées à partir de mesures en un point amont, défini en tenant compte du temps de transfert vers la baignade et du temps de réponse de la mesure des BIF.

Le temps de transfert change en fonction du débit et de la stratification thermique. Ces variations peuvent être calculées par des modèles hydrodynamiques, également utiles pour estimer la distribution spatiotemporelle des BIF.

Afin de proposer un système de suivi et de prévision de la qualité microbiologique d'une zone de baignade, deux axes de recherche ont été développés : (1) la mise en œuvre d'un modèle hydrodynamique 3D pour calculer les champs de vitesses et la température de l'eau et (2) la définition d'un proxy des BIF pour une évaluation rapide de la qualité microbiologique de l'eau.

Les objectifs spécifiques du doctorat sont les suivants :

i. Mise en œuvre du modèle hydrodynamique, TELEMAC-3D, couplé au module thermique WAQTEL ;
ii. Conception et déploiement d'un système de mesure de terrain ;
iii. Étude de l'influence de la stratification thermique sur la distribution spatiotemporelle des BIF ;
iv. Identification d'un indicateur des BIF de mesure rapide.

Le site d'étude, le bassin de La Villette, est situé à Paris (France). Il se compose d'un canal en amont (25 m de large, 800 m de long) et d'un bassin en aval (70m de large, 700m de long, 3m de profondeur). Le débit moyen est d'environ $2.1 m^3.s^{-1}$.

Le système de suivi se compose de deux points de mesure situés à 1000m en amont de la zone de baignade et en aval, à proximité de la baignade. Sur les deux sites, la température de l'eau est mesurée à trois profondeurs et la conductivité électrique à mi-profondeur (pas de temps 10 min).

Du printemps à l'automne, une stratification thermique de la colonne d'eau, interrompue la nuit a été mesurée. Les données de conductivité aux deux points de suivi ont permis d'estimer le temps de transfert d'un traceur passif, qui varie de 6h30 à 24h.

Pour connaitre leur niveau en temps réel, la relation entre BIF et fluorescence de la matière organique dissoute (MOD) a été étudiée. La spectroscopie de fluorescence combinée à la modélisation PARAFAC a été utilisée pour caractériser la MOD et l'activité microbiologique dans cinq sites en région parisienne. Les résultats ont montré une forte présence d'un composant protéique dans les échantillons de contamination microbienne plus élevée. Une relation a été établie entre ce composant et les BIF.

Pour une prévision à court terme (3 jours), une relation entre BIF et données hydrométéorologiques (débit, pluie et durée de temps sec) a été étudiée. À La Villette, une

relation entre *E. coli* et la pluie mesurée en amont, entre un et trois jours avant le jour d'intérêt, a été trouvée.

Les mesures ont été utilisées comme données d'entrée du modèle et pour sa validation, la concentration des BIF étant estimée à partir des relations établies entre fluorescence et données de pluie.

Les résultats du modèle hydro-microbiologique ont été comparés aux mesures. L'alternance de stratification thermique et mélange observée a été correctement reproduite. Son effet sur la distribution spatiale d'*E. coli* a été mis en évidence. La variation de la conductivité a été bien simulée. Les temps de transfert, correctement estimés, peuvent être appliqués à d'autres traceurs passifs.

Ce travail met en évidence la pertinence d'un modèle hydrodynamique 3D pour évaluer le transport des contaminants et les conditions sanitaires dans un cours d'eau.

Mots-clés : *Eaux urbaines, Système d'alerte précoce, Bactéries indicatrices fécales (BIF), Modélisation hydrodynamique 3D, Stratification thermique, Matière organique dissoute fluorescente.*

Acknowledgements

First of all, I would like to express my deepest gratitude to my thesis supervisor Brigitte Vinçon-Leite and co-supervisor Rémi Carmigniani, for their invaluable guidance, encouragement, and support throughout this journey. Their scientific advice, patience, guidance, and constructive feedback have been essential for the accomplishment of this work.

I extend my sincere gratitude to the members of my thesis committee, Sylvie Barraud, Laurent Moulin, and Minh-Hoang Le, for their insightful comments, suggestions, and advice, which strengthened the solidity of this work.

I would also like to thank Céline Casenave and Jean-Baptiste Burnet for their diligent and insightful review of this dissertation. Additionally, I express my gratefulness to Sarah Dorner and Paul Kennouche for their contributions as examiners, which have further strengthened the quality of this thesis. Engaging in discussions about my work with them four has been both enriching and encouraging.

I would like to express my appreciation to École des Ponts ParisTech and to OPUR project for funding my doctoral scholarship and providing all the resources to achieve the objectives of this work. Furthermore, I extend my gratitude to École des Ponts ParisTech for providing the academic and professional resources that were essential for the conduction of my scientific research.

I wish to thank *Ville de Paris* (*Service des Canaux* and *Service Technique de l'Eau et de l'Assainissement*), *Eau de Paris*, *Marin d'Eau Douce* and the *Syndicat Marne Vive* for their assistance and logistical support, especially on the obtention of field data.

All my gratitude to the teams at LEESU and LHSV, that warmly welcomed me and shared their solid scientific knowledge throughout these years. Thanks to Annick and Catherine for always being so helpful and being amazing paperwork problem-solvers, and to Louise, for being so supportive and for always being a positive and pleasant presence in our shared office.

Special thanks to Mohamed and Philippe for their willingness to share knowledge, their vital contributions, expertise and technical assistance in the field campaigns, laboratory analysis and equipment installation and maintenance, all essential to the success of this research. I also wish to thank the LEESU interns, part of the Brazilian crew, for their active and precious contribution to the progress of this work.

I am indebted to my colleagues and peers for their camaraderie, support, and stimulating debates, which have been invaluable in shaping my ideas and perspectives. To all my fellow doctoral and postdoctoral colleagues, my deepest gratitude for sharing all the coffees, happy hours, sport activities and philosophical conversations.

I would like to express heartfelt gratitude to Romain, my husband, who has always been a great support, providing the energy I needed to pursue this journey and always bringing light to my life. To my parents, Irene and Antonio, and to my brother Lucas, I am beyond grateful for their unwavering love and encouragement throughout this challenging endeavour. Their belief in me has been a constant source of motivation.

Lastly, I would like to express my deepest gratitude to all those who have supported me in ways seen and unseen. Your contributions, no matter how small, have made a significant difference in my journey towards completing this thesis.

Contents

List of Figures

List of Tables

Chapter 1. Introduction

Reintroducing urban bathing is part of environmental policies aimed at recovering the ecological quality of the river in wider Paris. Urban bathing mitigates the urban heat island effects of Paris and reduces the heat wave impact on the population. Furthermore, it is a way of reclaiming public space and reinforcing the links of the city dwellers with nature.

Open water sport events are increasingly popular in urban areas. Some open water competitions of the International Federation of Swimming are performed in urban freshwater bodies (FINA https://www.fina.org/open-water). For example, in July 2022, the second phase of Marathon Swim World Series took place in Paris, at the Ourcq Canal.

However, open water swimming can expose bathers to waterborne diseases. The prefectural decree of 1923, banning public swimming in the Seine River, is still in force today. An exception in Paris region, since 2017, is the public swimming area installed in La Villette basin during summer.

To control the sanitary risks, the European Directive of Bathing Waters 2006/7/EC (EU, 2006) classifies the water quality in function of the concentrations of indicators of faecal contamination known as faecal indicator bacteria (FIB). Two bacteria are used as reference parameters in the European classification: *Escherichia coli* (*E. coli*) and intestinal enterococci (IE). For inland waters, the threshold for a sufficient water quality for bathing is 900 MPN.100mL^{-1} for *E. coli* and 330 MPN.100mL^{-1} for IE, both values based upon a 90-percentile evaluation. The reference method for *E. coli* enumeration, in the laboratory, has a long response time, taking from 24h to 48h (ISO, 1998).

In some watercourses, the analysis of the microbiological water quality is performed on a daily basis, upstream of the bathing area, in order to anticipate the decision of a closure of the bathing area. However, hydro-meteorological conditions may impact the transfer time, such as flow rate variations and thermal stratification. Vertical thermal stratification causes differences in the flow velocity with depth, impacting the spatiotemporal distribution of FIB. Therefore, the arrival time and the duration of the contamination can vary with depth in the bathing area.

To improve the accuracy of the decision making and reduce the risk of having undetected contamination between two measurements, a more frequent monitoring time step is required. An indicator of faecal contamination available in real-time would be very beneficial for decision making about a possible closure or re-opening of the bathing area.

In this PhD thesis, two research lines have been developed in order to propose a system for monitoring and forecasting the microbiological quality of bathing areas: (1) the implementation of a 3D hydrodynamic model to compute the velocity field and the water temperature and (2) the definition of a FIB proxy for faster microbiological water quality assessment.

The specific objectives are as follows:

- To implement the 3D hydrodynamic model TELEMAC-3D, coupled with the WAQTEL thermal module;
- To design and install a monitoring system to obtain field data that can be used for input and validation of the hydrodynamic model;
- To investigate the influence of thermal stratification on the spatiotemporal distribution of faecal contamination;

- To identify a fast-measuring indicator of FIB.

This manuscript is divided into five chapters. The bibliography review (chapter 2) presents successively, the faecal contamination in inland waters, the water quality variables for the assessment of microbial contamination. Then, it presents the importance of autonomous monitoring system and describes the main characteristics of hydrodynamic and microbiological modelling.

Chapter 3 describes the study site, the design of the monitoring system, the water sampling protocol, the data processing, including PARAFAC modelling and then the hydrodynamic modelling.

Chapter 4 is devoted to result presentation. It includes four result sections: (1) the field collected dataset; (2) the dissolved organic matter characteristics; (3) the estimation of faecal indicator bacteria; (4) the hydro-microbiological modelling outcomes.

Finally, chapter 5 presents the conclusion and perspectives of this work.

Chapter 2. Bibliography review

This chapter describes the scientific background of the thesis subject and intends to update the recent advances in the literature. It is divided into three main topics: faecal contamination quantification, water quality variables related to microbial contamination and hydro-microbiological modelling.

2.1. Faecal contamination in inland water courses

2.1.1. Faecal indicator bacteria (FIB)

Polluted rivers can contain various types of harmful microorganisms such as viruses, bacteria, and protozoa. These micro-organisms primarily come from the waste of humans and warm-blooded animals that ends up in water. The risk to human health depends on how the water is used (drinking, recreational activities, bathing, etc.) and the concentration of pathogens in the water.

Waterborne pathogens can cause illness. These pathogens can cause outbreaks of gastrointestinal diseases like diarrhoea and dysentery, other diseases of the skin, eyes, respiratory tract, and even typhoid fever and cholera (Servais et al., 2007; US EPA, 2013a).

Detecting and counting all the potentially present pathogenic microorganisms in aquatic systems is challenging due to the wide range of pathogens, their low abundance, and the viability of detection methods. Therefore, routine monitoring of the microbiological quality of water still relies on the use of bacterial indicators found in faecal matter. These indicators should be abundant in human and animal waste, easily detectable, and not able to grow in natural waters.

The prediction of faecal bacteria in waterbodies is of significance importance. Faecal indicator bacteria (FIB) have been used as indicators of potential waterborne pathogens. A high level of FIB implies the presence of human or animal waste, which may be a source of viruses, protozoa and bacteria that are harmful for human beings.

The sources of faecal pollution on the environment can be from wastewater and sludge (human origin), slurry and manure (animal origin) and wildlife. This includes wastewater treatment plant effluent, sewage discharge, stormwater runoff, improper land application of manure or sewage (US EPA, 2013a).

The FIB are defined as of exclusive faecal origin, universally present in large numbers in faeces of warm-blooded animals and humans, capable of surviving in but do not grow in natural waters, and are readily detectable by simple methods (Environment Agency, 2002; Paruch and Mæhlum, 2012).

The commonly used bacteria as water quality criteria are intestinal enterococci (IE) and *Escherichia coli* (*E. coli*) (EU, 2006). Both commonly colonise the gastrointestinal tracts of warm-blooded animals, including humans, and are mainly present in the environment because of faecal contamination. Even if these bacteria are not necessarily pathogenic, several epidemiological studies support their use as indicators of a pathogenic source (Pruss, 1998).

The Bathing Water Directive 2006/7/EC of 2006 classifies the inland bathing water quality according to faecal bacteria indicators intestinal enterococci and *Escherichia coli* (Table 2.1). This classification is based on data collected over a four-year period. The World Health Organization (WHO) recommends a minimum of 20 annual samples per site, totalizing at least 80 samples over the classification period (WHO, 2018).

Table 2.1. Classification of Directive 2006/7/EC of 2006 for bathing water quality

Parameter	Excellent quality	Good quality	Sufficient	Reference methods of analysis
Intestinal enterococci (CFU/100 mL)	200 (*)	400 (*)	330 (**)	ISO 7899-1 or ISO 7899-2
***Escherichia coli* (CFU/100 mL)**	500 (*)	1000 (*)	900 (**)	ISO 9308-3 or ISO 9308-1

(*) Based upon a 95-percentile evaluation.
(**) Based upon a 90-percentile evaluation.
Source: (EU, 2006).

Both FIB are historically recommended for freshwater microbiological quality assessment (USEPA, 1986). Part of coliform bacteria group is resistant to high temperatures. This sub-group is called thermotolerant coliform bacteria and, as high temperature environments are supposed to suppress bacteria of non-faecal origin, it is also known as faecal coliforms (WHO, 1996). The main bacterium of this group is the *E. coli* and it is the only one that has an exclusively faecal origin and does not multiply appreciably in natural environment (Paruch and Mæhlum, 2012).

In freshwaters, *E. coli* is more frequently utilized (Harmel et al., 2016). In faecal contaminated waters, *E. coli* are found in greater concentrations than intestinal enterococci in all mammalian faeces and is the principal driver for fresh water quality classification (Jacob et al., 2015; WHO, 2018). Edberg et al. (2000) recommended to use *E. coli* as the best indicator, with IE as a complementary one.

Therefore, the detection of *E. coli* and IE contamination is important to minimise eventual health risks due to faecal pollution exposure and protect human health (Paruch and Mæhlum, 2012; WHO, 1996).

2.1.2. *E. coli* quantification methods

There are several ways of measuring *E. coli* concentration. The main technologies of quantification are the detection by enzymatic activity (or "defined substrate technology"), the measurement of total concentrations by optical methods, or a combination of both.

Different techniques can be used either in the laboratory or directly *in situ*. Historically, the enumeration techniques used in laboratory are culture-based, which consists of cultivating individual target organisms to have an amplification of the signal produced by the total population. This method requires time for the cultivation, needing more than 18 hours between sample collection and reporting results (ColiMinder VWMS, 2021; Lopez-Roldan et al., 2013).

The regulatory standard methods are culture-based. For example, NF EN ISO 9308 parts 2 (ISO, 2012) and 3 (ISO, 1998) are culture techniques based on the statistic method of the Most Probable Number (MPN), also called "multi fermentation tube technique".

The quantification of *E. coli* can also be performed by PCR-based techniques. They are fast, have high specificity, and are effective from surface water to wastewater. However, they

need standardization and specialized laboratory equipment, which limits their implementation for a rapid microbial water quality testing (Burnet et al., 2019). Also, they are easily affected by inhibitors or other interferents, and are difficult to integrate in an autonomous device (Lopez-Roldan et al., 2013).

Many of the *in situ* techniques are based either on the fluorescence of bacterial enzymatic activity or on optical principles. The latter are fast and reagentless, but they have low specificity (Lopez-Roldan et al., 2013).

The measured concentration of each technique can be expressed differently. The results of the standard methods are expressed as Most Probable Number per 100 mL ($MPN.100mL^{-1}$) or Colony Forming Unit per 100 mL ($CFU.100mL^{-1}$), that are numerically equivalent.

Other techniques have different units. Optical techniques can be expressed in particles or cells per 100 mL (Tatari et al., 2016). Specific enzymatic activity based techniques can be expressed on Modified Fishman Unit per 100 mL ($MFU.100mL^{-1}$), a unit that represents a direct measurement of the number of living organisms per sample volume (Koschelnik et al., 2015).

The following paragraphs describes in more details the methods of interest in this study. A summary table is presented afterwards (Table 2.2).

2.1.2.1. Laboratory analysis: multi fermentation tube technique

The multi-fermentation tube technique is a culture-based method. It uses aliquots of samples from three consecutive 10-fold dilutions and a culture media for each bacteria indicator (Ballance and Bartram, 1996).

This technique is based on Most Probable Number (MPN) methods, as well as other analyse methods of laboratory. According to the water sample source, a number of dilutions is done and distributed on specific support wells. These wells can be tubes or 96 well microplates, for example (Hamilton et al., 2005). The typical Time-To-Result (TTR) of current laboratory techniques are of 18 to 72 hours, without considering the time taken during sampling, transportation, handling and result interpretation (Angelescu et al., 2019).

The Most Probable Number (MPN) method applies the principle of the maximum likelihood to write a function that describes the probability of a bacterial density associated with a defined dilution degree "k" (Maul, 1982). This is the method used to create the reference MPN table and it is the base of laboratory analyse methods (Hamilton et al., 2005).

The bias of the real density estimation depends on the number of the aliquots used at each dilution. It becomes neglectable over 96 tubes, and this is why the commonly used on the certified methods has 96 aliquots or more (Maul, 1982).

The MPN based methods used on the legislation use a fluorescent media. Undiluted and/or diluted samples are added on micro wells or containers and are incubated for 24 h at 35 °C (NF EN ISO 9308-2) or for 36 to 48 h at 44 °C (NF EN ISO 9308-3). If in contact with faecal indicator bacteria, the media reacts and, under ultraviolet light, wells that both turned yellow and fluoresced are counted as *E. coli* positive (Hamilton et al., 2005; Prats et al., 2008; Tiwari et al., 2016).

2.1.2.1.1. Defined substrate technology with enzymatic fluorescence techniques

The multi fermentation technique can use a specific media culture with defined substrate technology. The introduction of chromogenic and fluorogenic media has been facilitating the recognition of colonies of target organisms.

The microbiological growth media contains enzyme substrates linked to a chromogen (colour reaction), fluorogen (fluorescent reaction) or a combination of both. The target population is characterized by enzyme systems that metabolize the substrate to release the chromogen/fluorogen. This results in a colour change in the medium and/or fluorescence under long-wave UV light (Prats et al., 2008).

Most devices using this technology are based on IDEXX Colilert, which is used for the simultaneous detection and enumeration of total coliforms and *E. coli* in water and wastewater based on the Most Probable Number (MPN) principle. This method uses two chromogenic nutrient indicators [i.e., ortho-nitrophenyl-β-D-galactopyranoside (ONPG) and 4-methylumbelliferyl-β-D-glucuronide (MUG)] as the major sources of carbon.

As coliforms grow, they use β-galactosidase (GLUC) to metabolize ONPG and change it from colourless to yellow. *E. coli* uses β-glucuronidase to metabolize MUG and create fluorescence by releasing of the fluorescent compound 4-methyl-umbelliferone (MUF) (Burnet et al., 2019; Hamilton et al., 2005) (Figure 2.1).

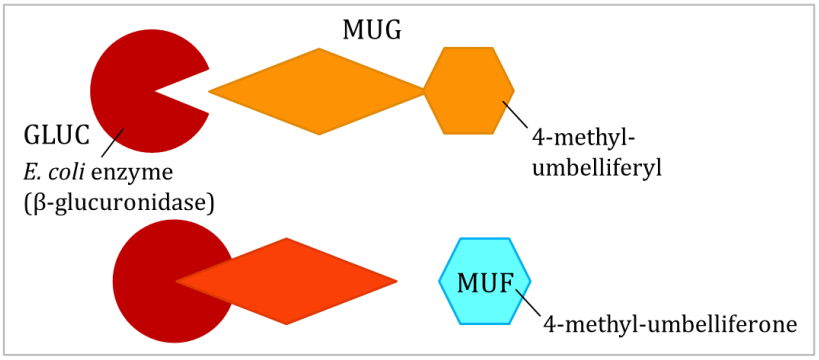


Figure 2.1. Scheme of defined substrate technology: enzymatic fluorogenic reaction (based on "Quanti-Tray System - IDEXX US," 2020)

2.1.2.2. *In situ* measurements

In situ devices are automatized samplers that provide *E. coli* concentration at the measurement point.

2.1.2.2.1. Fluidion ALERT system

- **Measure principle**

This system is based on the defined substrate technology with enzymatic fluorescence technique, presented on the precedent item. Each bioreactor has the substrate that provides enzymatic fluorogenic reaction with *E. coli*.

The fluorescence signal is monitored over time. The appearance time of fluorescence is measured and correlated with concentrations of *E. coli* in Colony-Forming Units per 100 mL (CFU.100mL^{-1}, numerically equivalent to MPN.100mL^{-1}). This appearance time ranges from **2 h to 12 h**. After calibration, for a concentration of 1000 MPN.100mL^{-1}, it takes around 9 h to have the results.

Complementary information is presented on Annex I.

- **Instrumentation description**

Each bioreactor allows to sample and make microbiological analysis individually, including the mixture with the bioreactive, incubation in a controlled temperature of 37-38 °C, monitoring of the optical characteristics of the sample (fluorescence) and the wireless transmission of the measures by a mobile phone network.

The equipment works with a battery, being fully autonomous and waterproof. The initiation of the sampling command can be established through a connection with an external sensor, allowing an automatic triggering of sampling and measurements operations by a mobile phone network, or through a USB cable, to download the data.

After the seven samplings are concluded, a technical maintenance is required to replace them. A technician is the responsible for changing of the vials, adding fresh reagent, decontaminating the sample tubing, and replacing the battery. This procedure takes less than 30 min.

- **Range and confidence interval**

According to the product data sheet, the range of measure is 4 to $5 \cdot 10^5$ CFU.100mL^{-1} (Colony Formation Units). Angelescu et al. (2019) concluded that, after a comparison with a reference method, the ALERT system instruments provide equivalent quantification accuracy to the reference microplate method over the range of measure after calibration.

2.1.2.2.2. ColiMinder

ColiMinder® (ColiMinder VWM Solution GmbH, Austria) is a *in situ E. coli* measurement device. It is based on the measurement of the enzymatic activity of the metabolism of the target microorganisms (ColiMinder VWMS, 2021).

While standard lab evaluation measures use proliferation to make bacteria visible and count the cell colonies, this device directly measures a specific metabolic activity via the fluorescence emitted by a product of the enzymatic reaction (Figure 2.2).

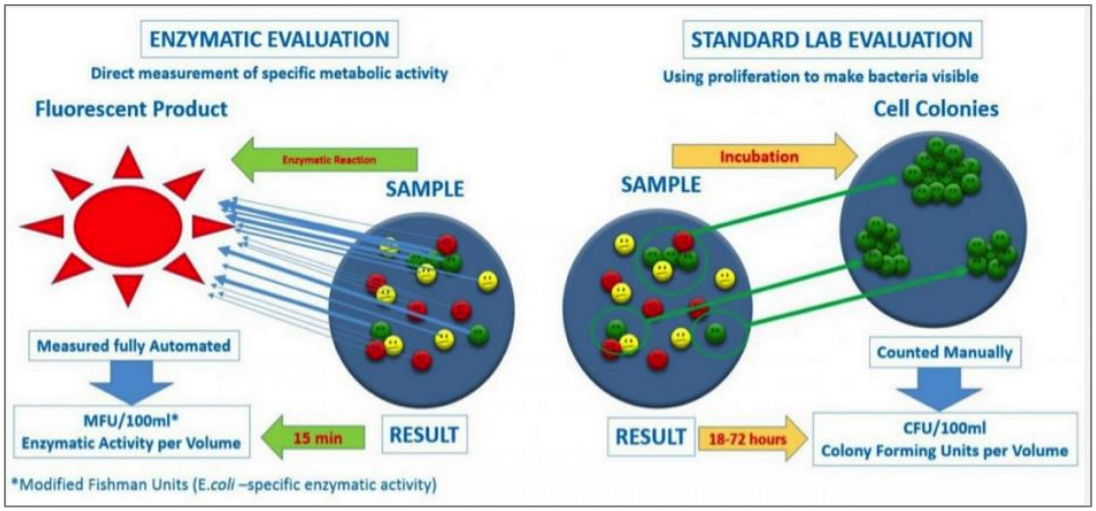


Figure 2.2. ColiMinder enzymatic evaluation and the standard lab evaluation of *E. coli* quantification (ColiMinder VWMS, 2021)

- **Measure principle**

ColiMinder measurement is based on a defined substrate technology. The activity of *E. coli* is associated to its enzyme known as GLUC (β-D-glucuronidase). This enzyme catalyses the MUG (4-methylumbelliferyl-β-D-glucuronide) substrate, forming a highly fluorescent reaction product called MUF (4-methylumbelliferyl) (Burnet et al., 2019; Koschelnik et al., 2015).

The fluorescence measurement of the *E. coli* specific enzymatic activity is expressed in Modified Fishman Units per 100 mL of water sample (MFU.100mL^{-1}).

- **Range and confidence interval**

The measurement range is from 1 to 1500 mMFU.100mL^{-1} (Koschelnik et al., 2015), with a precision of 0.8 mMFU.100mL^{-1} (Tatari et al., 2016).

However, there is no fixed conversion factor between MFU.100mL^{-1} and CFU.100mL^{-1}, with the factor being between 0:1 and 1:0 (Lackner et al., 2017).

Table 2.2. Synthesis of the presented *E. coli* quantification methods

Method	Measurement	Measure principle	Response time	Incubation temperature	*E. coli* unit
Multi tube technique	Manual	Substrate hydrolysis (MUG) with fluorescent analysis	24h	35 °C	MPN.100mL^{-1}
Fluidion	Automatic	Substrate hydrolysis (MUG) with fluorescence continuous measurements	2-12h (variable accordingly to concentration)	37 °C	CFU.100mL^{-1} (numerically equivalent to MPN.100mL^{-1})
ColiMinder	Automatic	Measures the GLUC activity (*E. coli* enzyme) with MUG substrate based on optical detection of MUF	30-60 min	44 °C	MFU.100mL^{-1}

2.2. Water quality variables for the assessment of microbial contamination

This section presents the water quality variables that should help to identify and track microbial contamination.

2.2.1. Fluorescent dissolved organic matter in natural waters

2.2.1.1. Dissolved organic matter

The dissolved organic matter (DOM) corresponds to any organic compound that passes through a 0.45 µm filter (Evans et al., 2005). It is a complex mixture of soluble organic compounds of various origins.

In freshwaters, dissolved organic matter (DOM) plays a key role in the ecosystem function. Part of DOM pool is associated to bacterial degradation and nutrient availability (Boavida and Wetzel, 1998; Wetzel, 1992), acting as a supplier of biologically accessible organic substances that sustains heterotrophic aquatic organisms. DOM may also attenuate light intensity in the water (Markager and Vincent, 2000; Morris et al., 1995), mobilize and

transport contaminants (Mladenov et al., 2010; Polubesova et al., 2007), and affect the concentration of dissolved oxygen in the water (McCabe et al., 2021). In addition, DOM can affect water treatment processes in drinking water supply plants (Wells et al., 2022).

Change in DOM composition may reflect the variation of the status of an ecosystem caused by land use and meteorological change (Williams et al., 2010; Wilson and Xenopoulos, 2009).

In urban waterbodies, during rain events, wastewater discharge and runoff affect the water quality and change the DOM composition. Humic-like components are associated with terrestrial organic matter such as run-off water, and protein-like components are associated with a high level of microbial activity, such as wastewater. Meng et al. (2013) concluded that anthropogenic inputs to urbanized surface waters are more likely to change seasonally and spatially. These inputs are indicated by organics compounds present on wastewater, such as humic-like compounds, and tryptophan-like and tyrosine-like proteins.

2.2.1.2. Fluorescence of dissolved organic matter

The fluorescent dissolved organic matter (FDOM) of a water sample can be characterized by a technique based on the simultaneous scanning of excitation and emission wavelengths. The sample is excited by a delivered wavelength. This excitation induces fluorescence that is emitted and which intensity is measured for a range of wavelengths (Frank et al., 2018a; Lawaetz and Stedmon, 2009). An excitation-emission matrix (EEM) is then obtained (Figure 2.3).

Fluorescence spectrophotometry has been widely used for characterizing DOM in surface waters (Cumberland et al., 2012a), ground waters (Frank et al., 2018b), marine waters (Parlanti et al., 2000), and for microbial quality assessment (Baker et al., 2015a).

The use of fluorescence spectrometry to identify microbiological contamination has been studied for several years (Bedell et al., 2020; Cumberland et al., 2012b; Frank et al., 2018a; Nowicki et al., 2019; Sorensen et al., 2018, 2015; Ward et al., 2021, 2020).

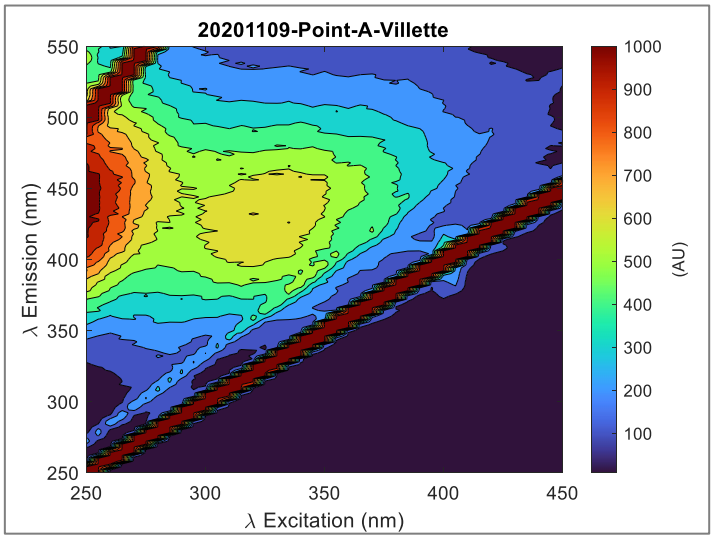


Figure 2.3. Typical fluorescence excitation-emission matrix (EEM) of La Villette canal (example from sample at Point A, upstream, 09/11/2020).

2.2.1.2.1. Fluorophores

Fluorescence spectroscopy can be used to assess DOM composition, DOM source, type, and microbiological activity (Coble, 1996). Fluorescent dissolved organic matter (FDOM) contains molecular groups called fluorophores. A fluorophore is a chemical compound that

emits fluorescence when excited by a light at specific wavelengths. Some compounds have a characteristic couple of excitation-emission wavelengths (Fellman et al. 2010) , such as the tryptophan.

The fluorophores identified in freshwaters are associated with different types of dissolved organic matter and can be mostly attributed to components with humic-like fluorescence (HLF), such as peaks A, C and M, and tryptophan-like fluorescence (TLF), like peak T. Peak A is associated with terrestrial components. Peak C is often linked to anthropogenic activities such as agriculture. Peak M is associated with marine humic-like components. Peak T may indicate microbial activity (Coble, 1996).

HLF intensity can be mainly associated to two peaks A and C. Peak A is the maximum at ex 260 nm and em 400-460 nm. Peak C is the maximum intensity in the intervals ex 320-360 nm and em 420-460 nm (Coble, 1996).

TLF intensity, hereafter peak T, is identified at an excitation (ex) wavelength range of 270-280 nm and emission wavelength (em) range of 300-360 nm (Frank et al., 2018a; Sorensen et al., 2018; Ward et al., 2021, 2020), being ex 275 nm and em 340 nm the most frequent on the studied literature (Coble, 1996).

Peak T is a protein-like component. It is described as the microbial contamination indicator related to faecal bacteria. It consists of the fluorescence peak of the aromatic amino acid tryptophan, often associated to microbial activity (Cumberland et al., 2012b; Sorensen et al., 2018; Ward et al., 2021, 2020).

Peak T measurements can be used as a risk assessment tool. Ward et al. (2021) mentioned that elevated TLF concentrations indicates favourable conditions for the persistence of bacteria like coliforms and other potential pathogens.

Studies have been showing strong correlations between peak T and *E. coli* and total coliforms (Baker et al., 2015b; Bedell et al., 2020; Cumberland et al., 2012b; Nowicki et al., 2019; Sorensen et al., 2018, 2015). According to Bedell et al. (2020), this is due because, compared to other bacteria, *E. coli* have the highest TLF per occurrence. Furthermore, they found that peak T was able to classify sources with *E. coli* concentrations >10 CFU.100mL^{-1}.

Among the studies, an uncertainty could be associated to the study site. The matrix composition of the sampling site may change among watercourses from different catchment areas, due to the complexity and heterogeneity of natural environments. In addition, multiple compounds can emit peak T fluorescence, increasing the uncertainty of peak T as a proxy of bacterial concentrations (Baker et al., 2007; Bedell et al., 2022; Sorensen et al., 2016).

When screening for faecal contamination, Ward et al. (2021, 2020) recommended that a potential influence of humic-like fluorescence (HLF) should be considered. They concluded that HLF may mask TLF at low concentrations. Therefore, the use of peak T is more suitable on the assessment of moderate and high level of faecal contamination.

Fluorescence measurement may be altered by temperature and pH. For a temperature variation of ±8 °C and between pH values 5 and 8, fluorescence intensities are not affected. For pH values <4.5 fluorescence intensities are quenched by 15% and at values >8.5 they are enhanced by 30% (Reynolds, 2003).

The temperature variation is particularly important for *in situ* measurements, where major variation of water temperature may occur. On the laboratory, a room with stable temperature (around 20 °C), the change in TLF signal is negligible (Nowicki et al., 2019).

Despite these drawbacks (potential shifts in fluorescence intensities with temperature, oxygen concentration or extremes of pH) (Fellman et al., 2010), fluorescence components still are a valuable alternative to be used as a rapid reagentless indicator of faecal contamination. Several studies have been made with encouraging results to use the presented fluorescence components, particularly protein-like fluorescence components like peak T, as a tool for the assessment of microbial water quality and correlate them with faecal bacteria (Baker et al., 2019, 2015b; Chen et al., 2017; Cumberland et al., 2012b; Frank et al., 2018a; Hudson et al., 2007; Sorensen et al., 2018, 2015; Ward et al., 2021).

2.2.1.2.2. Fluorescence indices

In addition to humic and protein-like fluorescence peaks, fluorescence intensity ratios can be used to assess the origin and transformation degree of DOM. Three fluorescence indices are commonly used: BIX, HIX and FI.

The first is the biological index (BIX) was introduced by Huguet et al. (2009). It was based on ratio β:α from Parlanti et al. (2000), also named freshness index, determined according to laboratory results of macro-algae degradation evolution in ultra-pure and sea water. BIX is an indicator of the contribution of recently autochthonous produced DOM, being the ratio between fluorophore intensities that represents more recently derived DOM, including freshly transferred from algae to the DOM pool, and more decomposed DOM.

The humification index (HIX) is related to the extent of humification of organic matter. It was introduced by Zsolnay et al. (1999), denoted $HIX_{Zsolnay}$, who assumed that humification is associated with an increase of molecular weight. The emission spectra of these more condensed fluorescing molecules tend to shift to longer emission wavelengths. It was used to estimate the degree of DOM aromatization in soil, and firstly applied to aquatic systems by Huguet et al. (2009). It is the ratio of the sum of the fluorescence intensities between the emission wavelengths of 435 nm and 480 nm and the sum of the intensities emitted between 300 nm and 345 nm, at excitation wavelength of 254 nm.

Ohno (2002) proposed a modified HIX calculation, denoted HIX_{Ohno}, introducing a second inner-filter effect correction. With this new ratio, HIX ranges from 0 to 1, also increasing with the increase of DOM aromatization degree. Since HIX values from Ohno's expression are independent of the concentrations of humic-like substances, it facilitates the comparison of results across different studies.

The fluorescence index (FI) was first introduced by McKnight et al. (2001). It is also known as f_{450}/f_{500} index, used by Huguet et al. (2009), and later modified by Cory and McKnight (2005). It allows to determine the source of DOM either microbial (≥1.8) or terrestrial (≤1.4) (Fellman et al., 2010).

2.2.1.2.3. PARAFAC analysis

PARAllel FACtor analysis (PARAFAC) is a multi-way method used to decompose fluorescence EEMs in independent components. It is a powerful tool to identify the unique and the main fluorescing components of a dataset, helping quantifying and characterizing DOM in different aquatic environments (Stedmon and Bro, 2008).

PARAFAC modelling has been contributing to the understanding of the processes involved on DOM concentrations and their characteristics (Stedmon et al., 2003). Over the past decades, PARAFAC analysis had been applied on the identification of fluorescence components as indicators of changes in aquatic environments.

These indicators can be used in the investigation of impacts of land cover (Yang et al., 2018), changes in coastal waters (Asmala et al., 2018) and urban surface waters (Marcé et al., 2021;

Meng et al., 2013; Romero González-Quijano et al., 2022), identification of pollution sources of surface waters (Wang et al., 2022), assessment of treatment of wastewater (Cohen et al., 2014) and drinking water (Wells et al., 2022), etc.

A PARAFAC component is characterised by a couple of loading vectors for excitation and for emission wavelengths. Each component represent a class of DOM of similar fluorophores (Meng et al., 2013).

In order to assemble and compare components of different datasets, the OpenFluor library, an online open-access spectral database was created (Murphy et al., 2014). In this database, similar components among the dataset can be identified. The comparison between components is made through the Tucker congruence coefficient (Tucker, 1951), given by Equation 2.1.

$$\theta(x, y) = \frac{\sum x_i \, y_i}{\sqrt{\sum x_i^2 \; \sum y_i^2}} \qquad 2.1$$

Where x_i and y_i represent the loading vector of the two components to be compared, x and y, for each wavelength element i.

A PARAFAC component is defined by a couple of loading vectors for excitation and for emission wavelengths. The congruence is computed for the excitation (θ_{ex}) and emission (θ_{em}) spectra separately. The final Tucker congruence coefficient is obtained by multiplying them (Equation **2.2**). Two spectra are considered similar when $\theta \geq 0.95$ (Murphy et al., 2014).

$$\theta = \theta_{ex} \cdot \theta_{em} \qquad 2.2$$

Through the years, datasets with different sample sources across the world were added in OpenFluor library, including samples from oceans, estuaries, streams, wetlands, wastewater, recycled water, treated water, among others. The diversity of datasets, allied with additional analysis from each individual study, allowed the establishment of the origin of similar PARAFAC components. For instance, terrestrial components, like humic-acid and fulvic-acid, are commonly associated with agriculture runoff, and protein-like components, such as amino acids like tryptophan and tyrosine, can be associated with wastewater effluent (Yang et al., 2018).

All the steps of PARAFAC modelling are detailed in Material and Methods section 3.4.3.

2.2.1.3. Correction and conversion of fluorescence data

The measurement of fluorescence intensity is influenced by its measuring sensor. The bias and errors of instrumental configuration result from problems on the optical components such as physical imperfections, alignment, and efficiency variation in different wavelengths (Murphy et al., 2010).

Fluorescence intensity is often measured in arbitrary units (AU), being instrument dependent. To neutralize the bias and errors related to instrumental configuration and obtain units of intensity that are comparable between instruments, a total of four to five steps of data correction and conversion is required: two for data correction, and two to three for data conversion.

The correction procedures are applied in every EEM of the dataset. Each EEM will be associated with a blank EEM, i.e., the EEM of a sample of ultra-pure water (deionized water disinfected by UV light). The blank EEM must be obtained as close as possible in time to the

EEM. Therefore, each battery of fluorescence measurement will require a blank EEM, to be associated with the EEMs of the same battery.

The first step is the instrument dependent spectral correction. The EEM and its correspondent blank are multiplied by the excitation and emission correction factors (C), usually given by the instrument fabricant. The second correction is related to the inner filter effect.

The Inner Filter Effect (IFE) is the absorption of part of the incident excitation light before reaching the fluorescence component in the sample, named "primary" IFE, and the re-absorption of part of the emitted light before it leaves the cell with the sample, named "secondary" IFE. Because of this effect, the measured fluorescence intensity may not be proportional to the concentration of its respective fluorophore (Kubista et al., 1994).

Turbidity can affect inner-filter error. The influence of turbidity is most significant when it is above 50 NTU (Bedell et al., 2022). Non-fluorescent impurities may interfere with the interaction between incident light, fluorescent components, and emitting light. In this case, samples need to be diluted and/or filtered. When the sample is diluted enough, a correction with absorbance measurements of the same sample can be applied (Parker and Barnes, 1957).

The next step is blank subtraction. It helps removing the water scatter signal of each EEM by subtracting the blank EEM from the same battery of measurements. It may also neutralize imperfections related to the instrument.

The last step is the calibration of fluorescence signal, or the conversion from AU to a standard fluorescence intensity scale. This step allows the comparison of datasets from different instruments and of measurements from the same instrument but conducted over time.

The most common fluorescence units are the Raman Unit (RU) and Quinine Sulphate Equivalent Unit (QSE or QSU). Their scales are based on the fluorescence of well-characterized standards, such as the ultra-pure water and the quinine sulphate (QS) (Figure 2.4), respectively.

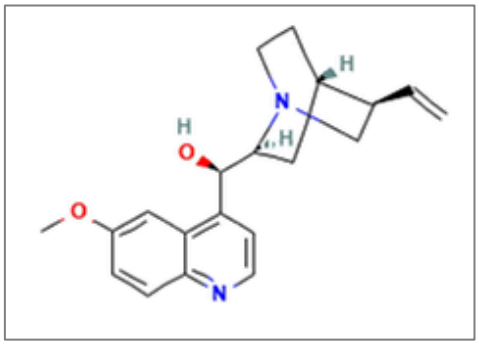


Figure 2.4. Chemical structure of quinine sulphate (QS) (image from https://pubchem.ncbi.nlm.nih.gov/)

The conversion to Raman Unit is also called "Raman normalization". Ultra-pure water has two scatter characteristic scatter peaks; Rayleigh and Raman. Both have first and second order scatter, all of them situated diagonally in the EEM as illustrated in Figure 2.5 (Bahram et al., 2007). Outside these scatters, the fluorescence intensity should be low, close to zero.

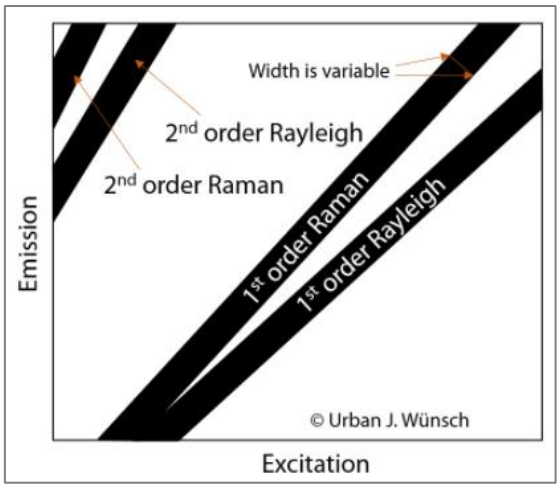


Figure 2.5. First and second order Rayleigh and Raman scatter peaks (from MATLAB© drEEM smoothing tutorial, Scatter excision in EEMs).

Rayleigh scatter is a type of elastic scatter. A direct scattering of the incident light results on a scattering emission appearing at the same wavelength as the incident excitation light. The Raman scatter, on the other hand, is a type of inelastic scatter. A fraction of the energy of the incident photons is transferred, losing energy to vibration in water molecules (Lawaetz and Stedmon, 2009). This energy loss has a fixed frequency of approximately 3400 cm^{-1} (Faris and Copeland, 1997). The scatter is then emitted at higher wavelength than the incident excitation light. Finally, unlike Rayleigh scatter, the Raman scatter peak has relatively low fluorescence intensity and may be overshadowed by fluorophores that fluoresce at the same wavelengths.

The conversion from AU to RU is based on a fixed property of water: the wavelength-dependent Raman cross-section of water. This property is directly proportional to the integral of the measured Raman peak (A_{rp}) (Faris and Copeland, 1997). Since the Raman peak height and width vary accordingly to different instrumental settings or different instruments (Lawaetz and Stedmon, 2009), A_{rp} can be used on the obtention of a standard fluorescence intensity unit. For that, the ultra-pure water blank EEM obtained at the same measurement battery of the EEMs to be converted can be used.

The integral of the Raman peak (A_{rp}) in the ultra-pure water blanks is calculated at excitation wavelength of 350 nm. The range of the emission wavelength, however, may vary depending on the author and the wavelength increment of fluorescence measurement (Murphy et al., 2013). Most commonly, the emission range is between 371 and 428 nm (Table 2.3). The fluorescence intensity in RU is obtained by dividing the values in AU by A_{rp}. The calculation is detailed in section 3.4.1.

Table 2.3. Excitation and emission wavelengths for calculation of the integral of Raman peak (A_{rp})

Excitation wavelength	Emission wavelength range	Source
350 nm	371 to 428 nm	(Lawaetz and Stedmon, 2009)
	380 to 426 nm	(Tedetti et al., 2012)
	374 to 426 nm	(Murphy et al., 2013)

The conversion of fluorescence intensity to QSU unit requires the running of standard curves using quinine sulphate or quinine bisulphate solutions (QS). At least three different

concentrations must be measured at excitation wavelength of 350 nm and emission wavelength of 450 nm.

The standards are prepared considering 1 QSU = 1 ppb quinine bisulphate in 0.05 M H_2SO_4 or in 0.1 N H_2SO_4, where N means normality of a solution, or equivalent concentration (Coble et al., 1993; Murphy et al., 2010). Other procedure uses $1.28x10^{-6}$ M (~1 ppm) of quinine sulphate dihydrate in 0.105 M $HClO_4$ (Velapoldi and Mielenz, 1980).

The slope $m_{350/450}$ is calculated from the linear regression of the QS concentration series. This slope must be corrected. It is multiplied by the correction factor at excitation wavelength of 350 nm and emission wavelength of 450 nm, the corresponding wavelengths of the measurement of QS dilution series (Murphy et al., 2010). The EEM in QSU can be obtained by dividing the corrected EEM by the slope $m_{350/450}$.

A calibration between fluorescence intensity in QSU (F_{QSU}) and in RU (F_{RU}) is possible. Murphy et al., (2010), in their interlaboratory comparison study, obtained the EEMs in QSU from the calibrated EEMs to RU. However, the slope QSU/RU, the multiplier for conversion between QSU and RU, is instrument-specific (Murphy et al., 2010). In addition, it may depend on the emission wavelength range used on Arp calculation for the obtention of EEMs in RU.

As an example, Lawaetz and Stedmon (2009) obtained QSU/RU = 13.04 (F_{QSU} = 13.04 F_{RU}), with A_{rp} integrated from λ_{Em} 371 to 428 nm and QS concentration range of 0-10 ppb. Tedetti et al. (2012) obtained QSU/RU = 71.43 (F_{QSU} = 71.43 * F_{RU}), with Arp integration from 380 to 426 nm. In their study, Murphy et al. (2010) found a QSU/RU range of 62.3-113.6, with A_{rp} integration from 381 to 426 nm and QS range of 0-100 ppb.

One main advantage of the Raman Unit calibration compared to the QS calibration includes to avoid the use of a reagent. Since no standards are required, the operational steps to obtain different QS concentrations and the risk of their degradation are removed, reducing the potential errors involved on the process. Raman normalization requires only ultra-pure water, a non-hazardous product available in most laboratories (Lawaetz and Stedmon, 2009).

All the steps of the EEM correction procedures were detailed by Murphy et al. (2010) and are presented in section 3.4.1.

2.2.2. Physical-chemical variables related to microbial contamination

Water quality variables to use as a fast and good proxy for the microbiological water quality survey are being studied over the past decades.

Considerable faecal contamination is expected after episodes of heavy rainfall that should bring them to the waterbody by runoff and stormwater inflow. Dissolved organic matter should also show a relation with microbial contamination (Frank et al., 2018a; Nowicki et al., 2019; Pronk et al., 2006; Sorensen et al., 2018, 2015). A major factor that influences the concentration of total and faecal coliforms is the rainfall (Frank et al., 2018a; Nnane et al., 2011; Seo et al., 2019). Electrical conductivity, or specific conductance, is one of the variables that indicates changing of water quality, particularly after heavy rainfall, and can be used to detect the arrival of a new mass of water.

2.2.2.1. Electrical conductivity

Electrical conductivity (hereafter, "conductivity") is defined as the ability of a substance to conduct electrical current. In water, a higher concentration of dissolved ions may increase the electrical currant that can be conducted.

Conductivity also varies with water temperature, increasing with higher temperatures. Therefore, the adoption of a water temperature of reference is needed, to allow comparison between measurements. The reference temperature for the specific conductance is of 25 °C. In this study, all the conductivity values correspond to the electrical conductivity at 25 °C, also kwon as "specific conductance".

According to US EPA (2013b), each water body tends to have a typical range of conductivity. Once this relatively constant range is identified, it can be used as a baseline to identify a change of water quality, for example caused by a stormwater inflow.

Continuous measurement allows to detect significant changes of the quality of water, being a potential indicator of a discharge or other source of pollution in the water body (US EPA, 2013b). In addition, when the conductivity is continuously measured at upstream and downstream of a river stretch, an estimation of the transfer time is possible by comparing when those significant changes were observed at both points (see section 4.1.5.3).

2.3. Autonomous monitoring system

Climate change has been affecting different regions around the world. Among its consequences, the availability and quality of continental water and the increase of extremes events of heavy rainfall and heat waves are of great concern (IPCC, 2023).

The elevated population density, high level of impermeabilization and low proportion of green areas contribute to the vulnerability of urbanized areas to heavy precipitation and elevated temperatures. However, climate change adaptations such as urban greening have been effective in reducing flood risks and urban heat (IPCC, 2023).

The opening of bathing areas in urban waterbodies can help local residents deal with heat waves, contributing to human well-being (Meyerhoff et al., 2010). Early warning systems can also help dealing with extreme weather events. Online monitoring systems have been developed for flood risk prediction (Krommyda et al., 2019), and for microbiological water quality monitoring (Offenbaume et al., 2020).

Long-term monitoring can be very useful on models and on the evaluation of local climate impact (IPCC, 2023). Monitoring water quality programmes are usually based on low-frequency measurements (EU, 2006), commonly weekly or monthly. However, water quality parameters can have wide spatial-temporal variations, particularly after certain weather conditions such as storm events (Bhurtun et al., 2019), that could be unnoticed with a large interval between measurements. Furthermore, processes that occurs over a shorter duration than the interval between measurements could not be detected.

High-frequency measurements can be important to assess sampling strategies and assign classes of water courses. Moreover, in long-term, they allow to detect temporal trends in several different scales, from hourly to yearly intervals (Skeffington et al., 2015), and improve long-term surveillance programs, especially for tracking local consequences of climate change in urban areas. Analysing an extensive database acquired for several months or years would improve the understanding of the environment and its evolution in time and space (Le Vu et al., 2011).

The availability of a wide range of sensors allows the obtention of high-frequency measurements from different variables related to physicochemical functioning of water bodies. The collection and exploitation of this data may enable not only the identification of tendencies due to climate change in general but also to better understand processes related to local extreme events like heat waves and heavy rainfall.

In addition, the implementation of a continuous monitoring system could help short-term decision-making. A short interval between measurements would reduce the risk of having an undetected contamination arriving from upstream. A short-term response would allow a fast decision-making concerning drinking water extraction or recreational activities.

If implemented in a watercourse, it would be possible to detect the arrival of a contamination at an upstream point of the area of interest, having enough time to close it if needed. Combined with a monitoring point at the site of interest, it will be possible to identify the arrival of the contamination and how long it would last in the area. The combination of upstream and downstream measuring points would allow the estimation of transfer time between them and to better understand the evolution of the measured variables in time (Kraus et al., 2017).

To improve the management of recreational activities in urban waters, high-frequency data can also be combined with modelling techniques. Models to estimate and predict the water quality of a bathing site from real-time measurements can be developed (Khac et al., 2018).

The reliability of these models lies on the quality of measured data. Procedures for data quality assurance and control are then required (Tiberti et al., 2021). All over the years, and with the popularisation and availability of cost-efficient sensors and high-frequency monitoring systems, methods were developed to help distinguish measurement anomalies from real environmental changes (Moatar et al., 2001; Mourad and Bertrand-Krajewski, 2002).

2.4. Numerical modelling

For a better representation of microbiological contamination in aquatic environments, it is important to integrate hydrodynamic and microbiological modelling (Sokolova et al., 2012). In this chapter, the hydrodynamic model adopted on this study is briefly discussed and some microbiological models for *E. coli* settling and decay rate are presented.

2.4.1. Hydrodynamic modelling

Hydrodynamics is the study of liquids in motion. The understanding of differential equations allowed to mathematically describe de motion of liquids. The combination of this equations with the development of computation methods resulted in powerful models capable of simulate the motion of liquids in different environments within few minutes.

The spatial dimension of a hydrodynamic model, from 1D to 3D, may depend on the context and objectives of its application. In a three-dimensional model, variables with three dimensions are computed, considering both horizontal and vertical variabilities. Spatial heterogeneities can be reproduced in horizontal and vertical plans.

Many hydrodynamic models were developed and used on the estimation of faecal contamination in water courses (Huang et al., 2017). Here, we are particularly interested in the transfer time of microbiological contamination, more specifically, of faecal contamination, from an upstream monitoring point to an area of interest downstream, a

bathing area. To obtain the spatiotemporal distribution, including vertical and horizontal distribution, of an indicator of faecal contamination, a 3D model is required.

The hydrodynamic modelling used on this study is entirely based on TELEMAC-3D, from TELEMAC-MASCARET modelling system.

This hydrodynamic model, coupled with a water quality module WAQTEL, has been used for studying the transport and fate in nearshore coastal waters (King et al., 2021) and in rivers (Van et al., 2022).

The choice of this model was based on several aspects. It is a free and open-source model with good hydrodynamic representation in multidimensions. In addition, it is fully coupled TELEMAC-3D with the water quality module WAQTEL, and/or with the biogeochemical library Aquatic EcoDynamics library (AED2), allowing enhancements on the computation of the transport of microbiological contamination. At each computation time step, the equations of all activated modules are solved, resulting in an evolution of multiple aspects of the simulation at the same time (EDF R&D and Telemac3D consortium, 2021a). In addition to the TELEMAC-3D hydrodynamic module, the second module of interest was WAQTEL, which includes the computation of the heat exchange between the water body and atmosphere. The expressions of the heat fluxes are presented in more detail in section 3.5.3.1.2.

Particularly, the thermal stratification is of interest because of the change of water density with the water temperature (Boehrer and Schultze, 2008), which may change the stability of the water column and lead to a change of water velocity and transfer time of contaminants.

More details about the thermal stratification and TELEMAC-3D software are presented on Annex II and section 3.5, respectively.

2.4.2. Microbiological modelling

The microbiological modelling considers the fate of the microorganisms, such as the faecal indicator bacteria (FIB). Once in water bodies, the decrease of bacterial concentrations in aquatics environments is due to a combined actions of biological (bacteriophages and lysis, for example) and physicochemical parameters (such as sunlight, temperature, and nutrients availability), and of sedimentation (Auer and Niehaus, 1993; Servais et al., 2007).

Mathematical models incorporates the main sources and sinks of this type of microorganism, such as discharges, mass transport and losses due to death and sedimentation (Auer and Niehaus, 1993). For a better representation of aquatic environments, it is important to consider a decomposition between free faecal bacteria and attached bacteria to suspended solids and to settled solids (Van et al., 2022). The following sections present models of FIB decay rate and settling.

2.4.2.1. *E. coli* decay rate

The decrease of faecal pathogens concentration in time can vary among the different characteristics of aquatic environments and according to the hydrodynamic properties of the water course. In water bodies with advection dominance, such as fast-flowing rivers, decay rate is not a critical parameter on the assessment of microbiological water quality for bathing areas. However, faecal indicator decay rate is very important when assessing dispersion dominant water bodies (Martin et al., 2023).

The decrease in time (t) of faecal indicator bacteria, such as *E. coli*, is assumed to follow an exponential first-order law (Equation 2.3)

$$C(t) = C(t = 0) \cdot e^{(-k \cdot t)} \quad 2.3$$

Where C is the bacteria concentration, and k is the decay rate at 20 °C.

Selméus (2018) presents three different approaches to describe the *E. coli* coliform rate in natural waters. The first approach was proposed by Mancini (1978). It describes the effect of light intensity, water temperature and salinity, being the model with less variables of the three.

The second degradation rate expression is proposed by Deltares (2019) in their Water Quality module "D-Water Quality", for the DELFT3D software. It is an empirical equation which depends on the water temperature, depth, intensity of UV radiation, salinity (more specifically, chloride concentration) and inorganic suspended matter concentration.

The third expression was based on the Chick and Martin model (1908) and was used by Erichsen et al. (2006). In addition of the aspects considered on the second expression, it takes into account the contributions from light and dark conditions, and it is valid between 12 and 24 °C. It also considers the water transparency (Secchi depth, SD).

The water quality module of TELEMAC-3D, named WAQTEL, considers the same degradation law, with the following expression for the decay rate (Equation 2.4):

$$k = \frac{2.3}{T_{90}} \quad 2.4$$

Where T_{90} is the amount of time necessary to consume 90% of the initial mass. Its value usually is between 20 h and 70 h (Hong et al., 2019; King, 2019; Poulin et al., 2013; Van et al., 2022).

2.4.2.2. *E. coli* settling

In aquatic environments, faecal bacteria can be attached to suspended particles and be removed from the water column through sedimentation, resulting in important implications on their fate in water courses (Servais et al., 2007).

The interaction between FIB and sediment results in three groups: free FIB, attached FIB to suspended sediment, and attached FIB to bed sediment. The effects of sedimentation and resuspension are related to the last two groups. In TELEMAC-MASCARET water quality module WAQTEL, the sub-module MICROPOL simulates the evolution of micropollutant in these three groups, also considering the micropollutants absorbed by suspended and bed sediment (EDF R&D and Telemac3D consortium, 2021b).

The evolution of suspended solids (SS) is assumed to follow the deposition and resuspension laws for cohesive sediment of Krones and Partheniades (Equations 2.5 and 2.6).

$$SED = \begin{cases} wSS\left(1 - {\tau_b}/{\tau_s}\right) & \text{if } \tau_b < \tau_s \\ 0 & \text{if } \tau_b \geq \tau_s \end{cases} \quad 2.5$$

$$RS = \begin{cases} e\left({\tau_b}/{\tau_s} - 1\right) & \text{if } \tau_b > \tau_r \\ 0 & \text{if } \tau_b \leq \tau_r \end{cases} \quad 2.6$$

Where SED and RS are, respectively, the deposition and resuspension (erosion) fluxes (in kg m^{-2}.s^{-1}), wSS is the settling velocity (in m.s^{-1}), e is the Patherniades constant, representing the rate of erosion characteristic of deposited suspended particulate matter

(in kg $m^{-2}.s^{-1}$), τ_b is the bottom shear stress (in Pa), and τ_s and τ_r are, respectively, the critical shear stress for sedimentation and for resuspension (in Pa).

The suspended solids (SS) and the bottom sediments (SF) are computed in function of the water level (h) (Equations 2.7 and 2.8).

$$SS = \frac{RS - SED}{h} \quad 2.7$$

$$\frac{\partial(SF)}{\partial t} = SED - RS \quad 2.8$$

The concentrations of micropollutants also include a decay rate using an exponential first-order law (similar to Equation 2.3). The model computed the concentrations of free micropollutants in water, micropollutants adsorbed by SS and adsorbed by bottom sediments.

Due to the similarity between micropollutants and FIB, this model is able to simulate the evolution of faecal contamination in time and space (Van et al., 2022).

Chapter 3. Materials and methods

In this chapter, the study site is described, along with the measurement equipment used, the treatment of fluorescence data, and the hydrodynamic model of the study site.

3.1. Study site

3.1.1. La Villette system

Located in the northeast of Paris (France), La Villette system (48.8918 N, 2.3855 E) is composed by a canal upstream that widens to a basin downstream. It is open to recreational activities, such as boating and the swimming. Since 2017, a free bathing area has been open during summer through "Paris Plage" program of Ville de Paris.

The upstream part of La Villette system corresponds to a canal of 25 m wide and 800 m long (La Villette canal). Then, it widens to 75 m wide, forming a basin of 700 m long (La Villette basin). The whole system has a total extension of 1500 m and approximately 3 m of depth. From 2012 to 2016, the average flow rate during summer (mid-June to mid-September) was 2.7 $m^3.s^{-1}$ (SAFEGE, 2017a).

Upstream, the Ourcq canal is divided in two branches connected by a roundabout, Saint-Denis canal and La Villette canal. The La Villette canal receives 2/3 of the flow from the Ourcq canal and the Saint-Denis canal receives about 1/3 of the flow (Vanhalst, 2018).

Downstream, on the right bank, a water pumping station withdraws water towards the non-drinking water network. The daily mean discharge was $1.64 \cdot 10^3$ $m^3.day^{-1}$ (1.9 $m^3.s^{-1}$) in 2021 (data from Ville de Paris). This network is mainly used to street cleaning, and park and garden watering. On the left bank, a navigation lock links La Villette basin to Saint-Martin canal, flowing downstream towards the river Seine.

The flow rate in La Villette system is very regulated. Upstream La Villette canal, the Ourcq canal is equipped with 10 navigation locks and two pumping stations along the 108 km stretch, from Port-aux-Perches to La Villette basin (Figure 3.1).

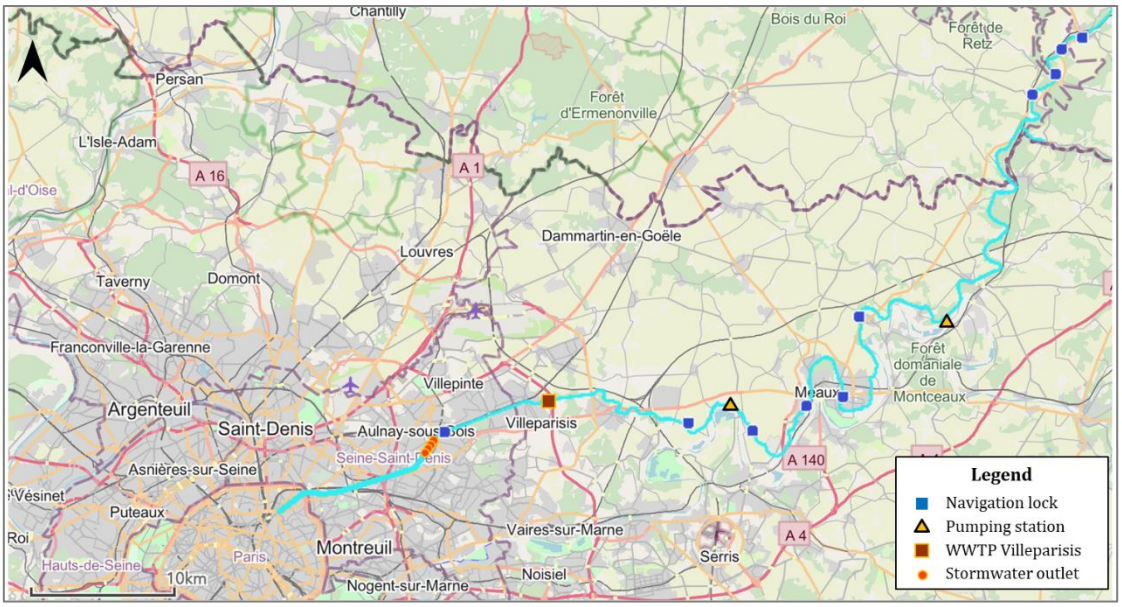


Figure 3.1. Ourcq canal upstream La Villette system (Source: https://capgeo.maps.arcgis.com, Ville de Paris).

The canal roundabout is 9.7 km downstream from a stormwater network outlet (SAFEGE, 2017a) and 20 km downstream from the WWTP of Villeparisis (Figure 3.1). The mean WWTP discharge in 2021 was 0.13 $m^3.s^{-1}$ (Portail assainissement collectif, 2021).

At the canal roundabout, a microbiological measurement station is installed. It is equipped with a Fluidion ALERT system (detailed in section 2.1.2.2.1) to monitor the FIB concentrations during summer. The measurements of *E. coli* and intestinal enterococci (IE) concentrations are conducted thrice a day, at 19h, 20h and 21h (Paris time). On the next morning, a mean of the measured concentrations is compared with the reference *E. coli* and IE bathing thresholds. If the thresholds are respected, the bathing area can be opened (Vanhalst, 2018).

The La Villette system, canal and basin, is presented in the Figure 3.2 and summarized on Table 3.1.

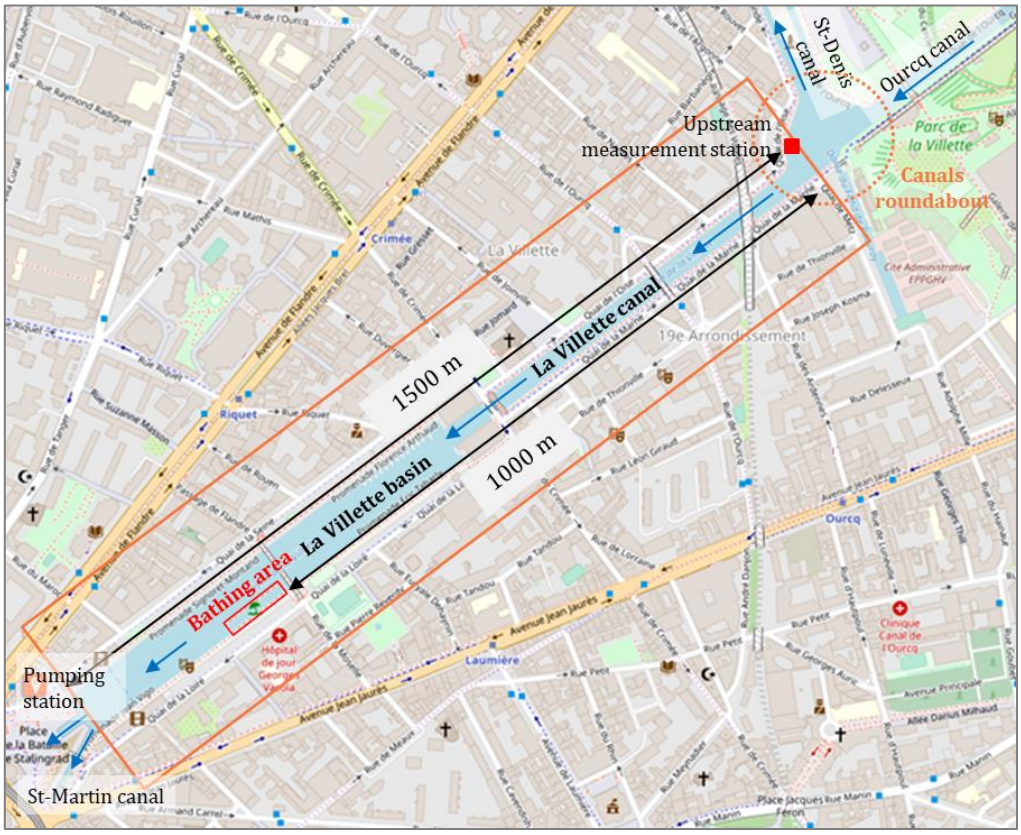


Figure 3.2. La Villette system (base map: www.geoportail.gouv.fr/carte)

Table 3.1. Characteristics of the study site (SAFEGE, 2017a)

Characteristics	Values
Canal	
Width	25 m
Length	800 m
Basin	
Width	75 m
Length	700 m
Distance between measurement station and bathing area	1000 m
Average flow rate during summer (from 2006 to 2016)	2.7 $m^3.s^{-1}$
Depth	Around 3 m
Fluidion sampling times (Paris time)	19h00, 20h00 and 21h00

3.1.1.1. Measurement points of La Villette study site

3.1.1.1.1. Continuous measurements

The three measurements points, named A, B, and B_0 (Figure 3.3) are equipped with sensors of continuous measures. Point A is located upstream (canal round-about). Point B is located downstream, at the right bank of the basin, close to the bathing area.

Point B_0 is located at the bathing area, at the left bank of the basin, attached to the bathing structure. Since the bathing structured is installed exclusively during summer, the measurements at point B_0 were from 01/08/2020 to 03/09/2020. After that, point B was used for continuous measurements at the basin.

The measurement points of the study site were equipped with water temperature sensors at three depths, electrical conductivity sensors and a water level sensor. The measuring system is detailed in section 4.1.

A detailed description of the location of the measurement points is presented in the following paragraphs.

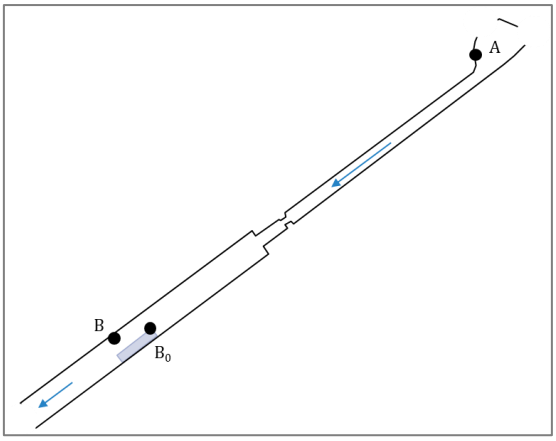


Figure 3.3. Measurement points of the field campaign.

The point A (48.891732 N, 2.3855 E) is the upstream point, located at the roundabout of the canals, on the inlet of La Villette system, at the right bank of the canal (Figure 3.4). It has the same location of the microbiological measurement station, where sensors are installed inside a protection cage (Figure 3.5).

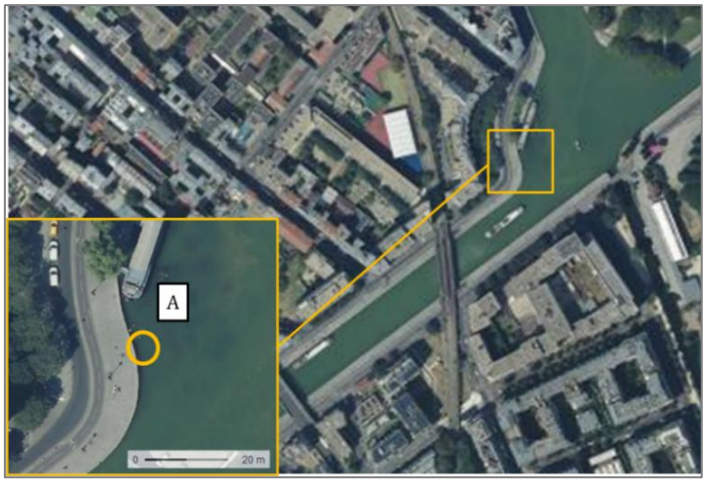


Figure 3.4. Location of point A (upstream).

Figure 3.5. Microbiological measurement station (point A).

The point B (48.885805 N, 2.373751 E) is located at the right bank of the basin, on the northeast extremity of Marin d'Eau Douce platform, close to the bathing area (Figure 3.6). This platform belongs to a company of electric boat renting, a leisure activity in operation almost all year round, closing only in winter.

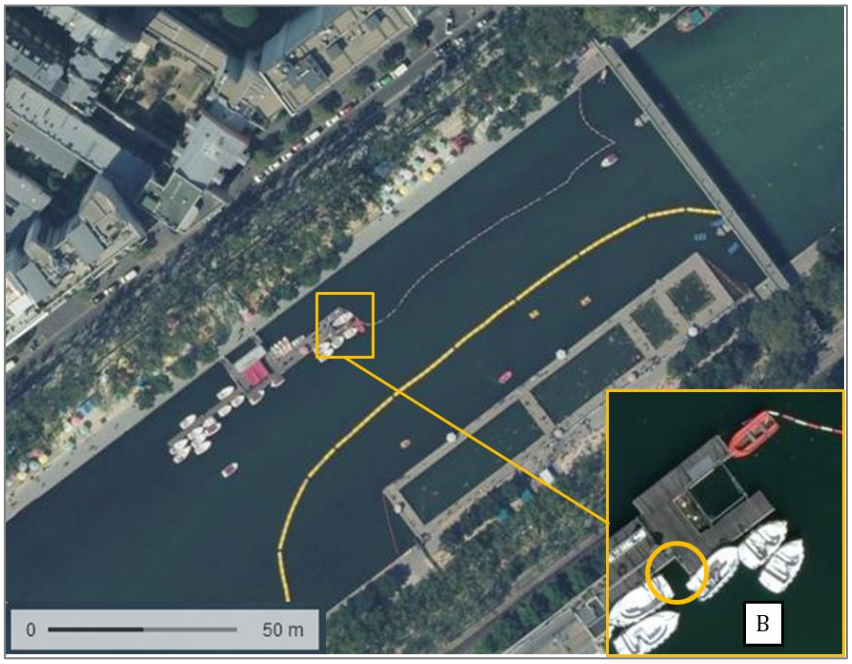


Figure 3.6. Location of point B (Marin d'Eau Douce platform).

The point B_0 (48.885998 N, 2.375096 E) is located at the extremity north of the bathing area platform, at the left bank of the basin. It is the point at the beginning of the bathing area the most at the centre of the basin, as it is shown on Figure 3.7.

The platform is installed exclusively during summer, as part of the "Paris Plage" program, usually from June to September. For this reason, point B_0 was only used as measuring point from 01/08/2020 to 03/09/2020.

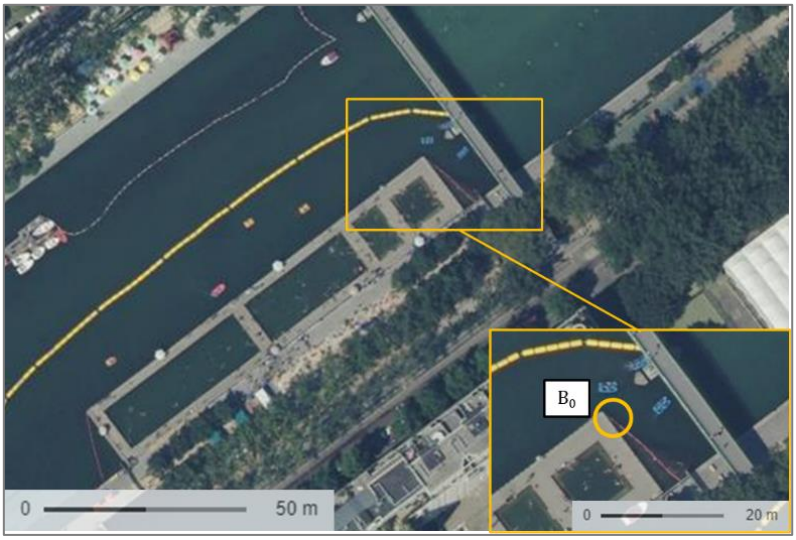


Figure 3.7. Location of point B_0 (bathing area platform).

3.1.1.1.2. Seasonal measurements

Between June and September 2021, some samplings were made at point C, between points A and B (Figure 3.8). It is located at Pont de Crimée, a few meters upstream of the widening to the basin (Figure 3.9).

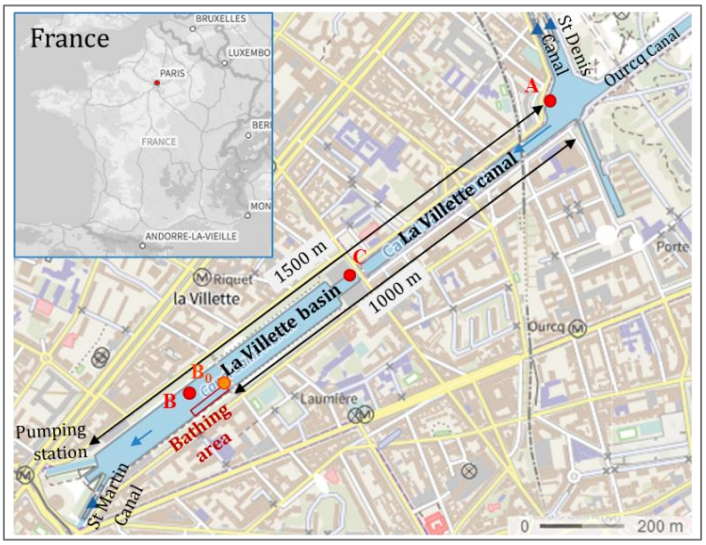


Figure 3.8. Location of La Villette system and the measurement points A, B_0, B and C (base map: geoportail.gouv.fr/carte).

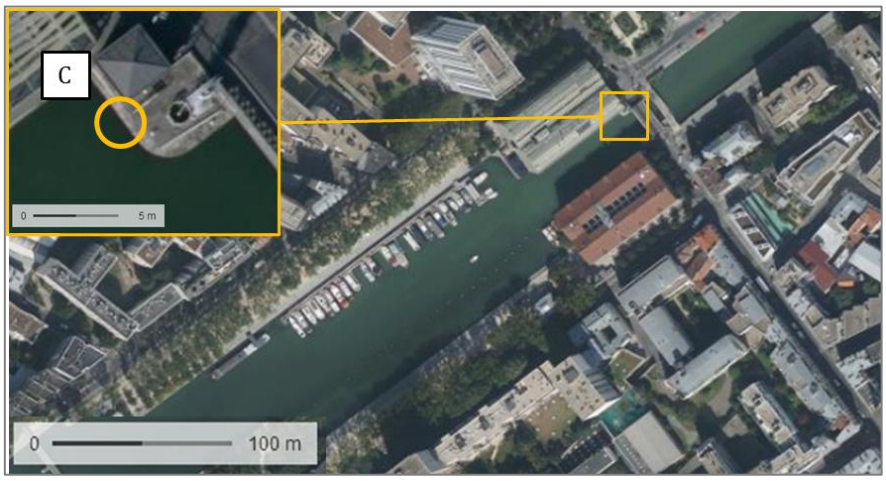


Figure 3.9. Location of point C (Pont de Crimée).

3.1.2. Complementary sampling points

To have a wider characterization of watercourses via FDOM, samples from different points in the river Marne were collected. In addition, samples from stormwater network and WWTP outlets were collected to have a wider range of fluorescence intensities and FIB concentration The aim was the detection of a fluorescence signal related to the microbiological contamination related to this type of outlets.

3.1.2.1. Natural surface water: the river Marne

The studied stretch of the river Marne is around 30 km long (48.8758 N 2.6721 E upstream, 48.8163 N 2.4414 E downstream). During the studied period (15/06 to 20/10/2022) the mean flowrate was around 36 $m^3.s^{-1}$ (HydroPortail, 2022). The samples were collected, in 19 sampling points along the river stretch. The point VGA is located in Saint-Maur-des-

Fossés (VGA sailing club). The other 18 points are part of *Syndicat Marne Vive* monitoring program (points SMV). Among SMV points, 6 correspond to cities candidates to open bathing sites as a legacy of the Paris Olympic Games in 2024. They were considered as "main points" due to a three times higher sampling frequency than the other 12 minor sampling points (Figure 3.10).

In this stretch are located two outlets of wastewater treatment plants (WWTP). The most upstream is WWTP St-Thibault-des-Vignes (hereafter WWTP-O), with 44 000 m^3 wastewater treated per day (0.46 $m^3.s^{-1}$) (Portail assainissement collectif, 2021). It discharges at 7.8 km upstream point SMV1.

The second is WWTP Paris Marne Aval (WWTP-PMA), with 35 000 m^3 wastewater treated per day (0.41 $m^3.s^{-1}$) (Portail assainissement collectif, 2021). It is located in Noisy-le-Grand, but discharges at 1.8 km upstream point SMV10, in Saint-Maur, on the second half of the stretch (Figure 3.10). In addition, in the studied stretch, 80 SN outlets and/or combined sewage overflows discharge in the river Marne.

The land use data of the river Marne sampling sites was obtained from the open data platform from *Institut Paris Region* (2022). A squared area of 1 km^2 immediately upstream the sampling points was considered to study the influence of their immediate surroundings. Due to the lack of samples from the SMV minor sampling points (n=4 at each point), only the 6 major points were explored (Figure 3.11). The main land use was impervious area, with average of 65% in the river Marne SMV major sampling points.

The distance between sampling points (Table 3.2) and the proportional land-use areas (Figure 3.12, Table 3.3) were calculated using an open-source geographic information system software (QGIS 3.30).

The sampling weather conditions were obtained from Météo-France stations: for La Villette basin Le Bourget weather station (48.96722 N, 2.42778 E, 9 km to the north), and in the river Marne, from Torcy weather station (48.86361 N, 2.65111 E, 5 km to the east from SMV1).

3.1.2.2. Stormwater network and WWTP outlets

Two stormwater outlets (points SO-N and SO-S, Figure 3.10) were sampled. In both catchments, with a surface area respectively around 500 ha and 1400 ha, the drainage network is separated and discharged in the river Marne.

The sampling was conducted at the outlet of their respective catchment. At SO-N, the water is directly released into the river Marne. At SO-S, downstream the sampling point, the water is screened then passes through two settling tanks before being released into the river Marne.

Samples of the treated water of WWTP St-Thibault were collected before its release into the river Marne (point WWTP-O in Figure 3.10). This WWTP is the around 8 km upstream the point SMV1 with no intermediate WWTP between them.

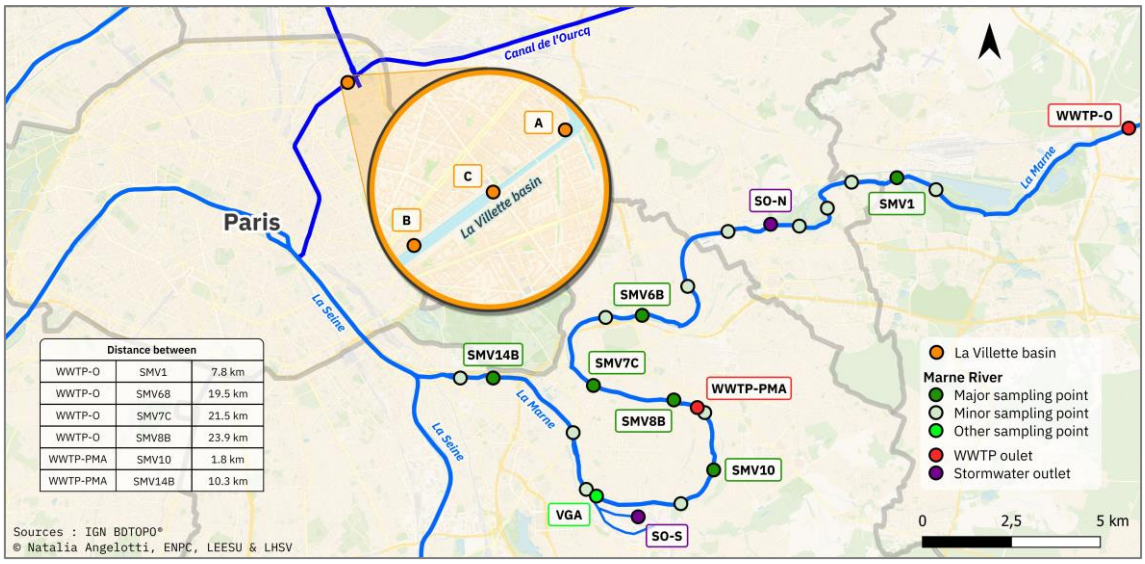


Figure 3.10. Sampling points. La Villette basin (orange), Marne River SMV (dark green) and VGA (light green), stormwater outlets SO-N, SO-S (purple) and WWTP outlet WWTP-O (red).

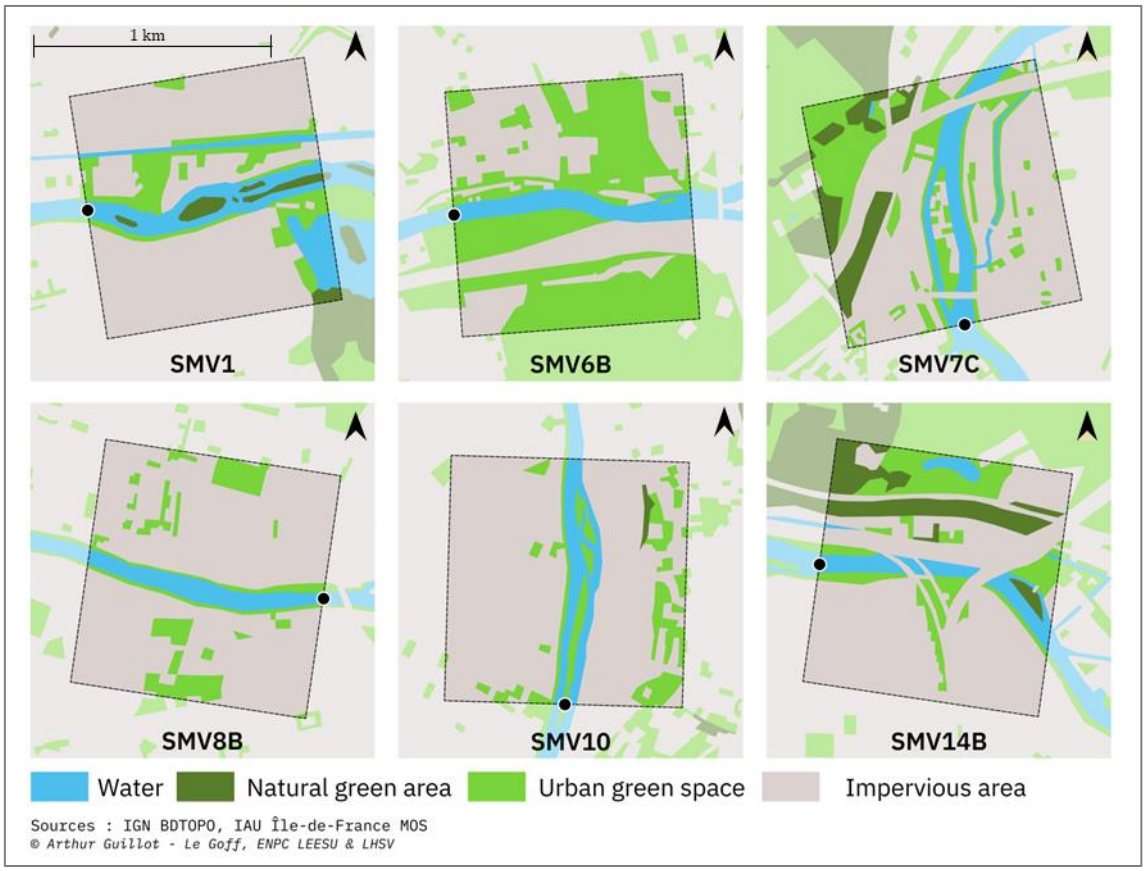


Figure 3.11. Land-use of La Villette basin and the river Marne main SMV points (Sources: IGN BDTOPO, IAU Île-de-France MOS, Arthur Guillot – Le Goff).

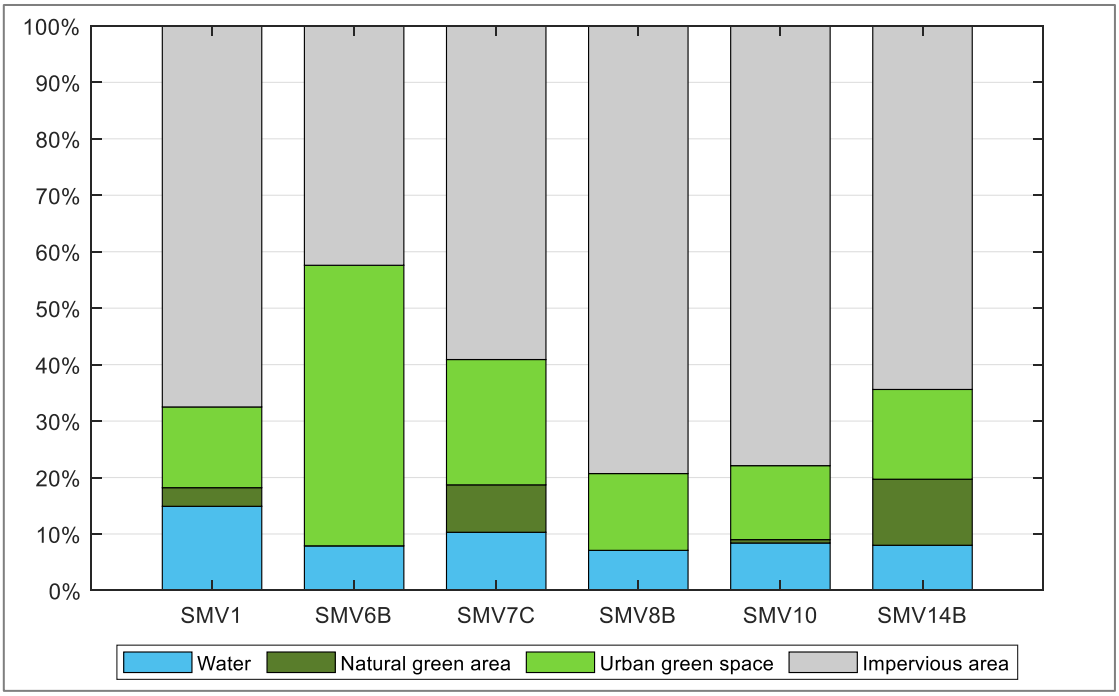


Figure 3.12. Land-use (%) of the squared area of 1 km² immediately upstream of the main river Marne sampling sites.

Table 3.2. Names, coordinates and distance from closest point of interest of sampling sites (other sampling point, stormwater network or WWTP outlet upstream). Latitude and longitude are given in decimal degrees North and East, respectively.

Site ID	Site name and city	Water type	Latitude (N)	Longitude (E)	Distance from the closest point of interest
SO-N	Urban centre of Noisy-le-Grand	Stormwater network outlet	48.852811	2.541739	~2km upstream WWTP-PMA.
SO-S	Sucy-en-Brie	Stormwater network outlet	48.771383	2.508284	Upstream of points VGA and SMV11.
WWTP-O	Station of Lagny-sur-Marne Saint-Thibault-das-Vignes	Wastewater treatment plant outlet	48.875812	2.672124	~7km upstream point SMV1
A	Upstream La Villette canal, Paris	Canal	48.891755	2.385479	~20 km downstream the WWTP from Villeparisis
B	La Villette bathing area, Paris	Canal	48.885802	2.373753	1.1 km downstream from point A
C	Crimée bridge, Paris	Canal	48.888371	2.379246	0.6 km downstream from point A
SMV1	Chelles	River	48.865599	2.583061	7.8 km downstream from outlet WWTP-O.
SMV6B	Nogent-sur-Marne	River	48.830919	2.477830	19.6 km downstream WWTP-O. 11.8 km downstream point SMV1.

Site ID	Site name and city	Water type	Latitude (N)	Longitude (E)	Distance from the closest point of interest
SMV7C	Joinville-le-Pont	River	48.819127	2.469216	21.3 km downstream WWTP-O. 1.7 km downstream point SMV6B.
SMV8B	Champigny-sur-Marne	River	48.810578	2.502057	23.9 km downstream WWTP-O. 2.7 km downstream point SMV7C.
SMV10	Saint-Maur-des-Fossés	River	48.792178	2.522323	1.8 km downstream discharge outlet of WWTP-PMA. 3.2 km downstream point SMV8B.
VGA	VGA Voile Club, Saint-Maur	River	48.788302	2.474838	5.8 km downstream discharge outlet of WWTP-PMA. Immediately upstream SMV 12 and 4.0 km downstream SMV10.
SMV14B	Saint-Maurice	River	48.816339	2.441363	10.3 km downstream discharge outlet of WWTP-PMA. 8.5 km downstream point SMV10.

Table 3.3. Land-use (%) and mean population density (people.km^{-2}) of the squared area of 1 km^2 immediately upstream of each sampling site. "Agriculture" and "Forest" (green) were grouped in "natural green areas".

Marne River sampling points	Dist. (km)	Water	Agriculture	Forest	Natural green areas	Urban green space	Impervious area	Mean pop. density (people.km^{-2})
SMV1	7.8	15%	0%	3%	3%	14%	68%	477.9
SMV6B	19.6	8%	0%	0%	0%	50%	42%	854.7
SMV7C	21.3	10%	2%	7%	9%	22%	59%	714.8
SMV8B	23.9	7%	0%	0%	0%	14%	79%	780.0
SMV10	27.2	8%	0%	1%	1%	13%	78%	437.2
SMV14B	35.6	8%	0%	12%	12%	16%	64%	742.7
Average	-	*9%*	*0%*	*4%*	*4%*	*21%*	*65%*	*667.9*

3.2. Measuring instrumentation

La Villette basin is monitored by LEESU though profiling and the deployment of high frequency *in situ* measurements. The high-frequency monitoring system measured water temperature, conductivity (specific conductance at 25 °C), turbidity and water level. The design of the monitoring system and the measuring equipment are presented in section 4.1.

3.2.1. Sensor calibration in laboratory

Before being installed, conductivity and turbidity sensors were calibrated. When not possible, sensors were sent to manufacturer for maintenance and calibration.

For conductivity calibration, five standard solutions were prepared: 0, 74, 147, 720 and 1413 $\mu S.cm^{-1}$. When for verification, only 0, 720 and 1413 $\mu S.cm^{-1}$ standard solutions were used. For turbidity calibration, eight standard solutions were used: 0, 20, 50, 100, 200, 400, 800 and 1000 NTU, with only 0, 100 and 800 NTU for verification.

In both parameters, 10 measurements were recorded for each standard solution. The standard deviation, average and error in percentage were calculated and compared to the precision of the measure. In the case of verification. if the deviation or error are greater than the precision, a calibration is required. In that case, when the calibration is done in the laboratory, the expected value is set on the equipment and another verification is conducted.

3.2.2. *In situ* validation of field data

Regular campaigns were performed at points A and B for equipment maintenance and field data validation. Water transparency is measured with a Secchi disk. Water temperature, conductivity and oxygen profiles are measured with a SeaBird (SEACAT SBE19, Sea-bird Electronics, Inc.) and WiMo (nke Instrumentation) probes (Table 3.4).

Table 3.4. Characteristics of reference probes and Secchi disk

Characteristics	SeaBird	WiMo	Secchi Disk
Model	SEACAT SBE 19-03 335 m	WiMo multiparameter sonde	26 cm of diameter
Fabricant	Sea-bird Electronics, Inc.	Nke instrumentation	-
Measured parameters	Temperature, conductivity, and pressure (depth), dissolved oxygen	Conductivity, temperature, water pressure, oxygen concentration, oxygen saturation, pH	Water transparency
Measurement range	Temperature: -5 to +35 °C Conductivity: 0 to 7 $S.m^{-1}$ Depth: up to 335 m Pressure: 500 psia DO: 120% of surface saturation	Temperature: -2 to +35°C Conductivity: 0 to 100 $mS.cm^{-1}$ Water pressure: 0-25 bars Oxygen concentration: 0-23$mg.L^{-1}$ Oxygen saturation: 0-250% pH: 0-14 pH units	0.2 to 5.0 m

Characteristics	SeaBird	WiMo	Secchi Disk
Accuracy	Temperature: 0.01 °C /6 months Conductivity: 0.001 $S.m^{-1}$ /month Pressure:0.25 % of full-scale range DO: 2% of saturation	Temperature: ±0.02°C Conductivity: ±0.5% of reading Water pressure: 0.15% Oxygen concentration: ±0.1$mg.L^{-1}$ Oxygen saturation: ±1% of reading pH: ±0.1 pH unit	0.05 m
Resolution	Temperature: 0.001 °C Conductivity: 0.0001 $S.m^{-1}$ Pression: 0.015 % of full-scale range	Temperature: 0.001°C Conductivity: 0.001 $mS.cm^{-1}$ Water pressure: 0.001 bar Oxygen concentration: 0.025$mg\ L^{-1}$ Oxygen saturation: 0.05% pH: 0.01 pH unit	0.05 m

3.3. Water sampling for FDOM measurement

3.3.1. Sampling method

Two types of samples were collected, grab samples and 24-hour integrated samples, during dry and wet weather. Wet weather is defined as a rainfall episode higher than 2 $mm.day^{-1}$ within 3 days.

All samples were covered in aluminium foil for protection against sunlight and stored at 4 °C after sampling and analysed within 24 h in the laboratory. All measurements are made in room temperature.

Only the turbidity of samples from the stormwater network outlets (points SO-N and SO-S) were higher than 10 NTU. To avoid interference on fluorescence and absorbance measurements (Bedell et al., 2022), the samples were filtered with a 0.45 µm glass microfiber filter (Whatman® Cat No 1825 047).

Due to high concentration of FDOM, samples from points SO-N, SO-S and WWTP-O were diluted with ultra-pure water with dilution factor of 1/5.

3.3.1.1. Punctual grab sampling

For the grab sampling, 100 mL of water were manually collected under the water surface at a pre-established schedule, covering wet and dry weather, in HDPE (high density polyethylene) flasks.

At La Villette basin, grab samples were collected at points A and B, twice a month from October 2020 to March 2023 (n=85 in each point). At point C, four grab samples were collected: one in 22/06/2021 and three in 16/12/2021 (Table 3.5).

At SMV points, grab samples were collected during summer 2022, three times a week in average, with a total of 189 samples. Six of the SMV sampling points formed a main group, where sampling collection was 3 times a week. A total of 4 cruises were conducted in all 18 SMV points on the same day.

3.3.1.2. Integrated sampling

During dry weather, 24-hour integrated sampling was made by an automatic sampler at an hourly time step. 24 samples (1L each) were collected under the surface.

In wet weather, 24-hour integrated samples were collected in La Villette basin (point C), the river Marne (point VGA), WWTP outlet (point WWTP-O), and SN outlet SO-S.

In point SO-N, the automatic sampler was activated according to the water level of the stormwater network. A five to nine-hour integrated sampling was made by the sampler with a 30 min timestep, accordingly to the duration of the increasing of water level. A total of 10 to 18 samples (1L each) were collected under the surface.

All the samples were stored at 5 °C and manually mixed on a 25 L flask before the lab analysis.

Mean 24-hour samples were collected at La Villette basin point C (n = 10), the river Marne point VGA (n = 18), the stormwater outlet points SO-N (n = 9) and SO-S (n = 8) and the WWTP outlet WWTP-O (n = 5). A total of 414 samples was collected (Table 3.5).

Table 3.5. Summary of sample collection

Point	Sampling period	Sampling type	Total samples (dry + wet weather)	Sample preparation	
				Dilution factor	Filtration
Villette A	10/10/2020 to 08/03/2023	Grab samples	85 (40 + 45)	1	No
Villette B	10/10/2020 to 08/03/2023	Grab samples	85 (40 + 45)	1	No
Villette C	30/06/2021 to 22/09/2021	24-hour integrated	10 (2 + 8)	1	No
Villette C	22/06/2021 and 16/12/2021	Grab samples	5 (5 + 0)	1	No
Marne VGA	15/06/2022 to 20/10/2022	24-hour integrated	18 (4 + 14)	1	No
Marne SMV	06/07/2022 to 08/09/2022	Grab samples	189 (125 + 64)	1	No
SO-N	04/05/2022 to 09/09/2022	24-hour integrated 5 to 9-hour integrated*	9 (4 + 5)	1/5	Yes
SO-S	19/09/2022 to 07/11/2022	24-hour integrated	8 (4 + 4)	1/5	Yes
WWTP-O	30/06/2021 to 04/10/2021	24-hour integrated	5 (3 + 2)	1/5	No

*Two 5-hour integrated, one 6-hour integrated, one 7-hour integrated, and one 9-hour integrated.

3.3.1.2.1. Faecal indicator bacteria data

Faecal indicator bacteria (FIB) are represented by *Escherichia coli* (*E. coli*) and intestinal enterococci (IE).

Sample analysis included *E. coli* and IE measurement of the 24-hour integrated samples from Villette C (n=10), the Marne VGA (n=18), SO-N (n=9), SO-S (n=8) and WWTP-O (n=5), and the grab samples from the Marne SMV (n=144), at the same day of the collection of the samples. FIB were measured in a total of 194 samples.

The sampling flasks were disinfected the day before sampling. The flasks were left to soak in a solution of 0.1% of sodium hypochlorite for two hours, and then rinsed at least 10 times with distilled water.

The enumeration of *E. coli* of the 24-hour integrated samples was conducted by Leesu (*"Laboratoire Eau, Environnement et Systèmes Urbains"*, in Creteil). The reference ISO methods were applied for *E. coli* and IE measurements, being, respectively, "ISO 9308-3 (1999): miniaturized method (MPN) for detection and enumeration of *E. coli* in surface and waste water", and "ISO 7899-1 (1998): Water quality — Detection and enumeration of intestinal enterococci — Part 1: Miniaturized method (Most Probable Number) for surface and waste water". The results are expressed as the most probable number per 100 mL (MPN.100 mL^{-1}) in 48h.

Additional FIB data of La Villette basin was obtained from Ville de Paris:

- ColiMinder activity, as an indicator of the evolution of FIB concentration in La Villette canal, from 02/07/2019 to 03/10/2019.
- Fluidion *E. coli* measurements at point A, from summer 2019, 2020, 2021 and 2022.

3.3.2. Measurement of fluorescence intensity and absorbance

Fluorescence measurements were performed on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies) with a 10 mm quartz cell, at room temperature of the laboratory (around 20 °C). A simultaneous scan of excitation wavelength (ex) from 200 to 450 nm with 5-nm intervals and emission wavelength (em) from 250 to 550 nm with 2-nm intervals was performed, with a scan rate of 1200 $nm.min^{-1}$.

The absorbance was obtained with a Lambda 35 spectrophotometer (PerkinElmer). The samples were analysed in a 10 cm quartz cell and scanned from 200 to 750 nm with 1-nm intervals, at a scan speed of 240 $nm.min^{-1}$.

In both fluorescence and absorbance analysis, an ultra-pure water sample blank was analysed in the beginning of each series of measurements. The cell was tripled-rinsed with ultra-pure water and with the water sample between measurements.

3.4. Fluorescence data treatment

The processing of the Excitation-Emission Matrix (EEM) spectra was conducted according to Murphy et al. (2013), using the drEEM toolbox for MATLAB® R2022b (MathWorks®, USA). The absorbance was used for inner filter effect correction (Parker and Barnes, 1957).

The EEM processing steps, methods for assessing the position of fluorescence peaks of reference, computation of fluorescence indices, and the obtention of PARAFAC modes and components are described in the next paragraphs.

3.4.1. Preprocessing steps

The preprocessing steps of the dataset were divided in two phases. The first phase included the following steps (Murphy et al., 2010).

1) Conversion of raw data into compatible drEEM toolbox data

All the raw data must be adapted to be compatible with the used toolbox. Additionally, the excitation wavelength range of all EEMs were cropped to 250 to 450 nm to avoid the region where high absorbance was detected (between 200 and 250 nm).

2) Application of spectral correction factor

The spectral correction factors are applied to all EEMs, samples and blanks included, to correct instrument spectral biases (Equation 3.1). In this case, the equipment did not require a spectral correction, so the correction factor C was equal 1.

$$EEM_{CF} = EEM_{Raw} \cdot C \quad 3.1$$

3) Inner filter effect correction

The method used for the inner filter effect correction (IFC) is the absorbance method, based on the work of Parker and Barnes (1957) (Equation 3.2).

$$EEM_{IFE} = EEM_{CF} \cdot IFC \quad 3.2$$

The IFC matrix was calculated from the total absorbance matrix (A_{tot}), with the following expression (Equation 3.3):

$$IFC = 10^{\left(\frac{PL}{2} * A_{tot}\right)}, \text{with } A_{tot} = A_{ex} + A_{em} \quad 3.3$$

The pathlength (PL, in cm) is given by the size of the cell used for absorbance measurement. For a 10 mm-sized cuvette, PL = 1 cm. The total absorbance matrix (A_{tot}) is obtained from the sum of the absorbance at each couple of excitation (A_{ex}) and emission (A_{em}) fluorescence wavelengths from the EEM. This method is accurate within 5% for absorbance at 254 nm is inferior to 1.5 A.

4) Raman normalisation

The Raman normalisation is the conversion of the fluorescence intensities from Arbitrary Units (AU) to Raman units (RU). For that, the EEM intensities are divided by the Raman peak Area (A_{rp}). The A_{rp} corresponds to, at a given excitation wavelength, the integration of scatter peaks between defined emission wavelengths (λ_{em1} and λ_{em2}), which values are rounded to the nearest 0.5 nm (Equation 3.4). If the given values are not found, the program will increment λ_{em1} down and λ_{em2} up by 0.5 nm. The integration is done by linear interpolating the fluorescence intensities with incrementation of 0.5 nm of emission wavelengths, with trapezoidal approach of the area under the curve.

$$A_{rp}^{\lambda_{ex}} = \int_{\lambda_{em2}}^{\lambda_{em1}} I_{\lambda_{em}} d\lambda_{em} \quad 3.4$$

A_{rp} was calculated at λ_{ex} = 350 nm with an integration range of emission wavelength from 371 to 428 nm (Lawaetz and Stedmon, 2009). The calibrated EEM is given by Equation 3.5.

$$EEM_{cal,RU} = \frac{EEM_{IFE}}{A_{rp}^{350}} \quad 3.5$$

5) Blank subtraction

The blank subtraction step consists of subtracting the correspondent EEM blank (B) of the EEM of the sample, measured at the same day (Equation 3.6). Both EEMs were already converted in Raman units. This step contributes to remove Rayleigh and Raman scatter peaks and to correct instrument biases.

$$EEM_{Final} = EEM_{cal,RU} - B_{cal,RU} \quad 3.6$$

6) Correction for dilution if necessary

If the samples were diluted before fluorescence and absorbance scans, a correction should be applied. The corrected EEMs should be divided by their respective decimal dilution factor D_{fac} (Equation 3.7), e.g., 0.2 for a 5-times diluted sample.

$$EEM_{Final,dil} = \frac{EEM_{Final}}{D_{fac}} \qquad 3.7$$

The second phase of pre-treatment was composed by two steps.

1) EEM smoothing and complete removal of 1st and 2nd order Rayleigh and Raman scatters.

Rayleigh and Raman scatter bands (Figure 3.13) are reduced with blank subtraction, but remaining traces can affect PARAFAC modelling. The aim of this step is to remove them.

The scatter bands were removed without interpolation, being replaced with missing data (NaN values). The extents of scatter removals were defined based on the spectral variance and standard deviation of FDOM fluorescence of the dataset. The widths of scatter removal were set so no local peaks of variance or the standard deviation of fluorescence were observed next to Rayleigh and Raman scatter bands.

The reminiscent scatter signals were removed by the smoothing function, without interpolation. They were numerically replaced by NaN on the following extents:

- 1st order Rayleigh: 25 nm above, 30 nm below, in the emission axis.
- 1st order Raman: 10 nm above, 15 nm below, in the emission axis.
- 2nd order Rayleigh: 20 nm above, 18 nm below, in the emission axis.
- 2nd order Raman: 20 nm above, 14 nm below, in the emission axis.

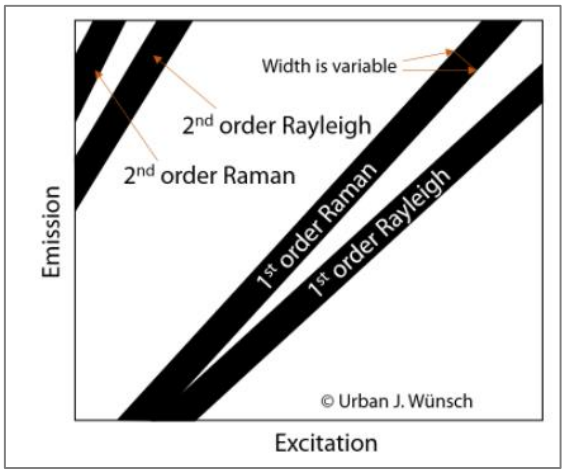


Figure 3.13. Scheme with the first and second order Rayleigh and Raman scatters (from Appendix A of Murphy et al., 2013).

2) Identification and removal of outliers

Firstly, the EEMs of all samples were viewed to identify potentially obvious outlier samples, such as accidentally included blanks, highly diluted samples which EEM are close to blank EEMs, samples with DOM concentrations high enough to reach the measurement range limit of the spectrophotometer.

The PARAFAC model was developed with a total dataset of 414 samples. 31 samples presented high residual values and were identified as outliers: 27 from La Villette basin, three (3) from the river Marne VGA and one (1) from SO-N. After their removal, the final

dataset contained 383 samples (n=158 La Villette basin, n=204 the Marne, n=8 SO-N, n=8 SO-S and n=5 WWTP-O).

Each EEM of the final dataset was then normalized to its total signal. In this part of the preprocessing phase in PARAFAC modelling, the weightings of samples with high and low concentrations become similar. Thus, the model should be able to identify chemical variations between samples independently of the magnitude of fluorescence intensity.

3.4.2. Fluorescence indices

The fluorescence intensities of the fluorophores from (Coble, 2007), and fluorescence indices used to describe DOM composition were obtained from the final corrected EEM, in RU ($EEM_{Final,dil}$).

The fluorescence intensities were linearly interpolated to a resolution of 1 nm in excitation and emission wavelengths to calculate and extract peaks and indices. Then, they were smoothed with the Savitzky-Golay method, with a span of 21 data points and polynomial of degree 2.

In addition, several spectral indices were computed and used to describe DOM composition. The biological index (BIX) was computed at an excitation wavelength of 310 nm (Equation 3.8), according to Huguet et al. (2009).

$$BIX_{ex\ 310} = \frac{I_{em\ 380}}{I_{em\ 430}} \qquad 3.8$$

Where $I_{em\ 380}$ and $I_{em\ 430}$ are the fluorescence intensities at 380 nm and 430 nm emission wavelength, respectively.

The humification index (HIX) was calculated at an excitation wavelength of 254 nm (Equation 3.9), according to Ohno (2002).

$$HIX_{ex\ 254} = \frac{\sum(Iem\ 435 \rightarrow 480)}{\sum(Iem\ 300 \rightarrow 345) + \sum(Iem\ 435 \rightarrow 480)} \qquad 3.9$$

Where $\sum(Iem\ 435 \rightarrow 480)$ is the sum of the fluorescence intensity between 435 nm and 480 nm of emission wavelength, and $\sum(Iem\ 300 \rightarrow 345)$ is the sum of the fluorescence intensity between 300 nm and 345 nm emission wavelength.

This index was compared with the HIX expression given by Zsolnay et al. (1999) (Equation 3.10).

$$HIX_{ex\ 254} = \frac{\sum(Iem\ 435 \rightarrow 480)}{\sum(Iem\ 300 \rightarrow 345)} \qquad 3.10$$

The fluorescence index (FI) was calculated as the ratio between the fluorescence intensity at excitation wavelength of 370 nm (Equation 3.11), according to Fellman et al. (2010).

$$FI_{ex\ 370} = \frac{I_{em\ 470}}{I_{em\ 520}} \qquad 3.11$$

Where $I_{em\ 470}$ and $I_{em\ 520}$ are the fluorescence intensities at 470 nm and 520 nm emission wavelength, respectively.

The positions of the fluorophores of interest (peaks T, M, C and A) and of the terms used to obtain the fluorescence indices BIX, HIX and FI on a fluorescence EEM are indicated in Figure 3.14.

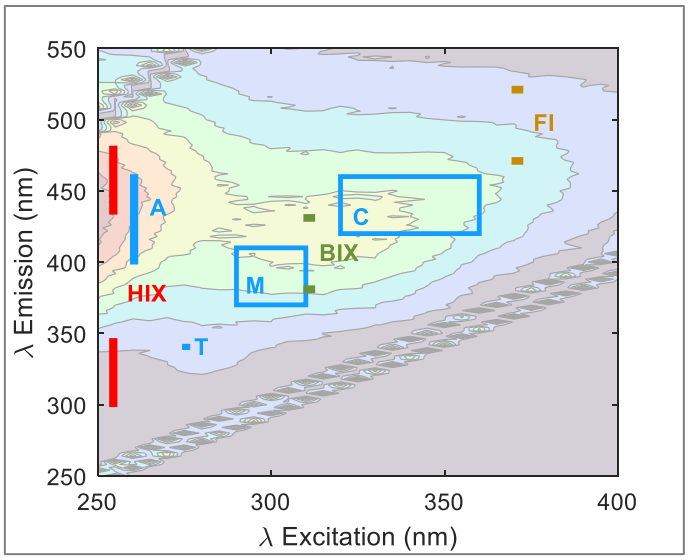


Figure 3.14. Typical Excitation-Emission Matrix (EEM). Location of the fluorescence peaks and fluorescence intensities used for indices computation are depicted in colour: blue for peaks T, M, C and A; red for HIX; green for BIX and brown for FI.

3.4.3. PARAFAC modelling

PARAFAC (PARAllel FACtor analysis) is a statistic tool applicable to data arranged in three (or more) arrays. The method is based on a generalization of Principal Component Analysis (PCA) to higher order arrays (Murphy et al., 2013). It can be used to decompose a EEM dataset into independent and distinct components that can be overlapped and combined to form the original dataset. The data signal is decomposed into a set of trilinear terms and a residual array, as in Equation 3.12.

$$x_{ijk} = \sum_{f=1}^{F} a_{if}\, b_{jf}\, c_{kf} + e_{ijk}, \text{ with i = 1, ..., } I\text{; j = 1, ..., } J\text{; k = 1, ..., } K\text{; f = 1, ..., } F \qquad 3.12$$

where x_{ijk} is the fluorescence intensity of the i^{th} sample at the j^{th} emission mode and at the k^{th} excitation mode; a_{if} is named "score" and is directly proportional to the concentration of the f^{th} analyte of sample i; b_{jf} is a scaled estimate of the emission spectrum of the f^{th} analyte at emission wavelength j; c_{kf} is linearly proportional to the specific absorption coefficient of the f^{th} analyte at excitation wavelength k; and e_{ijk} is the residual noise that represents the variability not accounted for by the model (Murphy et al., 2013).

The PARAFAC model was run and validated using drEEM toolbox in MATLAB® R2022b (MathWorks®, USA), following the recommendations of Murphy et al. (2013) and Stedmon and Bro (2008).

The region at excitation wavelengths below 250 nm was removed due to high levels of noise. Since negative concentrations and fluorescence intensities are chemically and physically impossible, a non-negativity constraint was applied to have realistic results.

The modelling in PARAFAC is composed by three main phases. Each phase can be revisited if needed.

1) Exploratory phase: outlier test and identification of most promising number of components

All EEMs of the dataset were normalised to its total signal, so the weightings of samples with high and low concentrations become similar. After EEM normalisation, the model should be

able to identify chemical variations between samples independently of the magnitude of fluorescence intensity.

A first exploratory test was done with 4 to 8-component models. In this phase of data analysis, it was possible to have a preliminary idea of how many components the PARAFAC model may contain and to identify additional potential outliers.

The models were initialized with SVD (singular value decomposition) and fit with a low convergence criterion of 10^{-2}. For each number of components, the best fit out of five initializations was showed.

The number of components was chosen accordingly to the percentage of explained variance, and the core consistencies of each preliminary model. Higher core consistence may protect against over-fitting, but not enough against under-fitting. In addition, the correlations between the components of each model were analysed. Models with high correlated components ($R^2 > 0.8$) were discarded.

The additional potential outliers may be identified by high values of sample leverage after EEM normalisation.

The leverage is a number between 0 and 1 and represents the deviation from the average data distribution. Other way to identify outliers was by the error residuals (error = data – model). Here, all sample leverages of the normalised final dataset were inferior to 0.1, with mean Sum Squared Residuals (SSR) under 20.

2) Model validation

PARAFAC models with F=4 to 8 components (see Equation 3.12) were fitted to find the best number of components. The 7-component PARAFAC model was validated through a split-half analysis, a validation with multiple split-half tests to confirm if the model is identically produced in independent subsample groups of the final dataset.

The validation method was an alternating '$S_4C_6T_3$'. This means that four split groups (S) were created. Each sample was alternately assigned to one of the four splits. The splits were assembled in six different combinations (C) to produce three split-half comparison tests (T).

3) Interpretation of the results

Finally, the obtained components were compared to the results in the OpenFluor database (http://openfluor.org). (Murphy et al., 2014). Each component was related to one type of DOM (humic-like, protein-like) and associated with components from the open database for qualitative comparisons.

The fluorescence intensity at maximum loadings of emission (b_{if}) and excitation (c_{if}) modes was used to quantify components in a sample, denoted C_{if}. For a given sample i and a component f, C_{if} is given by Equation 3.13.

$$C_{if} = \max_{jk}(a_{if}\ b_{jf}\ c_{kf}) \qquad 3.13$$

The relative percentage of each component f was calculated by dividing the correspondent Cf value by the sum of Cf of each sample i (Equation 3.14).

$$\%C_{if} = \frac{Cif}{\sum_{f=1}^{F} Cif}, \text{ with f = 1, ..., } F \qquad 3.14$$

Cf represents its correspondent PARAFAC component f. Hereafter, the component and its fluorescence intensity will both be denoted Cf.

3.4.4. Data analysis

The fluorescence data was analysed in five main steps. First, the fluorescence indices of all samples were compared. Second, the PARAFAC components of all samples were compared and associated with fluorophores from Coble (1996). Third, fluorescence indices from the river Marne samples collected at the same day were used to investigate the variation of DOM composition with land cover in the river Marne. Fourth, fluorescence indices from La Villette basin samples were compared to assess the temporal variation of DOM composition, comparing samples from dry and wet weathers. Fifth, the relationship between FIB and protein-like PARAFAC components was obtained through a linear regression model.

Differences among the studied sample groups were determined by a Kruskal-Wallis test, a non-parametric one-way ANOVA test. This multiple comparison test among more than two groups determinates which group differs from the others by using average ranks of the medians of the groups (Helsel et al., 2020). For La Villette basin samples, Kruskal-Wallis test was performed to investigate the variability of fluorescence indices between dry (n=85) and wet (n=98) weather samples.

Principal components analysis (PCA) was applied to reduce the dimensionality of fluorescence data (PARAFAC components and indices) and visualize the correlated variables among water types and among sampling sites.

3.5. Hydrodynamic modelling of La Villette system

This chapter presents the hydrodynamic TELEMAC-3D model of La Villette system. The software version used in the simulations was v8p3 for Linux.

3.5.1. Description of the model

TELEMAC-MASCARET modelling system was developed by Electricité de France (EDF) and partners. Nowadays, it is managed by a Consortium with four core organisations in addition to EDF (further information in: https://www.opentelemac.org/index.php/22-introduction/83-welcome-to-telemac). It is an open-source tool used in free-surface flows, such as river and maritime hydraulics. The modules are based in finite-element method. The simulated space is discretised in grids of triangular elements.

TELEMAC-3D is the three-dimensional (3D) hydrodynamics model of TELEMAC-MASCARET. The 3D mesh of the model consists of a series of layers between bed surface and water surface layers. The placement of the layers is flexible to adapt to the best vertical distribution of the layers according to the needs of the user.

The code solves 3D free surface Navier-Stokes equations including transport of active and passive tracers (Annex III). A detailed description of the mathematical and computational approach of the model can be found in TELEMAC-3D Theory guide (EDF R&D and Telemac3D consortium, 2021c) and User manual (EDF R&D and Telemac3D consortium, 2021a).

TELEMAC-3D has been coupled with the water quality module WAQTEL. One of the modules of WAQTEL is the thermic module, where the water-atmosphere heat exchanges are considered.

In the coupling, the hydrodynamic model TELEMAC-3D is responsible for computing the main physical processes, such as velocity field and tracer transport, and WAQTEL manages

the thermic process, computing the heat exchange with the atmosphere. Changes in water temperature affect the water density and, consequently, the hydrodynamics. In order to consider these effects of water temperature evolution, the equations from both models TELEMAC-3D and WAQTEL are computed at each time-step, evolving together in time and space.

3.5.2. Input files

TELEMAC-3D and WAQTEL require a number of input files. For a computation of La Villette system simulations, five files are required:

- The **steering file** (mandatory): contains the configuration of the simulation;
- The **geometry file** (mandatory): contains the information regarding the mesh;
- The **boundary conditions file** (mandatory): contains the type of each boundary point (solid or liquid);
- The **liquid boundaries file** (optional): contains the information on the prescribed values at the liquid boundaries;
- The **previous computation file** (optional): provides the initial state of the calculation in the case of a restart calculation;
- The **meteorological data file** (optional): contains the meteorological information for the computation of the water-atmosphere exchanges.

The **steering file** is an ASCII file which contains all the selection of computational options data. It serves as the computation dashboard, and it includes a set of keywords to which values are assigned. if a value is not defined, a default value will be assigned.

Each model requires its corresponding steering file, and software reads them at the beginning of the computation. In the model of La Villette system, two steering files are required: the TELEMAC-3D steering file, identified as the main steering file, and WAQTEL steering file. The configuration and parameters of both are presented in section 3.5.3.

The **geometry file** defines the computation domain. The geometry file contains the geometric data of the water body, that is the bathymetry and the horizontal shape of the system.

The geometry of La Villette system was defined from the bathymetry at 6 section profiles carried out by Fernando Bezerra in 2018 (Bezerra, 2019), with an echo sounder (798ci HD SI Combo) equipped with a sonar (700 Series™).

The Lambert coordinates of the geometric contour of La Villette system were obtained with QGIS (version 3.22.6) and the plug-in GMSH (version 2.16). The BlueKenue 64 (3.3.4 version) software was used to apply the bathymetry on the mesh and create the geometry file of the water body (Figure 3.15).

The geometry file contains the bottom topography and the computational mesh of La Villette system. The 2D grid of each layer is composed by 3932 nodes, 7289 horizontal triangles of 5 m side. The 3D mesh is composed of 7 horizontal parallel layers of 0.50 m depth. The layer 1 follows the bottom level and layer 7 follows the free surface. The method of horizontal planes is the most suitable method of discretization of the vertical axis to better represent temperature stratification zones (EDF R&D and Telemac3D consortium, 2021c).

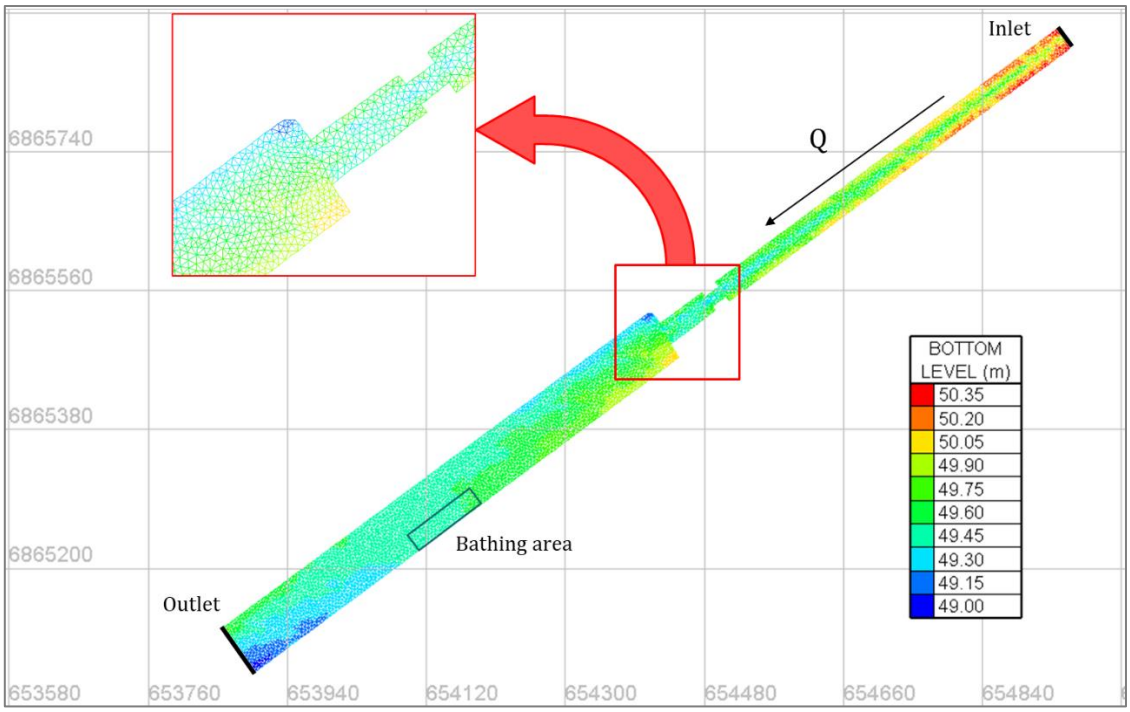


Figure 3.15. 2D grid of La Villette system of the bottom elevation, with the public bathing area indicated by the blue rectangle.

The **boundary conditions file** contains the type of each lateral boundary point, defined as liquid or solid. The points of the inlet and outlet boundaries (Figure 3.15) were defined as liquid boundaries, requesting the association of flow condition information. The side points of La Villette system were defined as solid walls. This file is generated by BlueKenue 64 (3.3.4 version) software.

Once the boundary conditions are set and defined as solid or liquid boundaries, the variables of the liquid boundaries need to be prescribed. The flow condition information should be declared in terms of water level, tracers and velocities or flow rate.

In this study, three tracers are computed: the water temperature, the electrical conductivity and an indicator of faecal contamination. The prescribed values should be declared on the liquid boundaries file, and the free values are obtained with the simulation results.

In La Villette system, the first liquid boundary corresponds to the inlet (upstream) and is defined as an open boundary with prescribed water level and tracer values. The second liquid boundary corresponds to the outlet (downstream) and is defined as an open boundary with prescribed flow rate and free tracer values.

The **liquid boundaries file** contains the timeseries of prescribed values of the liquid boundaries. TELEMAC-3D do not convert units, so the user must respect the standard unit. The indication of units on the file are only for user information (Figure 3.16).

The file must present the time of prescribed input values, declared in seconds elapsed from the start of the simulation. In La Villette model, the declared values are the outlet flow rate (in $m^3.s^{-1}$), and the inlet surface level (in m), and tracers. Here, the tracers are water temperature (in °C), electrical conductivity (in $\mu S.cm^{-1}$), and, when available, an indicator of faecal contamination (in $MPN.100mL^{-1}$).

These values should be declared with the following headers:

- T: time of prescribed data, in seconds. It should start with zero and be incremented according to the time step of the prescribed values. Here, the time step is of 1h (3600s);
- Q(2): flow rate at boundary 2 (outlet), in $m^3.s^{-1}$;
- SL(1): surface level at boundary 1 (inlet), in m;
- TR(1,1): tracer 1 at boundary 1 (inlet). Here, it corresponds to water temperature, in °C;
- TR(1,2): tracer 2 at boundary 1 (inlet). Here, it corresponds to the electrical conductivity, in $\mu S.cm^{-1}$;
- TR(1,3): tracer 3 at boundary 1 (inlet). Here, it corresponds to the indicator of faecal contamination, when available, in $MPN.100mL^{-1}$.

```
# QSL FILE
# start 2022-08-31 00:00:00
# end: 2022-09-23 23:00:00
# station: La_Villette
T           Q(2)     SL(1)    TR(1,1) TR(1,2)
sec         m3/s     m        degC    uS/cm
0.00        -2.18    52.43    19.75   620.67
3600.00     -2.24    52.47    21.72   612.00
7200.00     -2.27    52.48    21.67   612.00
10800.00    -2.28    52.49    21.61   612.67
14400.00    -2.29    52.50    21.56   612.33
18000.00    -2.29    52.49    21.52   613.00
```

Figure 3.16. Example of liquid boundaries file (QSL file)

The **previous computation file** is a result file of a previous simulation used to initialize a new computation. The last timestep recorded in this file defines the initial conditions of the new computation.

In La Villette model, the previous simulation is computed with liquid boundaries data from three days before the period to be simulated and does not consider the water-atmosphere heat exchange. It allows to start the new simulated period with a velocity field closer to the field conditions of La Villette system.

The **meteorological data file**, also identified as "ASCII atmospheric data file", contains the timeseries of the data measured in a meteorological station close to the study site.

The file must present the time of prescribed input values, declared in seconds elapsed from the start of the simulation (Figure 3.17). The required meteorological data are the wind magnitude (WINDS, in $m.s^{-1}$) and direction (WINDD, in degrees), air temperature (TAIR, in °C), atmospheric pressure (PATM, in hPa or mbar), cloud cover or nebulosity (CLDC, in octas), cumulated rainfall (RAINC, in mm), and relative humidity (HREL, in %).

For La Villette model, the meteorological data were obtained from the MeteoFrance station of Roissy (49.015278 N, 2.534444 E), at Charles de Gaulle airport. This is the closest station with all required data available, located 18 km northeast of the bathing area (Figure 3.18). All data were downloaded with an hourly time step.

```
# ASCII METEOFILE
# start 2022-08-31 00:00:00
# end: 2022-09-23 23:00:00
# station: 7157
T           WINDS    WINDD    TAIR     PATM     CLDC     RAINC    HREL
sec         m/s      deg      degC     hPa      octa     mm       %
0.00        16.00    360.00   20.20    1019.10  0.00     0.00     60.00
3600.00     15.00    10.00    19.40    1019.40  0.00     0.00     63.00
7200.00     17.00    360.00   18.60    1019.30  0.00     0.00     64.00
10800.00    19.00    10.00    17.60    1019.00  0.00     0.00     67.00
14400.00    17.00    10.00    16.60    1019.10  0.00     0.00     71.00
18000.00    13.00    360.00   15.90    1019.40  1.00     0.00     73.00
```

Figure 3.17. Example of meteorological data file (atmospheric data file).

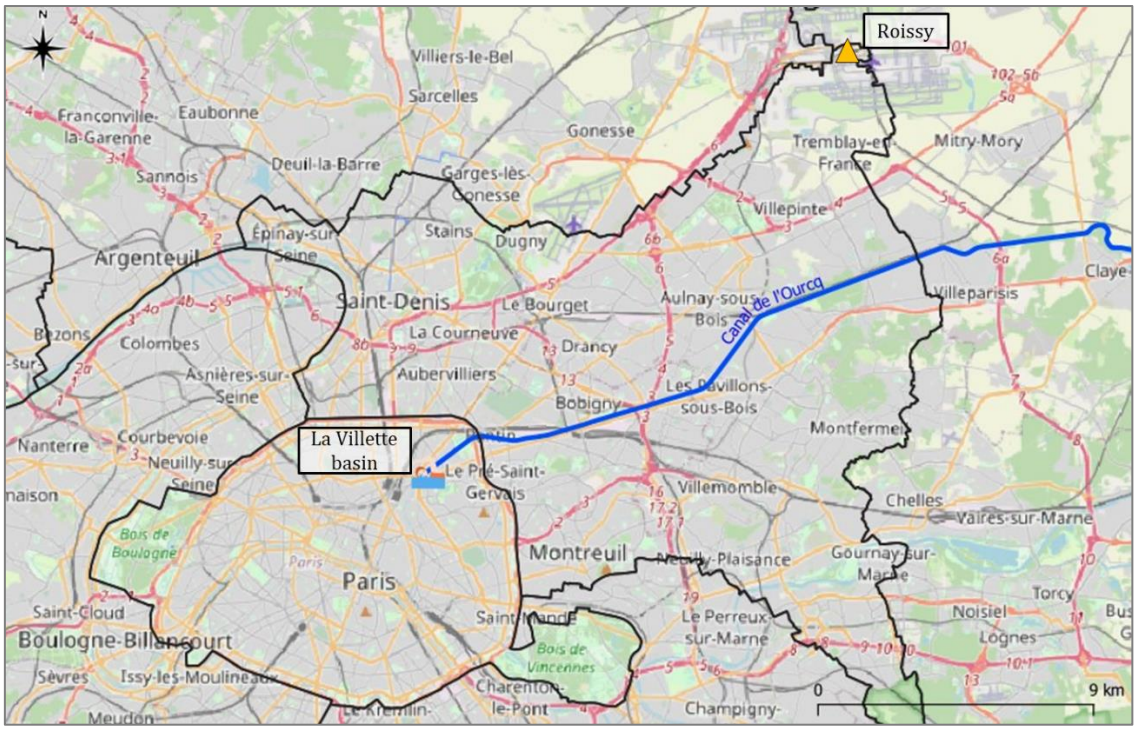


Figure 3.18. Location of Roissy airport meteorological station, 18 km from La Villette bathing area (base map from QGIS).

The input files must be declared in both steering files, and the WAQTEL steering file must be declared in the main steering file (Figure 3.19, black arrows). The output variables include the 3D water velocity, the water surface level, and the tracers of the liquid boundaries file (Figure 3.19, blue box).

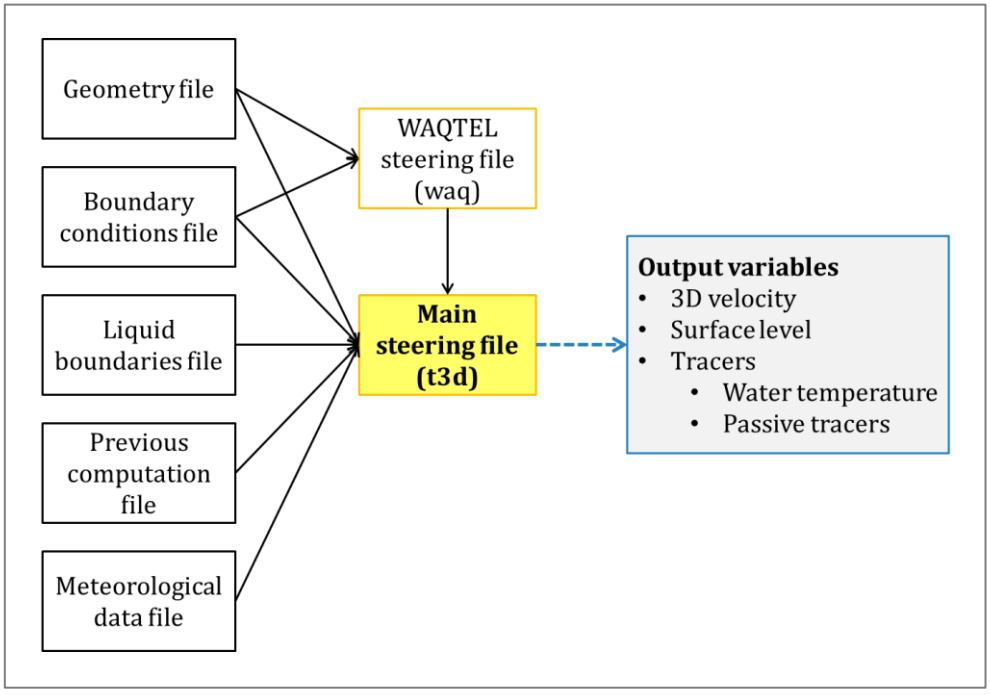


Figure 3.19. Scheme of input files for TELEMAC-3D modelling coupled with WAQTEL.

3.5.3. Modelling parameters

The installation of TELEMAC-3D comes with a folder of examples of the available modules of the software. The model set-up of La Villette system was based on the TELEMAC-3D examples of "canal" and "heat exchange" without wind and rain influences. The settings and the activated modules are described on the following paragraphs and all the information is from the Telemac3D User Manual (2018).

The variables of the 3D computation results are the water elevation, the velocities on the three axes (ENU axes), and two tracers.

3.5.3.1. Main configuration

3.5.3.1.1. Hydrodynamics configuration

The default initial condition sets the free surface water elevation, the water velocities, and the tracers equal to zero. In a first moment, to have an initial condition of water velocities field closer to the field data, a previous 3D computation file is considered.

The previous 3D computation file is a hydrodynamic pre-simulation of three days immediately before the period of interest to be simulated. It considers the inlet water elevation and the flow rate at outlet.

The initial conditions of flow rate and the tracers (water temperature and the passive tracer of each simulation) correspond to field data at the beginning of the simulated period.

Even with the pre-simulation file, some iterations were needed before stabilization of the tracers. We considered the first 6 iterations (i.e., the 1st hour) of the simulation as the stabilization period.

The boundary conditions consist of having prescribed water surface elevation (H) and tracer values at inlet (upstream), and prescribed flow rate (Q) and free tracer values at outlet (downstream). The outlet flow rate was calculated with the Manning-Strickler equation from the water level measured at upstream (detailed in section 4.4.1.2.2).

To consider the liquid boundaries file on the computation, the prescription of initial surface elevation and flow rates values at each boundary are required. The free surface elevation and the flow rate are initialized with, respectively, the measured and estimated values at the beginning of the simulated period. Since there is no tidal area on La Villette system, the setup "TIDAL FLATS" is deactivated.

The friction of the bottom follows the Strickler law. In La Villette system, during summer, the height of the macrophytes can achieve the water surface (Figure 3.20), configuring a build-up channel with highly presence of vegetation. The indicated friction coefficient for this configuration is $K_S = 2\ m^{1/3}.s^{-1}$ (Chow, 1959). The friction on the sidewalls follows the Nikuradse law, with friction coefficient $k_S = 0.02$ m, representing the size of asperities.

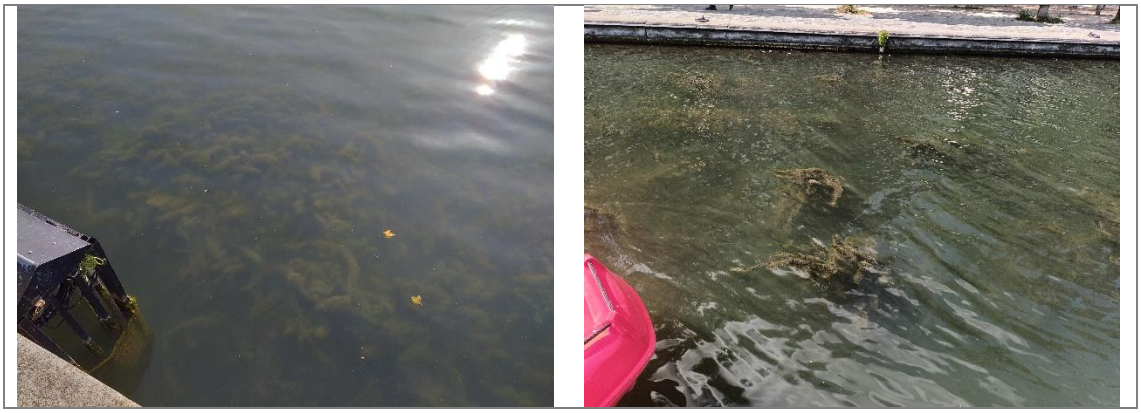


Figure 3.20. Macrophytes in La Villette system at upstream, point A (left), and downstream, point B (right).

The horizontal turbulence model uses a constant viscosity coefficient. The horizontal diffusion coefficient of velocities is set to 10^{-6} $m^2.s^{-1}$. The vertical turbulence follows the Nezu and Nakagawa mixing length model (Equation 3.15).

$$L_m = \kappa z \sqrt{1 - \frac{z}{h}} \qquad 3.15$$

With von Karman constant κ = 0.41, z the distance to the bed and h the water depth. The default value for the coefficient for vertical diffusion of velocities is of 10^{-6} $m^2.s^{-1}$ and the Viollet model set as damping function.

The advection scheme of velocities follows the method of characteristics. In this method, advection and diffusion steps are independent. Here, the value of one velocity is equal to its value in the previous instant traced back on the path taken during the time step. For the advection of depth, the software automatically selects the conservative scheme.

The mass-lumping rates for depth, diffusion, and velocities are equal to 1.0, resulting in diagonal mass matrices for those variables. This enables to shorten the computation time and results in smoothed solutions.

The propagation solver follows a conjugate gradient method, implying that the matrix of the system to solve is symmetric. The accuracy for propagation and for vertical velocity are both equal to the default value of 10^{-6}.

The total number of tracers is three: the water temperature , the electrical conductivity, and an indicator of faecal contamination (FIB). Because the water temperature affects the flow through the hydrostatic pressure gradient, it is considered an active tracer. The pressure is the sum of the hydrostatic pressure and the dynamic pressure. The hydrostatic pressure (p_h) is in function of the water density (Equation 3.16).

$$p_h = \rho g(Z_S - z) + \rho_0 g \int_z^{Z_S} \frac{\Delta\rho}{\rho_0} dz \qquad 3.16$$

Where ρ is the water density (in $kg.m^{-3}$), g is the acceleration due to gravity (in $m.s^{-2}$), Z_S is the free surface elevation (in m), z is the vertical space component (in m), ρ_0 is the average water density (1000 $kg.m^{-3}$ for freshwater), and $\Delta\rho$ the variation of water density around the reference density.

Therefore, the selected density law of the water considers a variation of ρ according to the temperature (Equation 3.17).

$$\rho = \rho_{ref}\left[1 - T_w\left(T - T_{ref}\right)^2 10^{-6}\right] \qquad 3.17$$

Where ρ_{ref} is the reference density of water (999.972 kg.m^{-3}) at the reference water temperature T_{ref} = 4 °C, and T_w the water temperature (in °C).

The electrical conductivity and the indicator of faecal contamination are considered passive tracers.

The initial and prescribed values of tracers were defined as the respective first values of the input liquid boundary file of the simulated period, and the value of atmospheric pressure is 100 000 Pa.

The coefficient for vertical diffusion of the tracers is 10^{-4} m^2.s^{-1}, and the maximum number of iterations for diffusion is 1000.

The water-atmosphere exchanges include the wind effect, variable in time and constant in space, and take into account the atmospheric pressure and the influence of precipitation or evaporation.

3.5.3.1.2. Water quality models: WAQTEL module

The model is coupled with the water quality module WAQTEL. To consider the heat exchange balance with the atmosphere, the thermic process module is activated, considering the complete balance of exchanged heat fluxes that affects the energy balance. These fluxes (in W.m^{-2}) are illustrated in Figure 3.21 and are the following:

- Solar radiation (RS)
- Atmospheric radiation (RA)
- Water surface radiation (RE)
- Sensible heat or Conductive heat (CV)
- Latent heat, or evaporation (CE)

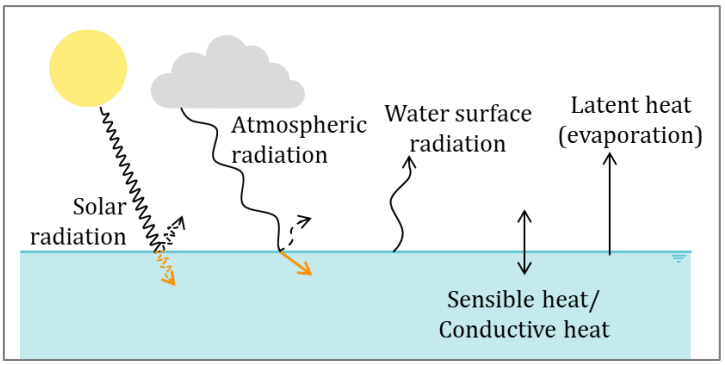


Figure 3.21. Scheme with heat fluxes between atmosphere and water surface (redrawn from Piccolroaz et al. (2013)).

The complete balance of heat fluxes is calculated at the free surface for water temperature (Equation 3.18).

$$K \frac{\partial T}{\partial z}\Big|_{z=\eta} = \frac{RA - RE - CE - CV}{\rho_{water} C_P} \qquad 3.18$$

Where K is the coefficient of molecular diffusion in water (in m^2.s^{-1}), T is the water temperature (in °C), z is the distance from the bed, η is the free-surface elevation, ρ_{water} is the water density (kg.m^{-3}) and C_P is the specific heat of water (4180 J kg^{-1}.°C^{-1}).

The penetration of solar radiation in the water column depends on the water turbidity. It is considered in the source term of the advection-diffusion equation of the water temperature (S) (Equation 3.19).

$$S = \frac{1}{\rho_{water} C_P} \frac{\partial Q(z, RS)}{\partial z} \quad 3.19$$

Where Q(z, RS) is the residual RS at z.

The solar radiation (RS) depends on time and location of the site and the cloud coverage (Equation 3.20).

$$RS = AA.\sin(ang)^{BB}\,(1 - 0.65C^2)(1 - Alb) \quad 3.20$$

Where:

- AA (W.m^{-2}) and BB (dimensionless) are related to luminosity and sky colour. These coefficients are chosen with respect of the considered area. The default values are for a mean pure sky, with AA = 1080 W.m^{-2} and BB = 1.22.
- ang (rad) is the angular height of the Sun. It depends on the geographical position of the site and changes with day and hour.
- C is the nebulosity, in tenths;
- Alb is the water albedo for short waves that may vary every month.

The residual solar radiation (W.m^{-2}) is calculated from *in situ* measurements of Secchi depth (SD, in meters) and is written from Beer-Lambert's law (Equation 3.21).

$$Q(z, RS) = RS.\exp\left(-\frac{1.7(z_s - z)}{SD}\right) \quad 3.21$$

Where z_s (m) is the free surface elevation.

The atmospheric radiation (RA) is the long wave radiation emitted by the atmosphere like a black body. It corresponds to the reemission of a part of direct solar energy and are determined by cloud coverage and albedo at the free surface (Equation 3.22).

$$RA = (1 - alb_{lw})\varepsilon_{air}\sigma(T_{air} + 273.15)^4(1 + Nua.C^2) \quad 3.22$$

Where:

- alb_{lw} is the albedo for long radiative waves and is equals to 0.03;
- ε_{air} is the air emissivity, and $\varepsilon_{air} = 0.937 \cdot 10^{-5}(T_{air} + 273.15)^2$.
- σ is the Stefan-Boltzmann's constant, equals to $5.67 \cdot 10^{-8}$ W.m^{-2}.K^{-4};
- C is the nebulosity, in tenths;
- Nua (dimensionless) represents the type of the cloud. It ranges from 0.04 (Cirrus cloud) to 0.24 (Stratus cloud). A mean value of 0.17 (Alto Cumulus cloud) is used.

The radiation emitted from a water body (RE) is calculated considering that the water surface behaves like a grey body (Equation 3.23).

$$RE = \varepsilon_{water}\sigma\left(T_{surface} + 273.15\right)^4 \quad 3.23$$

With $T_{surface}$ as the water temperature at the surface and ε_{water} = 0.85 as the value for water emissivity in La Villette system. The water emissivity is given with the keywords COEFFICIENTS FOR CALIBRATING SURFACE WATER RADIATION and depends on the location and obstacles surrounding the water body.

The latent heat flux (CE) comes from evaporation and is calculated with Equation 3.24.

$$CE = \rho_{air} Le\left(T_{surface}\right) f(u_2)(HS_{water} - HS_{air}) \quad 3.24$$

With:

- ρ_{air} is the air density, given by $\rho_{air} = \frac{100\ P_{atm}}{(T_{air}+273.15).287}$, where P_{atm} is the atmospheric pressure in hPa;
- $Le(T_{surface})$ ($J.kg^{-1}$) is the latent heat of evaporation at the temperature at the surface. $Le(T_{surface}) = (2500.9 - 2.365\ T_{surface}).10^3$;
- $f(u_2)$ is a function of the wind velocity at 2 m high (u_2, in $m.s^{-1}$). In 3D, $f(u_2) = b(1+u_2)$. In La Villette system, $b = 0.0028$;
- a and b are empirical coefficients to be calibrated. Their values are very close, and the default is 0.0025;
- HS_{water} is the specific humidity of the saturated air at the surface temperature;
- HS_{air} is the specific humidity of air.

Finally, the sensitive heat flux (CV) corresponds to the heat transfer by convection and is given by Equation 3.25.

$$CV = \rho_{air} C_{P_{air}} f(u_2)\left(T_{surface} - T_{air}\right) \qquad 3.25$$

Where $C_{P_{air}}$ is the air specific heat equals 1005 $J.kg^{-1}.°C^{-1}$.

All the expressions and values were presented on the WAQTEL and TELEMAC-3D Theory Guide v8p3 (EDF R&D and Telemac3D consortium, 2021c).

The exchanges with lateral boundaries and with the bed are neglected. The effect of the wind, atmospheric pressure and influence of rain or evaporation are taken into account.

The Secchi depth was defined accordingly to the mean of the field measurements taken during the simulated or at its closest date.

The summary of the configurations and parameters of the modelling is presented on Table 3.6.

Table 3.6. TELEMAC-3D modelling setup information synthesis

Characteristic	Description	Values
TELEMAC-3D version	Version for Linux	v8p3
Input data		
Geometry and bottom topography	Bathymetry by Fernando Bezerra (2018).	-
Mesh **Layers** **Thickness**	Horizontal triangles	5 m side 7 layers 0.50 m depth.
Boundary conditions	Upstream (inlet) Downstream (outlet) Side points	Prescribed water surface level (H) and tracer. Prescribed flow rate (Q). No flow condition information.
Flow rate (m^3s^{-1})	Assumed to be the same at inlet and outlet.	Manning-Strickler equation, from the measured water level upstream.
Water temperature	Timeseries of water temperature at inlet of La Villette system.	*In situ* measurements at mid-depth at upstream (point A)
Meteorological conditions	Timeseries of wind magnitude and direction, air temperature, air pressure, relative humidity, nebulosity, and precipitation.	Data from Roissy Airport station

Characteristic	Description	Values
Modelling parameters		
3D simulation results	Water elevation, velocities, water temperature and passive tracers (conductivity, indicator of faecal contamination).	
Prescribed initial conditions	Free surface with constant elevation. Need to declare initial surface elevation and flow rate to activate the reading of the liquid boundaries file.	Last values of the 3 days pre-simulation.
Tidal flats option	Deactivated (no tidal area).	-
Bottom friction	Strickler law	K_s = 2 $m^{1/3}.s^{-1}$
Sidewalls friction	Nikuradse law	k_S = 0.02 m
Horizontal turbulence	Constant viscosity coefficient (default). Coefficient for diffusion of velocities =	 10^{-6} $m^2.s^{-1}$
Vertical turbulence	Nezu and Nakagawa mixing length model. Damping function of Viollet. Coefficient for diffusion of velocities =	 10^{-6} $m^2.s^{-1}$
Number of tracers	An active tracer (water temperature), and two passive tracers (conductivity and indicator of faecal contamination).	3
Density law	Variation of density according to the temperature.	-
Prescribed tracer values	Initial tracer values are the same as the liquid boundaries file after pre-simulation of 3 days, when T=262800 sec.	Values in liquid boundaries file when T=262800 sec.
Atmospheric pressure		100 000 Pa
Coefficient for vertical diffusion of tracers		10^{-4} $m^2.s^{-1}$
Maximum number of iterations for diffusion		1000
Coupled with	WAQTEL, the water quality module.	
Water quality process	Thermal module. Only heat exchange with atmosphere is considered (with bed and lateral boundaries are neglected).	
Water-atmosphere exchanges	Wind effect, air pressure, and influence of rain or evaporation are taken into account.	
WAQTEL steering file	The Secchi depth is the mean field measurements of the simulated period, when available. Otherwise, the default value is adopted. Lightness of the sky: mean pure sky. Water emissivity ε_{water} = Coeff. To calibrate wind velocity b =	Secchi depth (default) = 0.9 m 0.85 0.0028.

3.5.3.2. First and final configuration

The model configuration evolved during the development of this study. The initial configuration was developed before the installation of the measuring instrumentation. The final configuration was established with the validation of the model. A comparison between the first and the final configuration is presented in this section and summarized in Table 3.7.

The availability of water level at upstream enabled the modification of the boundary conditions of the original configuration. The previous conditions considered a prescribed flow rate and tracer at inlet (upstream) and prescribed constant surface level at outlet (downstream). The new conditions consider a prescribed surface level and tracer at inlet, and a prescribed flow rate and free tracer at outlet.

To optimize the computation time of the simulation, a sensitivity analysis of the mesh resolution and the number of horizontal layers was conducted by Maria Júlia Costa Medeiros (Costa Medeiros, 2021). Grids with elements of 4 m, 5 m, 6 m and 1.2 m size were compared. The results showed that the mesh with 5 m-sized elements was the minimum resolution necessary to assure satisfactory results.

Simulations with 3, 5, 7 and 10 horizontal layers were compared. No significant differences were found among the simulation results of the previous configuration, with 10 layers, and with 7 layers. Therefore, a number of 7 horizontal layers was adopted for the modified configuration to reduce computation time.

In the initial configuration, the bottom friction coefficient was 60 $m^{1/3}.s^{-1}$, the default value for Strickler law, and no friction on the sidewall was considered. In the modified configuration, the bottom friction coefficient is 2 $m^{1/3}.s^{-1}$, representing the elevated presence of vegetation in the bottom, and the Nikuradse law was applied for the friction on the sidewalls, with friction coefficient of 0.02 m.

On the heat exchanges between wate and atmosphere of the initial configuration, no wind effect, atmospheric pressure nor influence of precipitation or evaporation were taken into account. In the modified configuration, they were all considered. In addition, the water emissivity for surface water radiation was 0.92 in the first configuration. After calibration, it is 0.85 in the modified configuration.

These modifications were made with the contribution of two interns of Leesu, Maria Júlia Costa Medeiros (September to December 2021) and Arthur Guillot – Le Goff (February to July 2022).

Table 3.7. Summary of modification in the model configuration

Parameter	Initial configuration	Final configuration
Boundary 1	Outlet (downstream). Open boundary with prescribed H (constant) and free tracer (variable).	Inlet (upstream). Open boundary with prescribed H (variable) and tracer (variable).
Boundary 2	Inlet (upstream). Open boundary with prescribed Q (variable) and free tracer (variable).	Outlet (downstream). Open boundary with prescribed Q (variable) and free tracer (variable).
Size of mesh element	5 m	5 m
Number of horizontal layers	10	7

Parameter	Initial configuration	Final configuration
Law of friction and friction coefficient for the bottom	Strickler law K_S = 60 $m^{1/3}.s^{-1}$ (default)	Strickler law K_S = 2 $m^{1/3}.s^{-1}$
Law of friction and friction coefficient for sidewalls	No friction.	Nikuradse law. k_S = 0.02 m
Water-atmosphere exchanges (wind, atmospheric pressure and rain or evaporation effects)	No wind effect, no atmospheric pressure and no influence of precipitation or evaporation taken into account.	Wind variable in time, atmospheric pression and influence of rain of evaporation taken into account.
WAQTEL calibration coefficients ***Surface water radiation*** ***Wind velocity***	 Water emissivity ε_{water}=0.92 -	 Water emissivity ε_{water}=0.85 b= 0.0028

Chapter 4. Results

This chapter presents the main results of this PhD thesis. It is divided into four main topics: (1) the implementation of the field monitoring system and the collected dataset; (2) the dissolved organic matter characteristics of urban waters from fluorescence analysis; (3) the estimation of faecal indicator bacteria from fluorescent dissolved organic matter; (4) the hydro-microbiological modelling outcomes.

4.1. Field monitoring system

This section describes the monitoring system designed and installed on the La Villette study site, at points A (upstream) and B (downstream). The trends and patterns of the data collected since August 2020 are presented.

The monitoring system was implemented in two steps. From August 2020 to June 2021, a Basic monitoring system was deployed. From May 2021 onwards, a more complete monitoring system was implemented. From May to June 2021, simultaneous measurements from both systems were recorded to check the data consistency between both systems. Regular field campaigns were conducted at least once a month, for the maintenance of the sensors and to validate field sensor data.

The data collected during 32 months, from August 2020 to March 2023, are explored to illustrate:

(i) *in situ* validation of field data by comparing the sensor data with measurements from the reference probes;
(ii) the common types of data alteration, resulting in a non-validation of field measurements;
(iii) the possible corrections of field data, particularly of water level and conductivity;
(iv) the seasonal and daily thermal regimes of La Villette basin;
(v) the evolution in time and space of conductivity, with an application to estimate the mean transfer time between upstream and downstream points.

4.1.1. Basic version of the monitoring system

This first version of the monitoring system was deployed from August 2020 to June 2021.

4.1.1.1. Measuring *in situ* equipment

At point A, from 07/08/2020 to 10/06/2021, a conductivity and a temperature sensor (NKE STPS 150 – PR, nke instrumentation) was installed within the cage of the microbiological measuring station (see Figure 3.2). This multiparameter sensor was fixed at the same depth as the microbiological probe, at 0.25 m from the surface (Figure 4.1-a). The measurement timestep was of 5 min.

Downstream, at point B_0, a sensor line was installed from 01/08/2020 to 03/09/2020. It was moved to point B (see Figure 3.3) on 11/09/2020. This line was composed by temperature sensors (NKE SP2T, nke instrumentation) located at three depths: 0.50 m, 1.20 m, and 2.00 m (Figure 4.1-b). The measurement timestep is of 10 min.

The measurements were manually recovered twice a month, during the field campaigns. This on-site data transfer was carried out via a radio data pencil (point A) and an induction

data pencil (point B) to a portable computer. The characteristics of the sensors installed upstream and downstream are presented in Table 4.1.

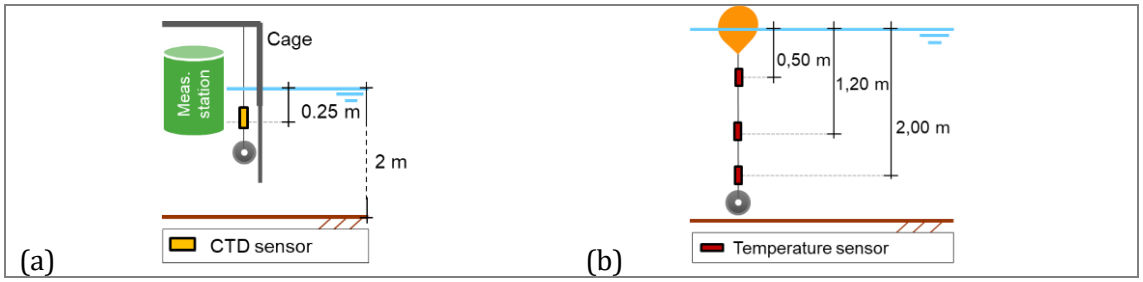


Figure 4.1. First basic configuration: (a) CTD sensor at point A and (b) temperature sensor at point B.

Table 4.1. Characteristics of the *in situ* devices of the basic monitoring system.

Equipment	CTD sensor	Temperature sensor
Model	NKE STPS 150 – PR (radio)	NKE SP2T (induction)
Fabricant	nke instrumentation	nke instrumentation
Serial number	ID: 35013. Radio number: 15172 (0.25 m)	Surface: 32007 (0.50 m) Middle: 32009 (1.20 m) Bottom: 28076 (2.00 m)
Point of measure	A (upstream)	B_0 (bathing area platform) B (Marin d'Eau Douce platform)
Date of installation	07/08/2020 to 10/06/2021	Point B_0: from 01/08/2020 to 03/09/2020 Point B: 11/09/2020 to 20/05/2021
Parameters	Salinity*, temperature, depth**	Temperature, depth**
Timestep	5 min	10 min
Range	Salinity*: 2 to 42 $g.L^{-1}$ at 20°C Temperature: -5 to +35°C	Temperature: -5 to +35 °C
Precision	Salinity*: 0.01 $g.L^{-1}$ Temperature: 0.05 °C	Temperature: In range 0 +20°C :0.05 °C Out of this range: 0.1 °C

* The salinity measurements (in $g.L^{-1}$) were converted to conductivity at 25°C (in $\mu S.cm^{-1}$) based on the Practical Salinity Scale of 1978 (Lewis, 1982). Equations in Annex IV-a.
** Out of service. Parameter not considered.

4.1.1.2. Collected data from August 2020 to June 2021

The data set obtained from August 2020 to June 2021 was validated against field data. Profiler probes of reference, SeaBird (SBE) and WiMo (nke), were used to validate the continuous measurements and to obtain complete vertical profiles (Annex IV-b).

The range of conductivity was between 600 and 900 $\mu S.cm^{-1}$ (Figure 4.2). From 07/08/2020 to 27/09/2020 the measurements were between 600 and 650 $\mu S.cm^{-1}$. Then, the values increased almost continuously until the 04/11/2020, where it had stabilized between 800 and 900 $\mu S.cm^{-1}$. Since the end of December 2020, large variations between 600 and 900 $\mu S.cm^{-1}$ were measured.

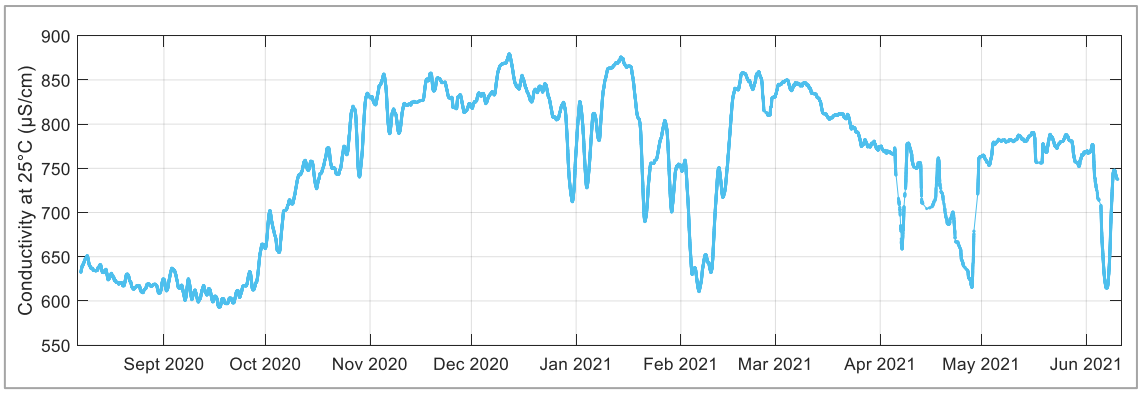


Figure 4.2. Conductivity at point A from 07/08/2020 to 10/06/2021.

Upstream, the water temperature has been measured since 01/08/2020. The data range was [0.8-26.7] °C. The highest value, 26.7 °C, was measured on 12/08/2020 (Figure 4.3-top).

Downstream, from 01/08/2020 to 03/09/2020, the water temperature was measured at point B_0, immediately upstream the bathing area, on the left bank. The range of the water temperature was of [18.5-26.9] °C (Figure 4.4). Daily stratification was observed between 05-20/08/2020. The water column mixes every night.

Since 11/09/2020, the water sensors have been moved to point B, on the opposite bank of the bathing area (see Figure 3.3). Between September 2020 to June 2021, the data range was of [1.1-21.9] °C. The minimum value (1.1 °C) was measured on the 14/02/2021 and the maximum value (21.9 °C), on 16/09/2020 (Figure 4.3).

Figure 4.3. Water temperature at points A (a) and B (b) from 01/08/2020 to 20/05/2021.

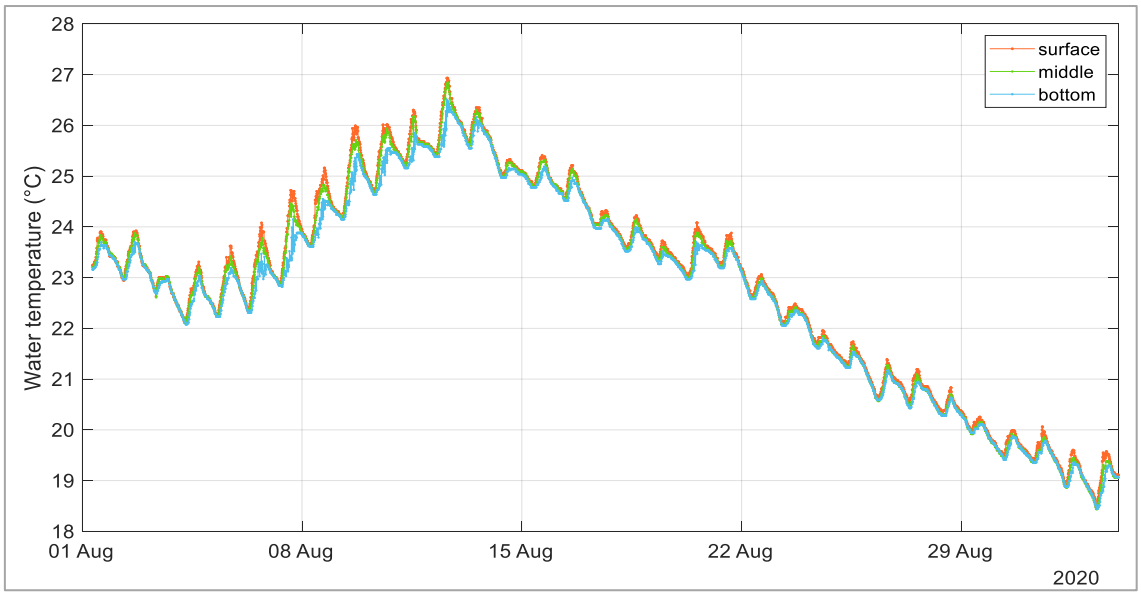


Figure 4.4. Point B_0: water temperature at surface (0.50m), middle (1.20m) and bottom (2.00m) depths, from 01/08/2020 to 03/09/2020.

Temperature stratification of the water column at point B was observed on late September 2020. Temperature differences from 0.4 to 1.4 °C between the surface and the bottom were measured during the day. From 23 to 29/09/2020, a drop of the water temperature erased the stratification, which reappeared at lower temperatures (around 15 °C) at the end of September (Figure 4.5).

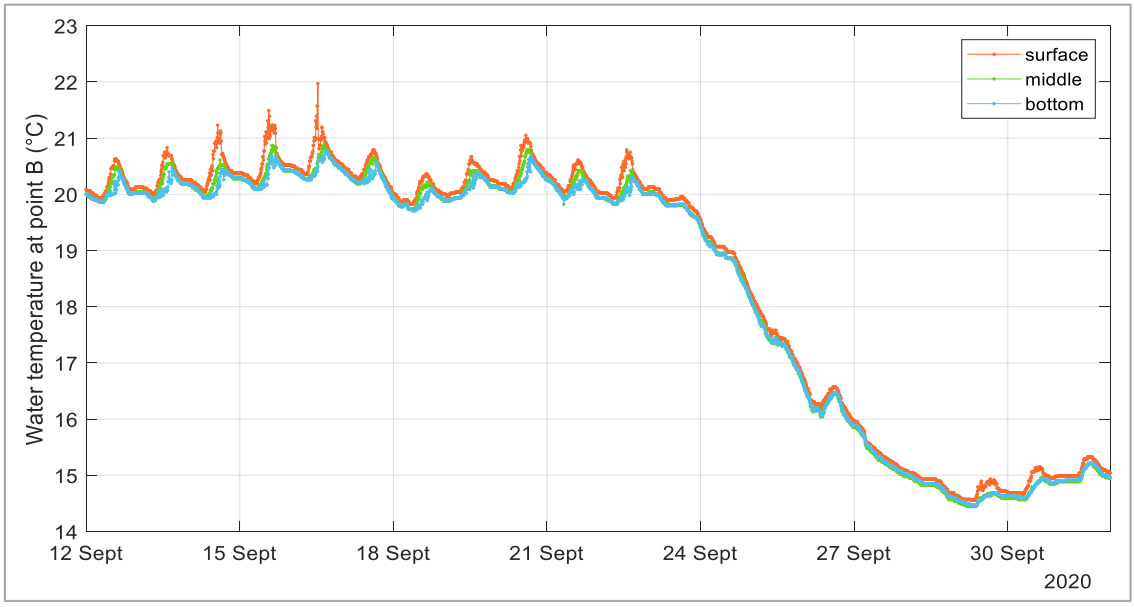


Figure 4.5. Water temperature at point B, from 12/09/2020 to 02/10/2020.

4.1.2. Design of the complete monitoring system

Since May 2021, a complete version of the autonomous monitoring system was installed. The data were automatically transmitted and sent daily to a database on an internal server at Leesu-ENPC. Similar systems were installed at points A (upstream) and B (downstream).

4.1.2.1. Measuring equipment

4.1.2.1.1. In situ equipment

At each monitoring point, two physical-chemical variables were measured: temperature, conductivity (specific conductivity at 25 °C). In addition, turbidity and water level were measured at point A, upstream. The temperature sensor CS225 (Campbell Scientific®) was used for measuring the water temperature at three depths: at point A, 0.25 m, 0.95 m, 1.75 m and at point B, 0.50 m, 1.20 m, 2.00 m (Figure 4.7).

At point A, the multiparameter probe MS5 (OTT®) was used for measuring conductivity, water temperature and turbidity at 0.95 m. A water pressure sensor CP5240 3mCE (Hitec®) was used to measure the water level (Figure 4.7-a). At point B, the multiparameter probe DS5 (OTT®) was used for measuring conductivity and water temperature at 1.20 m (Figure 4.7-b). The characteristics of all measuring devices are presented in Table 4.2. A synoptic table of the sensor measurements is presented in Figure 4.6.

Table 4.2. Characteristics of *in situ* measuring devices

Equipment	Multiparameter sensor	Multiparameter sensor	Temperature sensor	Water level sensor
Fabricant	OTT	OTT	Campbell Scientific	Hitec
Model	OTT MS5	OTT DS5	CS225	CP5240 3mCE
Monitoring point	A (upstream)	B (bathing area)	A (upstream) B (bathing area)	A (upstream)
Date of installation	20/05/2021	05/05/2021	05/05/2021	05/05/2021
Parameters	Conductivity, temperature, turbidity	Conductivity, temperature,	Temperature	Air pressure, water pressure
Timestep	20 min	20 min	20 min	20 min
Range	Conductivity: 0 to 100 $mS.cm^{-1}$ Temperature: -5 to 50 °C Turbidity: 0 to 3000 NTU	Conductivity: 0 to 100 $mS.cm^{-1}$ Temperature: -5 to 50 °C	-55 to +85°C	0.5 to 300 m
Resolution	Cond.: 0.001 $mS.cm^{-1}$ Temp.: 0.01 °C Turbidity: 0.1 NTU	Conductivity: 0.001 $mS.cm^{-1}$ Temperature: 0.01 °C	0.01 °C	0.001 m
Precision	Conductivity: ±1% of reading; ±0.001 $mS.cm^{-1}$ Temp.: ±0.10°C Turb.: ±1% up to 100 NTU, ±3% from 100–400 NTU, ±5% from 400–3000 NTU	Conductivity: ±1% of reading; ±0.001 $mS.cm^{-1}$ Temperature: ±0.10°C	Typical: ±0.2°C (-40° to +85°C), includes lifetime drift Worst Case: ±0.4°C (-40 to +85°C), includes lifetime drift	±0.05% of full scale

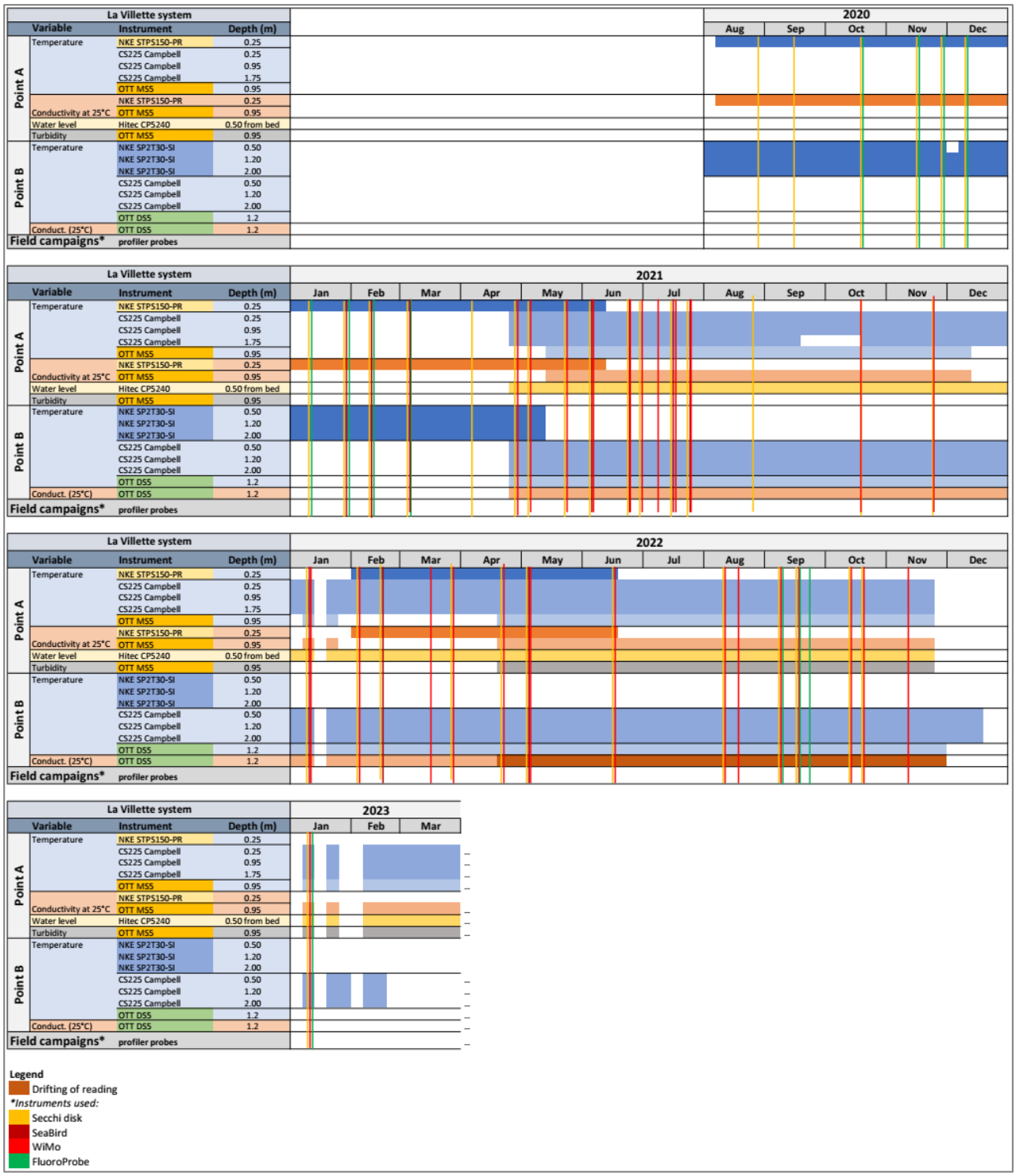


Figure 4.6. Synoptic table of La Villette monitoring system, from August 2020 to March 2023. Data used in this thesis.

4.1.2.1.2. Regular field campaigns

Regular campaigns were performed at points A and B for validation of sensor data and for cleaning and maintenance of the instruments. The follow-up of the field campaigns is presented in Annex IV-c. In addition, a profiler probe was used to obtain a vertical profile of the water column. Water temperature and conductivity profiles were measured with a WiMo (nke Instrumentation®) probe. Water transparency was measured with a Secchi disk (Table 3.4, section 3.2.2). All the measured profiles are presented in Annex IV-d.

4.1.2.2. Implementation of the monitoring system

At point A, all the instruments were mounted directly on the monitoring station container, a non-floating point. At 0.50 m from the canal bed, a water pressure sensor was installed to measure the variation of water level (Figure 4.7-a). At point B, the instruments were attached to a floating buoy (Figure 4.7-b).

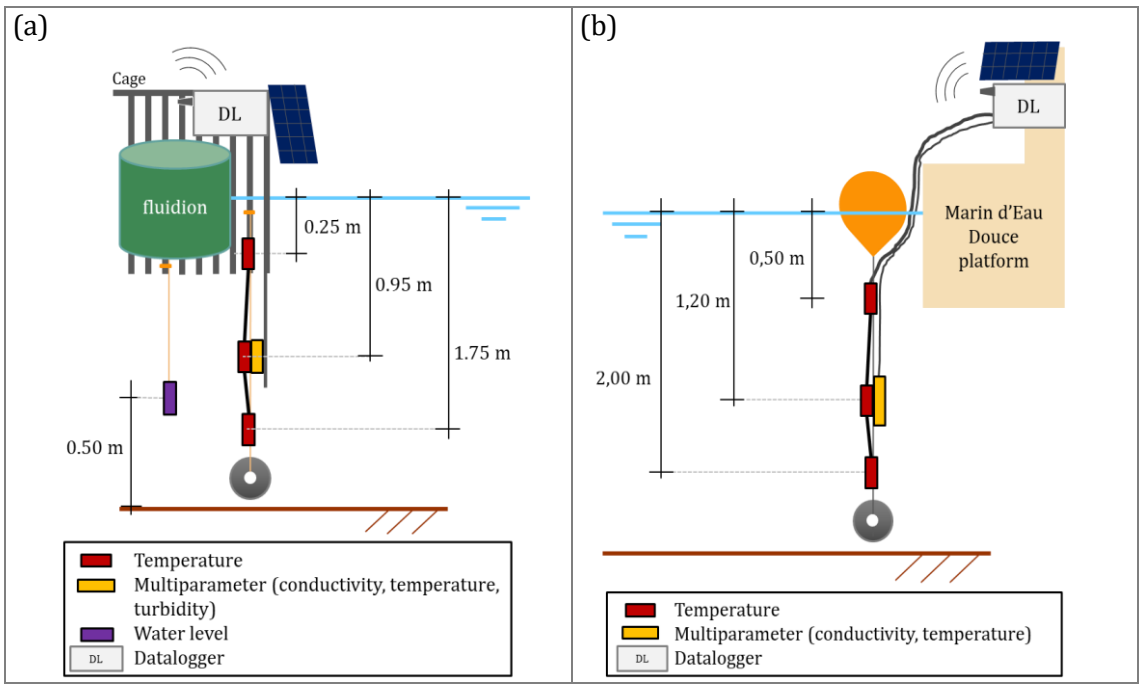


Figure 4.7. Configuration of monitoring system of: (a) conductivity, turbidity, temperature, and water level sensors at point A; and (b) conductivity and temperature sensors at point B.

For the remote data transmission, the system was equipped with a datalogger (CR300 Campbell scientific®), a PMW (Pulse Width Modulation) controller (BlueSolar PMW Light Charge Controller, Victron energy®), a batterie (AGM 12V 14Ah, Victron energy®), a photovoltaic panel (PVRPC1201016), a surge protector (SGB3, Campbell scientific®), an antenna (RM Series, MobileMark®), and a SIM card with GPRS (General Packet Radio Service) internet subscription (Table 4.3).

Table 4.3. Components of the devices for the data transmission of the autonomous measuring station.

Component	Reference	Fabricant
Datalogger	CR300	Campbell scientific
PMW controller	BlueSolar PMW Light Charge Controller	Victron energy
Battery	AGM 12V 14Ah	Victron energy
Photovoltaic panel	PVRPC1201016	-
Surge protector	SGB3	Campbell scientific
Antenna	RM Series	MobileMark
SIM card	M2M subscription of 5 Mb/month	SFR

Through a Python script, the datalogger triggers the sensor devices to start measuring, it converts the unit of the measurements, it stores the data in its memory, and it transmits data by email. In addition to the measured variables, the battery level is also stored. The datalogger and the probes are powered by a 12V-8AH battery. The battery is powered by a 10W photovoltaic panel of 35.5 x 25 cm.

The power supply of the battery is controlled by the PWM controller. PWM can also be used to control the power supplied to the data logger, optimising the performance of the battery and solar panel system. In turn, the datalogger supplies the sensors with 12V DC via a programmable switch. To protect the battery lifespan, the system enters in a hold mode when the charge is under 12.2 V, interrupting the measurements and data transmission. The system was designed to function for three days without battery recharging.

All the electronics are held inside a waterproof box (Figure 4.8-a). Each component is mechanically assembled on a metal plate fixed to the bottom of the casing (Figure 4.8-b). The sensors and photovoltaic panel are connected to the casing via cable plugs. Finally, to ensure that the box is at atmospheric pressure, a membrane made of a waterproof but breathable material (Gore-Tex®) is fixed to one wall of the box.

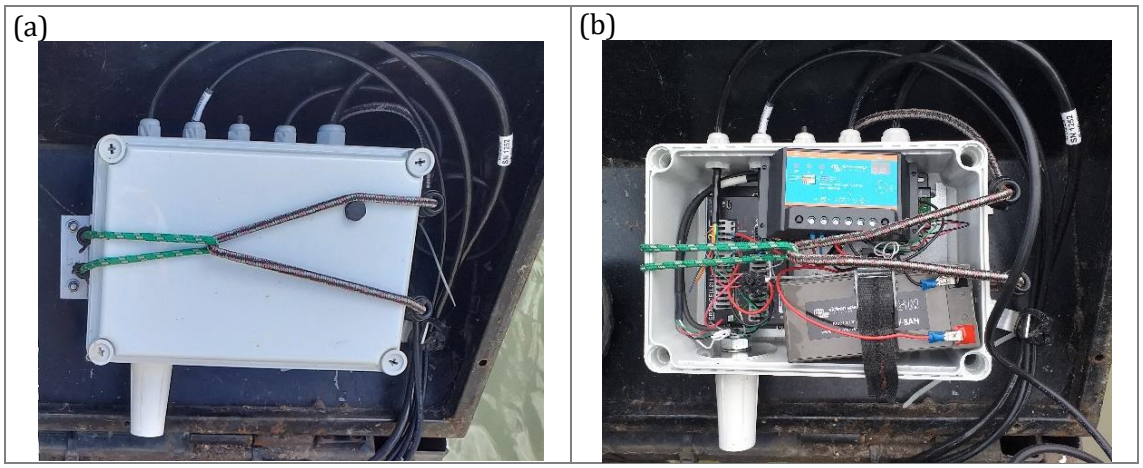


Figure 4.8. (a) Waterproof box and (b) all the electronics fixed inside the box.

During winter, the battery recharging via photovoltaic panel was limited due to the little sunlight exposure. To reduce battery consumption, the timestep was modified from 10 to 20 min after 24/11/2021 at both points. Nevertheless, the energy recharging was still not enough to cover the energy consumption during winter. Therefore, periods of more than a week of missing data were observed every year between November and February.

The data are transferred via a wireless modem and the GPRS internet by email once a day. The data are received in an email inbox (@enpc.fr) and read by a Python script for format standardization. They are then registered in a local file server by a Python script in scheduled times. Once formatted and stored in the server, the data can be visualized and exported in CSV format via the software Data Base Plot (DB-plot), developed by Philippe Dubois, metrology engineer at Leesu. The data transfer platform (Figure 4.9) is similar to the wireless real-time data transfer presented by Khac et al. (2018).

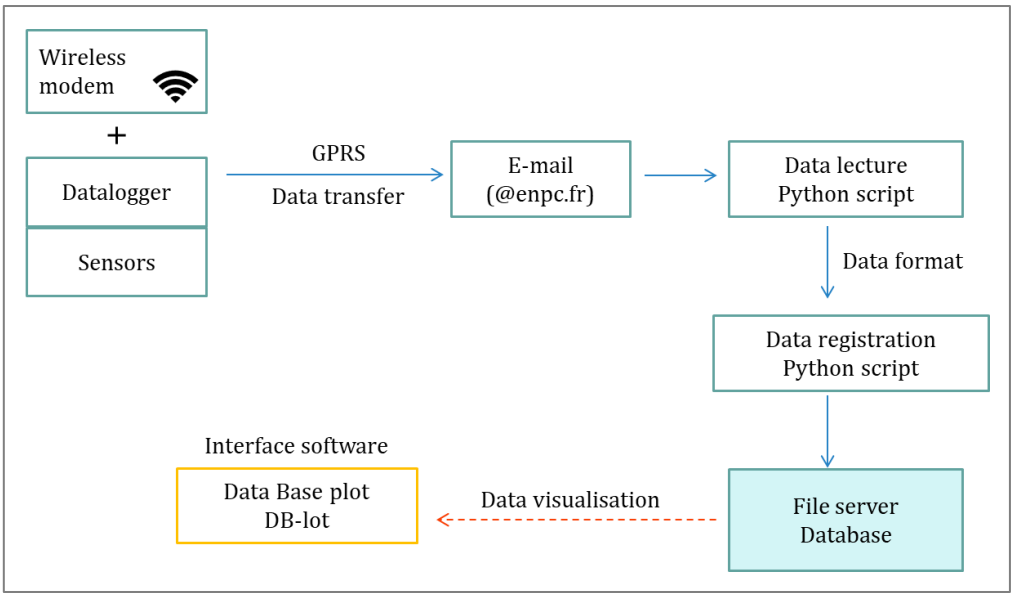


Figure 4.9. Scheme of data transfer platform

Eventually, the presented autonomous monitoring system was designed to be of easy implementation. The measurements were taken at fixed depths and the sensors were activated exclusively for measurements, reducing the energy cost (Le Vu et al., 2011).

4.1.3. Sensor maintenance and validation of field data

A profiler probe of reference, WiMo (nke), was used to validate the continuous measurements. The 10 cm-graduated rope of the Secchi disk was used to measure the depth to be compared with the continuously measured water level (Table 4.4).

Table 4.4. Dates of field campaigns

Date	Points	Secchi Disk	WiMo	Date	Points	Secchi Disk	WiMo
05/05/2021	A, B			15/03/2022	A, B		
20/05/2021	A, B			29/03/2022	A, B		
10/06/2021	A, B			22/04/2022	A, B		
22/06/2021	A, B			05/05/2022	A, B		
28/06/2021	A, B			15/06/2022	A, B		
05/07/2021	A, B			12/08/2022	A, B		
16/07/2021	A, B			18/08/2022	A, B		
23/07/2021	A, B			09/09/2022	A, B		
25/08/2021	A, B			15/09/2022	A, B		
19/10/2021	A, B			22/09/2022	A, B		
24/11/2021	A, B			14/10/2022	A, B		
12/01/2022	A, B			17/10/2022*	A, B	4	4
02/02/2022	A, B			09/11/2022	A, B		
16/02/2022	A, B			11/01/2023	A, B		

* 4 measurements at each point during the day

A major issue in high-frequency measuring is the reliability of the data and the frequency of field interventions for sensor maintenance (Tiberti et al., 2021).

Regular cleaning was required to avoid a biofouling formation and assure good quality data. In La Villette system, one indicator of biofouling formation, and therefore the need of cleaning, was conductivity. After some weeks, an increasing number of outliers was observed. They disappeared after cleaning intervention (Figure 4.10). The estimated

optimal duration was between 6 and 8 weeks, to preserve good quality data and reduce the quantity of field interventions.

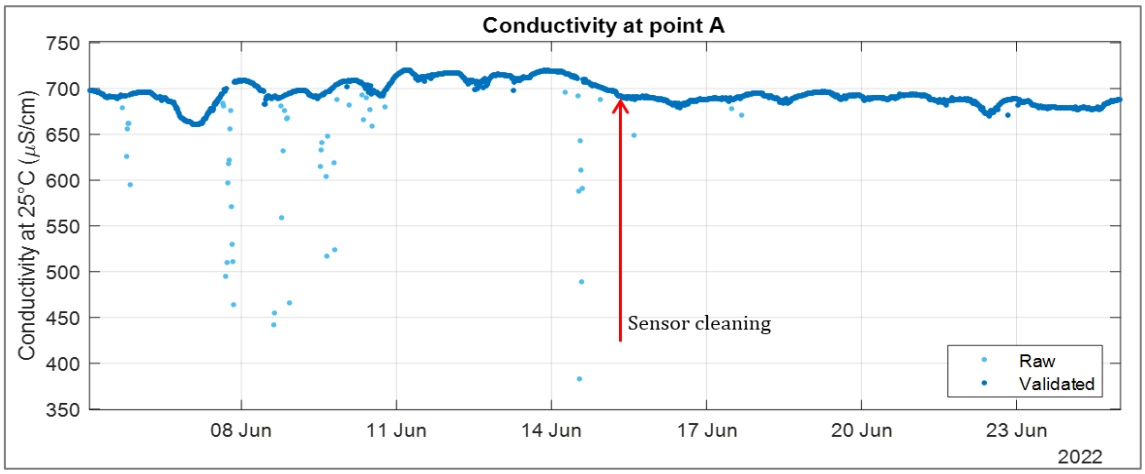


Figure 4.10. Conductivity outliers (light blue) before and after sensor cleaning (red arrow).

Before the final validation of the field data, some criteria were used for data pre-validation: (1) verification of the battery level to allow the correct measuring of the sensors; (2) physical meaning of the data range; (3) realistic values considering the history of measurements, current season, etc.

Known events that could impact the measured variables (e.g., weather events, such as a heavy rainfall), are considered to check whether sudden variations in the timeseries are outliers. Since outliers are more likely to occur after a long period without maintenance, it is important to consider the maintenance dates in data validation (Mourad and Bertrand-Krajewski, 2002; Tiberti et al., 2021).

The last step for data validation was based on device redundancy: a profiler probe was used to measure the same variables as *in situ* sensors. A discrepancy between both measurements will detect possible outliers. Moreover, *in situ* probes can detect a shift in the sensor timeseries. The required sensor calibration in the laboratory will be performed only when needed. This avoids unnecessary periods of lack of measurements.

The turbidity sensor data at point A were not validated (Figure 4.11). Therefore, the measurements were not explored.

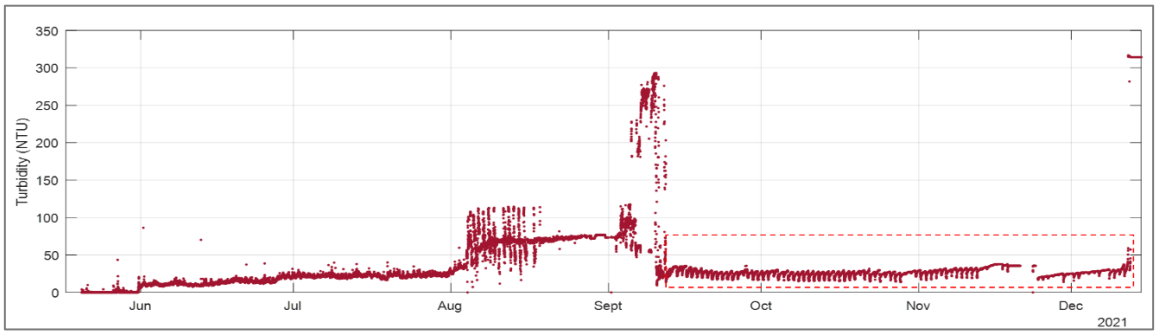


Figure 4.11. Turbidity measured at point A from 20/05/2021 to 12/12/2021, with regular oscillations from 13/09/2021 (red rectangle).

The sensor data of conductivity and water temperature were validated with the reference probe, WiMo (nke), once or twice a month, at upstream and downstream, plotting the values at the same depth (Figure 4.12, Figure 4.13, Figure 4.14).

At points A and B, the conductivity from the reference probe WiMo and OTT sensors was highly correlated (Figure 4.15, RMSE<8 µS.cm^{-1}, R^2=0.99). Similarly, at both measuring

points, the water temperature from WiMo and from Campbell sensors at three depths were very well correlated (Figure 4.16, RMSE≤0.21°C, R^2=1).

All the measurements of conductivity and water temperature can be considered as a reliable representation of the real respective values.

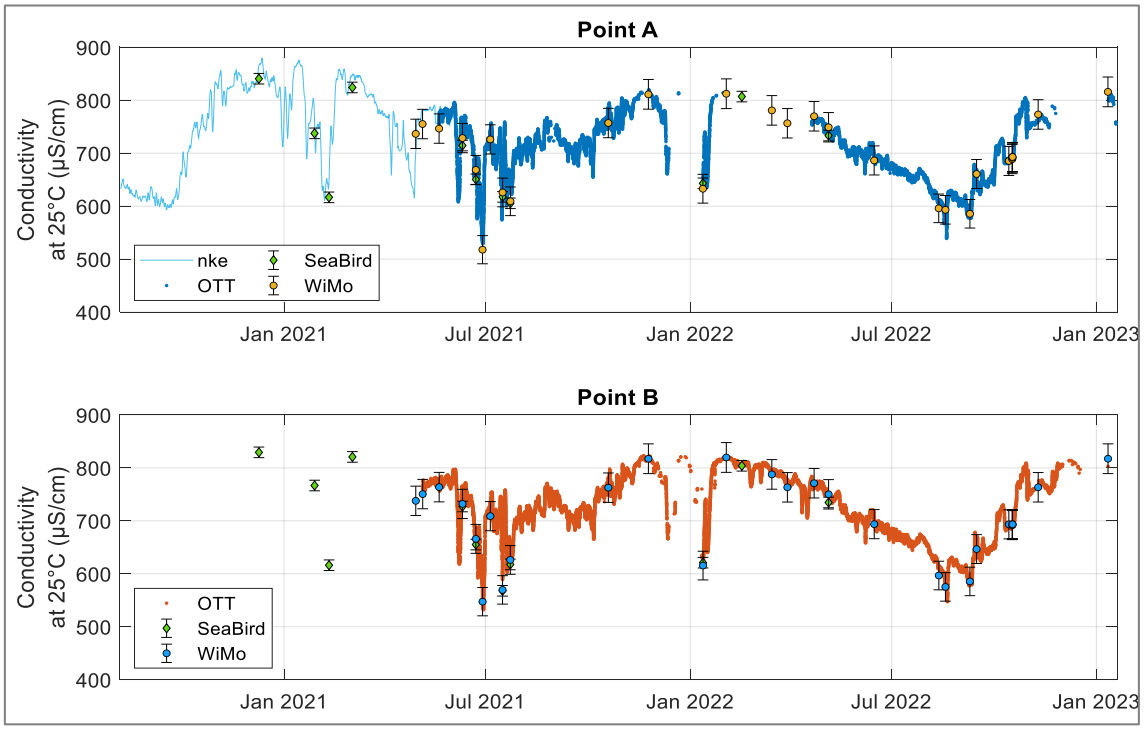


Figure 4.12. Conductivity at points A (top) and B (bottom), measured by nke and OTT sensors and WiMo probe, from August 2020 to January 2023.

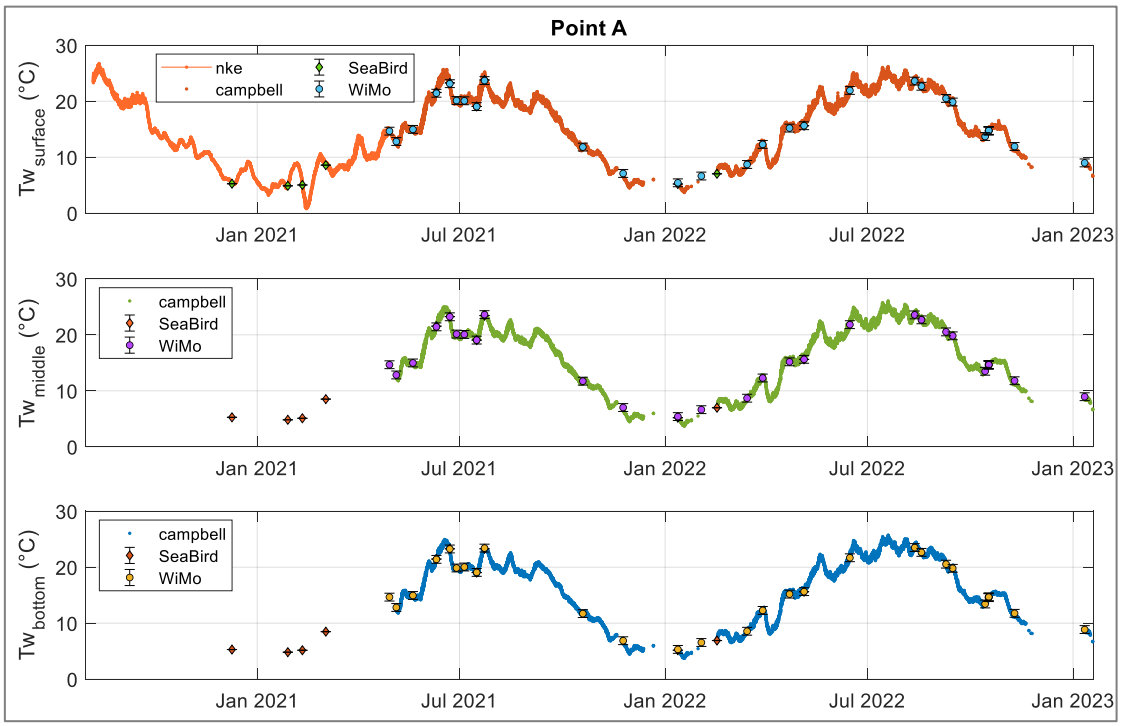


Figure 4.13. Water temperature on surface, middle and bottom at point A, measured with Campbell sensors and with WiMo probe (WM), from August 2020 to January 2023.

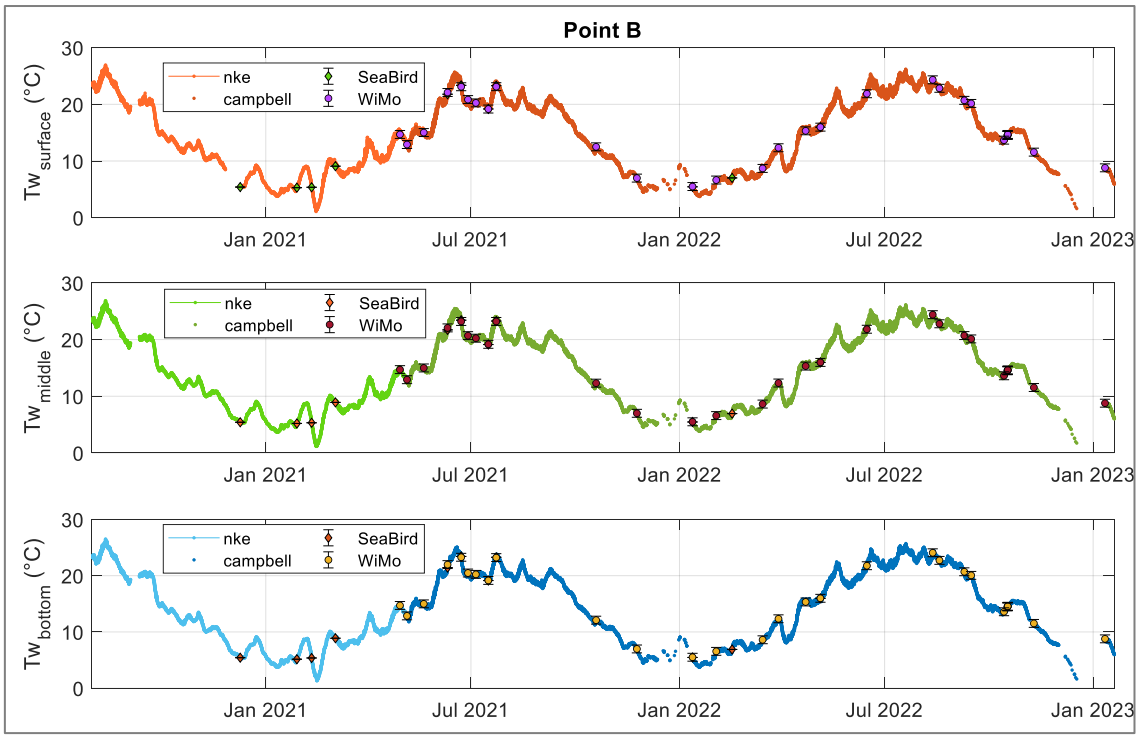


Figure 4.14. Water temperature on surface, middle and bottom at point B, measured with Campbell sensors and with WiMo probe (WM), from August 2020 to January 2023.

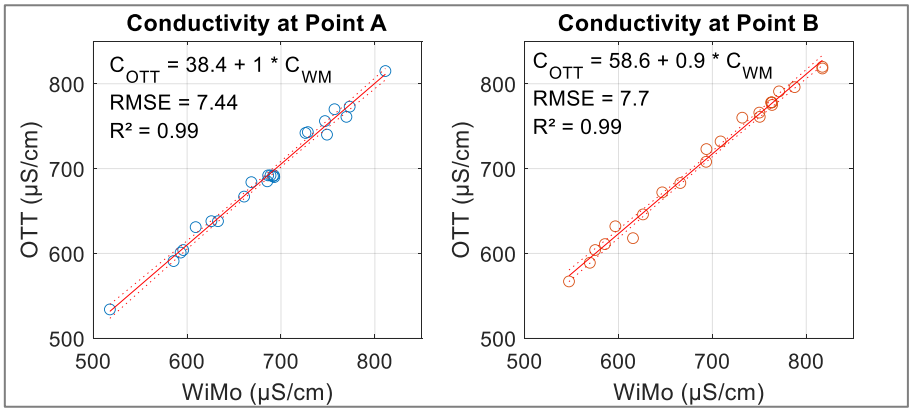


Figure 4.15. Linear regression between reference probe WiMo and OTT conductivity sensors at points A (left) and B (right).

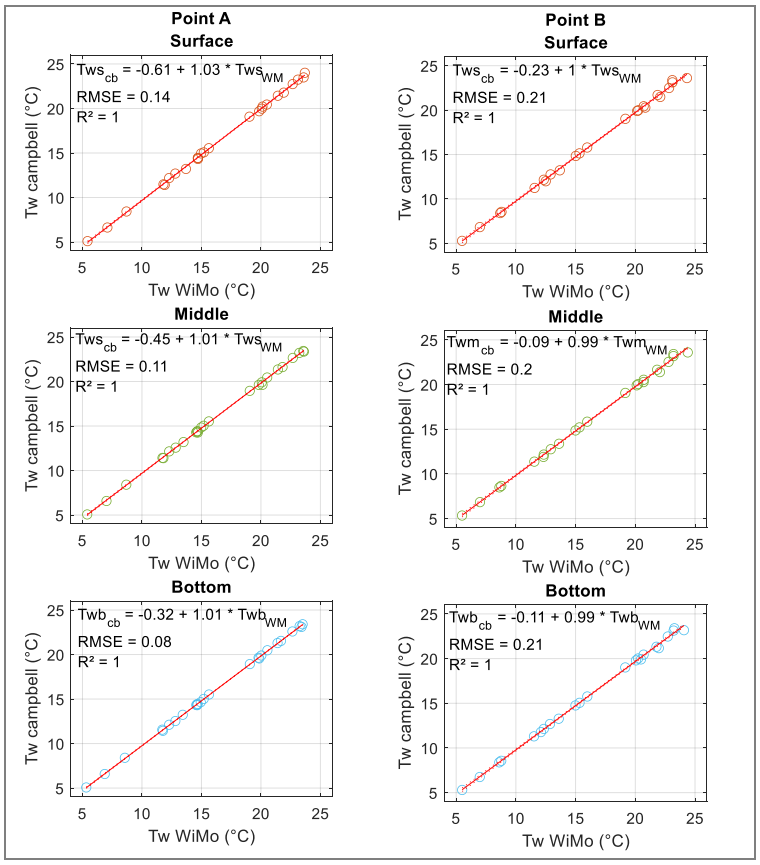


Figure 4.16. Water temperature: Linear relationship between reference probe (WiMo) and temperature sensor (Campbell) measurements at three depths (top, middle, bottom) in points A (left) and B (right).

The water level was validated with the punctual measurements with a 10cm-graduated rope (Figure 4.17-a). The punctual measurements were considered indicative values, due to the resolution of the graduated rope (0.05 m), 50 times greater than the resolution of the water level sensor (0.001 m). All the sensor measurements were within the incertitude range of the punctual measurements (Figure 4.17-b).

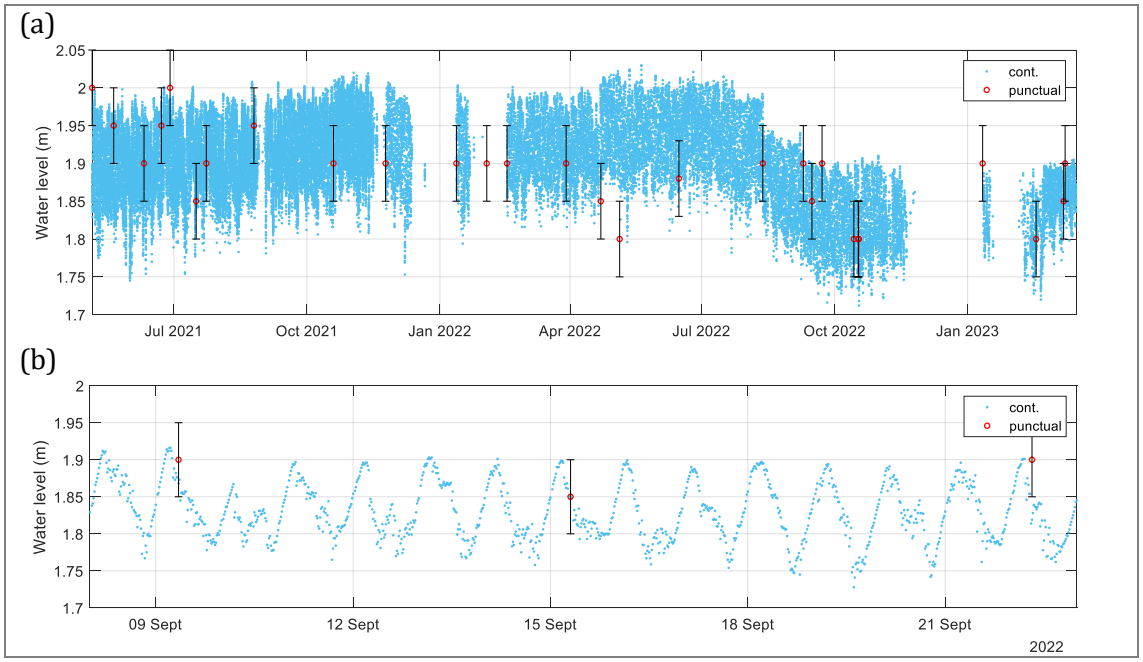


Figure 4.17. Water level at La Villette upstream (point A). Sensor measurements (in blue) and punctual measurements with graduated rope (red), from: (a) 05/05/2021 to 16/03/2023, and (b) 08 to 23/09/2022.

4.1.4. Correction of field data

Tiberti et al., (2021) described criteria for data quality control of an automated high frequency monitoring in a lake. The authors proposed a procedure to remove, correct or retain values. A similar procedure was adopted in the quality control of the present data.

In our monitoring system, the most common types of data alterations could be divided in two groups: (1) sudden increase or decrease of value in a short time period; and (2) drifting of measurements.

A sudden data variation was defined as a variation higher than the equipment precision (EP) between two time-steps, n-1 (C_{n-1}) and n (C_n). To be considered as an outlier, this variation should not be present in the following time-step, n+1 (C_{n+1}). In other words, values of n-1 and n+1 are closer than n-1 with n or n with n+1 (Equation 4.2).

$$C_n \text{is outlier if} \begin{cases} |C_{n-1} - C_n| > |EP| \\ |C_n - C_{n+1}| > |EP| \\ |C_{n-1} - C_{n+1}| < |C_{n-1} - C_n| \\ |C_{n-1} - C_{n+1}| < |C_n - C_{n+1}| \end{cases} \quad 4.1$$

Some variations are due to field interventions. They can be predicted, such as programmed interventions for maintenance and cleaning of sensors. In this case, the values are considered as outliers and removed. The gap can be filled with a linear interpolation between the consecutive time-steps before and after the intervention.

Another cause related to field interventions is due to a displacement of the sensors. Some shifts could occur due to unexpected actions, or a problem with the sensor attachment. When a shift in the vertical position of a sensor is identified, the values should be flagged.

The drifting of data was identified with reference probes measurements. Then, it was corrected with these reference values or by recalibrating the sensor in the laboratory or by the fabricant.

4.1.4.1. Outlier treatment

The outliers can be identified by a sudden variation of data in a short time-step. Data from consecutive time-steps of 20 min were compared. In La Villette system, if the data variation was higher than the precision of the sensor, it was considered as an outlier. Once identified, these outliers were flagged for further checking.

In our data, shifts between consecutive time-steps superior to 5 µS.cm^{-1} for conductivity and 0.15 °C for water temperature were flagged and possibly removed. Some examples of outlier processing are presented in the following paragraphs.

- *Variations due to sensor maintenance*

All the field campaigns were registered in a follow-up file (Annex IV-c). Cleaning sensors could last more than one measurement time-step (20 min). Therefore, sensors registered one measure outside of the water, resulting in a sudden increase or decrease of values between two time-steps in all chain of sensors (Figure 4.18). All measurements taken during sensor cleaning were excluded.

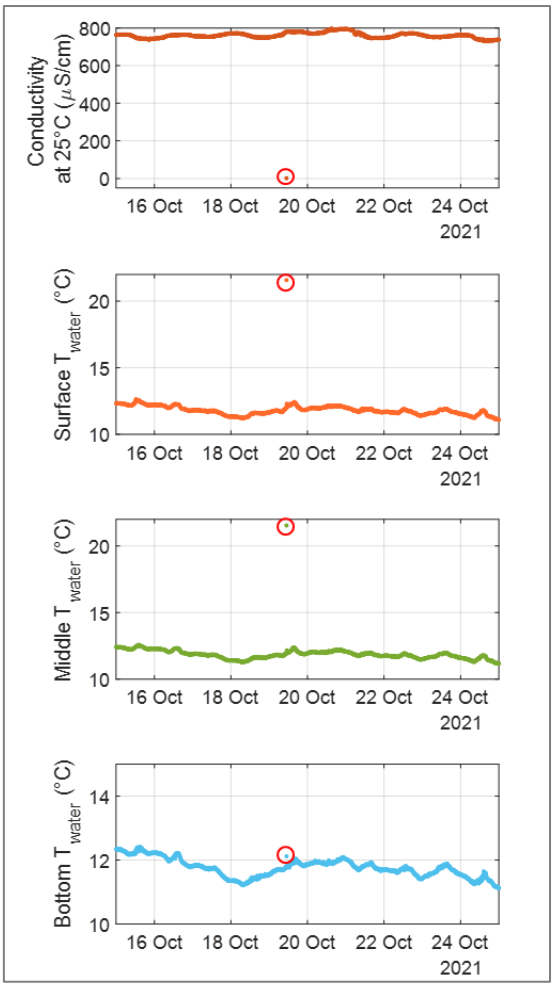


Figure 4.18. Conductivity and water temperature at three depths measured at point B from 15 to 25/10/2021, with field intervention on the 19/10/2021 (red circles).

- *Limited disturbance of the timeseries*

The disturbance can last several time-steps. The values change in a short period, and suddenly come back to the similar range of values (Figure 4.19). These variations may result from biofilm formation or perturbation caused by local small fauna. Thus, a regular cleaning of the sensors is essential for the acquisition of reliable data.

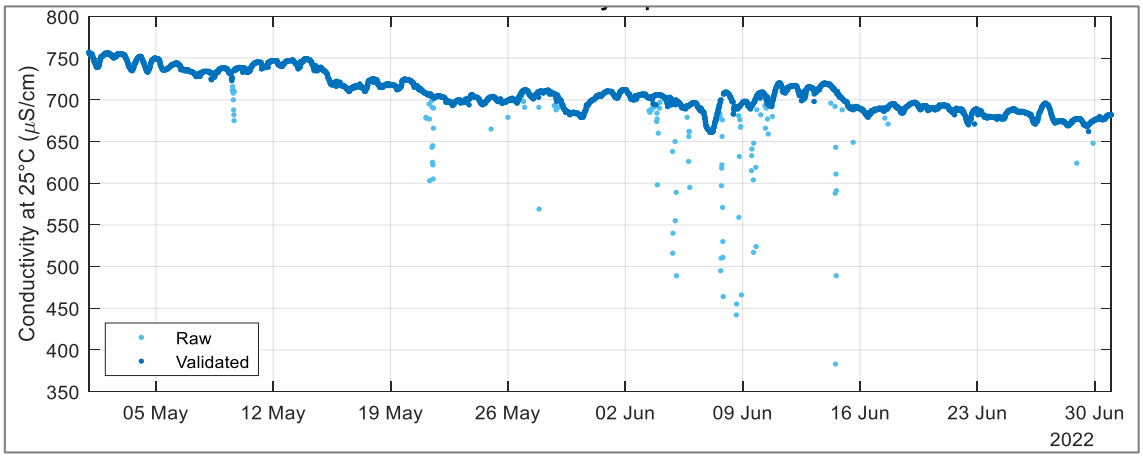


Figure 4.19. Conductivity at point A: raw data (light) and validated data (dark), from 01/05/2022 to 01/07/2022.

- *Shift of the timeseries*

At point A, from 20/09/2021 to 19/10/2021, the water temperature at the bottom was higher than at the surface and the middle of the water column, especially during night mixing (Figure 4.20). This was due to a fixing problem of the sensor: the attachment of the bottom sensor broke, and the sensor sank to the canal bed. The sensor was relocated at 1.75m depth in the field intervention of the 19/10/2021. The difference between surface and canal bed temperatures during the night might be due to the influence of the macrophyte canopy on the heat flux.

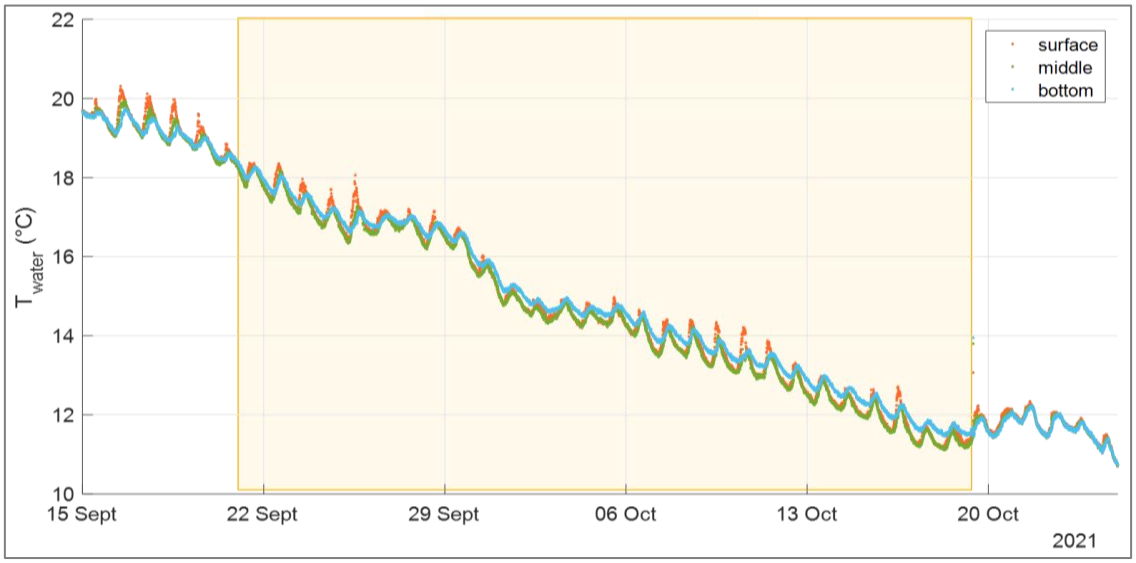


Figure 4.20. Water temperature at point A at the surface (orange), middle (green) and bottom (blue). From 20/09/2021 to 19/10/2021 (yellow box), the bottom sensor was not at the right depth (1.75 m), but on the bed of the canal.

4.1.4.2. Water level correction

On the 23/07/2022, the water level sensor was shifted downwards, increasing the measured water level. This shift was due to an unexpected action for the maintenance of other sensors installed in the same place (Fluidion sensors for FIB measurements, Ville de Paris). The sensor was put back at the right place on the 12/08/2022. From 23/07 to 12/08/2022, the water level was overestimated (Figure 4.21).

For the data correction, we considered the difference between the level measured immediately before and after the rise of the values (0.123 m), and before and after the restoration of the sensor depth (0.190 m). The mean of the two values (0.156 m) corresponds to the sensor displacement. The timeseries was corrected by deducting this value (Figure 4.22).

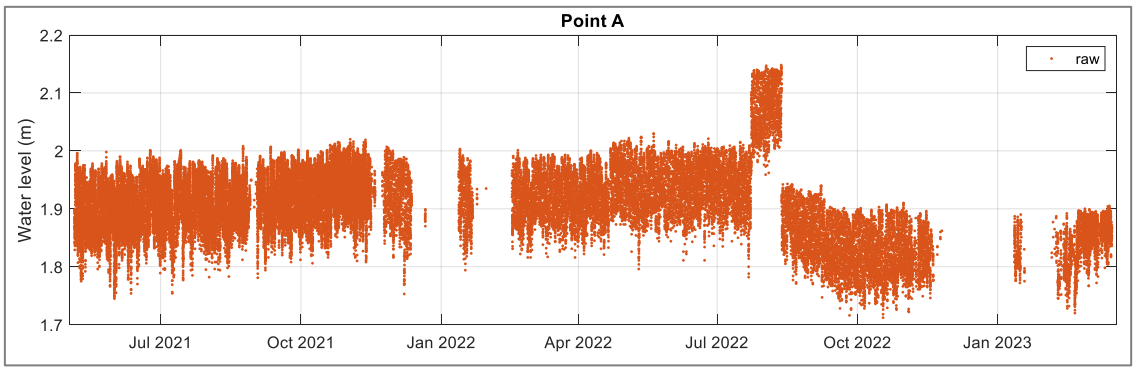


Figure 4.21. Water level measurements at point A before correction (20/05/2021 to 16/03/2023).

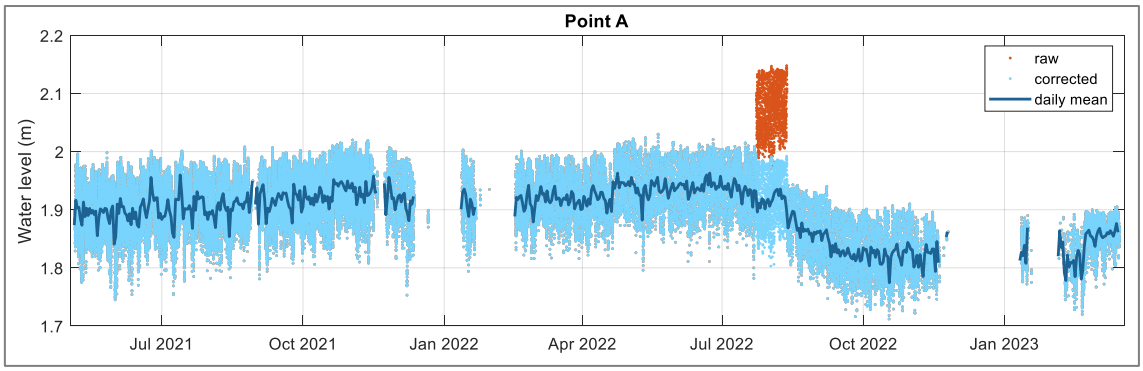


Figure 4.22. Raw and corrected water level data at point A (20/05/2021 to 16/03/2023).

4.1.4.3. Conductivity correction

In January 2022, the multiparameter sensor at point A (MS5 OTT) sensor was defective, causing an elevated energy consumption. It was sent to the company for maintenance. Afterwards, the conductivity at point A was lower than at point B and closer to WiMo reference values (Figure 4.23). Therefore, a correction of the conductivity shift was necessary. It was computed from the relationship between WiMo reference measurements and the correspondent OTT timeseries. The final relationship was based on a 13 corresponding data at point A and 10 at point B (Table 4.5).

A good relationship was obtained at both points (R^2=1.00, Figure 4.24). The mean difference between the values of OTT_A and OTT_B was -18 µS.cm^{-1}, value used for the final correction (Figure 4.25).

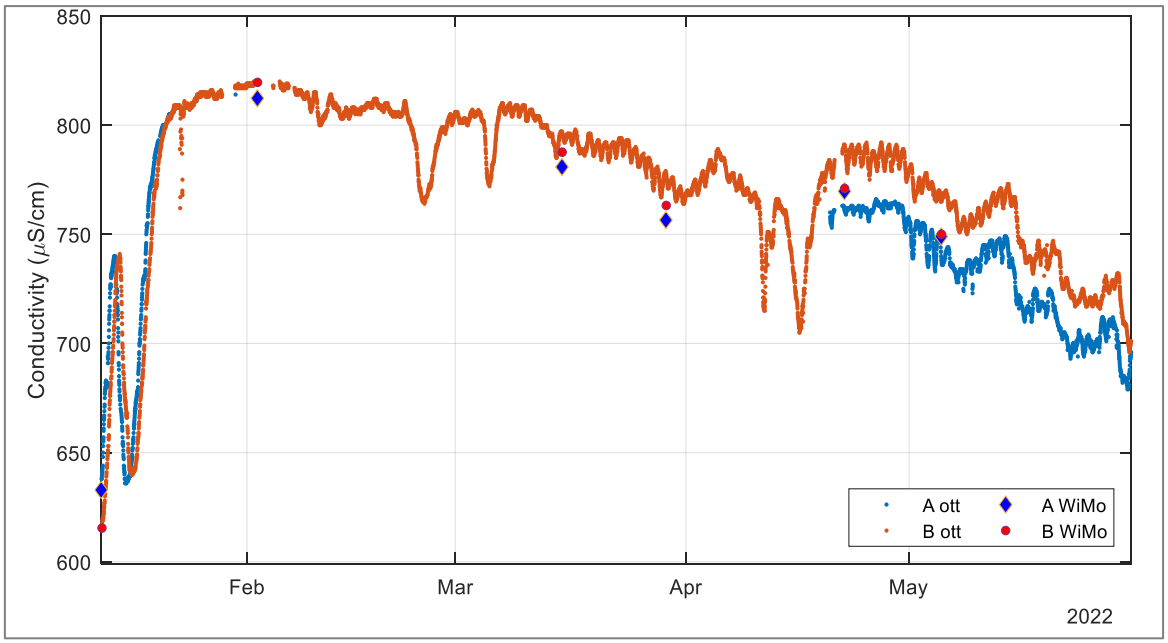


Figure 4.23. Conductivity at points A and B (11/01/2022 to 31/05/2022) before shift correction.

Table 4.5. Conductivity measurements from the reference probe (WiMo) and from continuous measurement sensors (OTT) at points A and B, from 22/04/2022 to 11/01/2023.

Point A	Conductivity (µS.cm^{-1})		Point B	Conductivity (µS.cm^{-1})	
Date	WiMo	OTT MS5	Date	WiMo	OTT DS5
22/04/2022 07:40	770	761	22/04/2022 08:00	771	791
05/05/2022 07:40	749	740	05/05/2022 08:40	750	766
15/06/2022 07:20	686	692	15/06/2022 08:40	694	723
12/08/2022 08:00	596	604	12/08/2022 10:20	597	632
18/08/2022 08:40	593	601	18/08/2022 09:00	575	604
09/09/2022 08:40	586	591	09/09/2022 08:00	586	611
15/09/2022 07:40	661	667	15/09/2022 08:40	647	672
14/10/2022 07:40	685	685	14/10/2022 08:20	693	708
17/10/2022 09:20	689	692	17/10/2022 10:10	692	-*
17/10/2022 11:00	693	692	17/10/2022 11:40	694	-*
17/10/2022 13:20	691	691	17/10/2022 13:40	692	-*
17/10/2022 15:20	693	690	17/10/2022 16:00	693	-*
09/11/2022 16:20	773	773	09/11/2022 17:00	763	778
11/01/2023 09:40	816	-	11/01/2023 11:20	817	820

* Problem with saving and transmission of data.

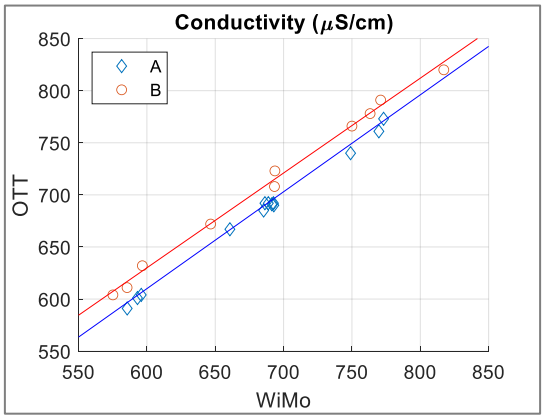


Figure 4.24. Relationship between conductivity from WiMo and OTT at points A (blue, $OTT_A = 0.93 \cdot WiMo_A + 52$) and B (red, $OTT_B = 0.91 \cdot WiMo_B + 84$).

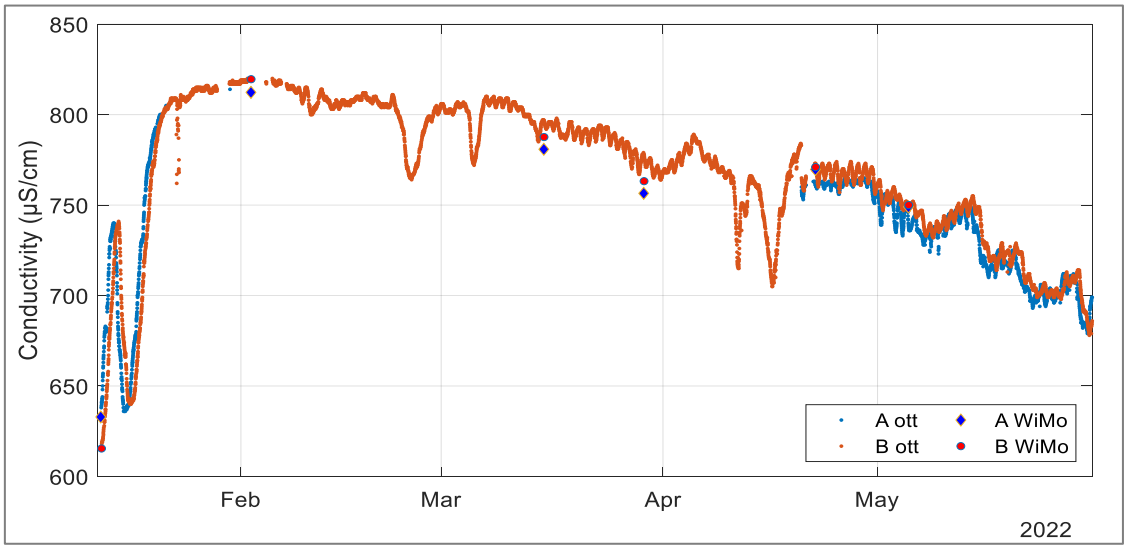


Figure 4.25. Conductivity at points A and B (11/01/2022 to 31/05/2022) after data correction.

4.1.5. Time evolution of the validated timeseries

The trends and patterns of validated timeseries are presented in this section.

4.1.5.1. Water level and Secchi depth

The time evolution of the water level at point A (upstream) was assessed over the studied period. The water transparency evolution was estimated through Secchi depth data at points A and B measured during the regular field campaigns.

From May 2021 to July 2022, the mean water level measurement was 1.92 m. From July 2022 to September 2022, it dropped of around 0.10 m (Figure 4.26). From October 2022 to January 2023, it raised again to 1.85 m. In 2022, in the Ile-de-France region, a drought period occurred from spring to summer (Météo-France, 2023), which may explain the water level decrease.

On a daily basis, the water level variation is around 0.20 m. The highest values are usually measured in the early morning and the lowest values in the afternoon. The regular variation of water level reflected the fluctuations of the pumping station and the lock operation.

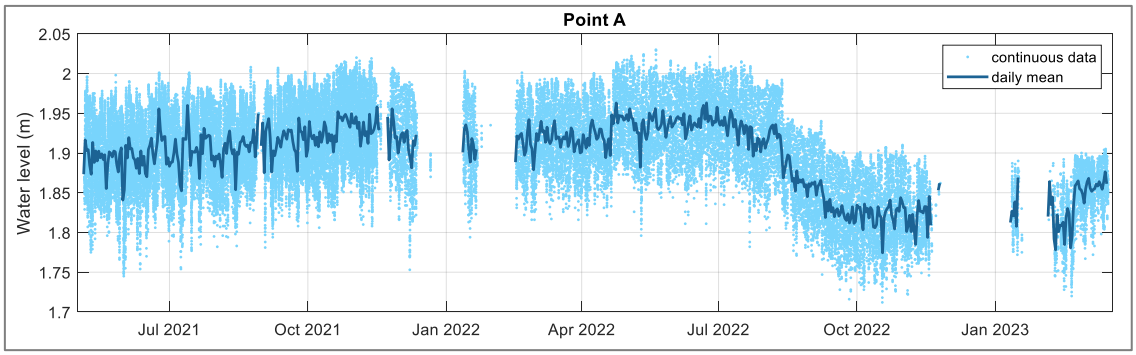


Figure 4.26. Corrected water level measurements at point A, from 20/05/2021 to 18/01/2023.

The Secchi depth values from August 2020 to March 2023, were very similar at upstream and downstream (Figure 4.27).

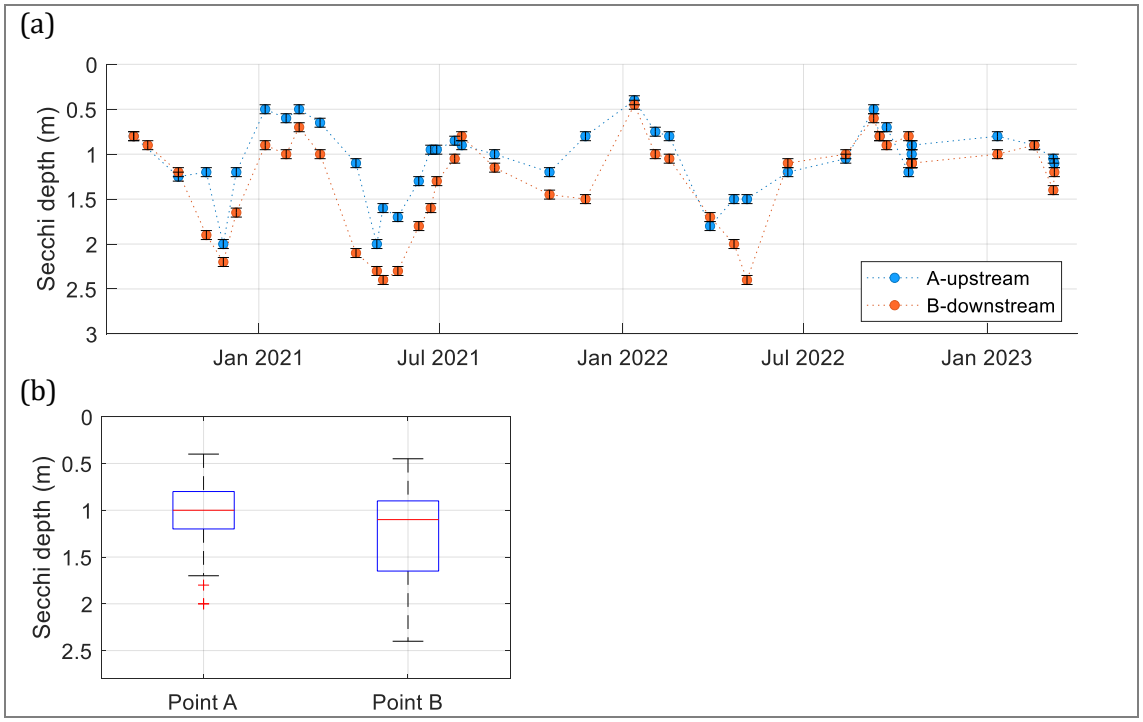


Figure 4.27. (a) Secchi depths at points A and B; (b) Boxplot of the data from Aug 2020 to Mar 2023.

4.1.5.2. Thermal stratification

The thermal regime in La Villette basin was assessed at the daily and seasonal timescales. From August 2020 to March 2023, the temperature range at both monitoring points was [0.82-26.93] °C (Figure 4.28-a, Figure 4.28-b).

The daily thermal regime is illustrated for a high temperature period, from the 5th to the 23rd of June 2021 (Figure 4.29). At points A and B, the water column was stratified during the day and mixed every night. At point B, the surface (0.50 m) and middle (1.20 m) water temperature were very close. This suggested that the temperature gradient is located between middle (1.20 m) and bottom (2.00 m) layers. To assess with a finer vertical resolution the stratification of the water column, additional temperature sensors could be located at 1.00 m, 1.70 m, and 2.50 m.

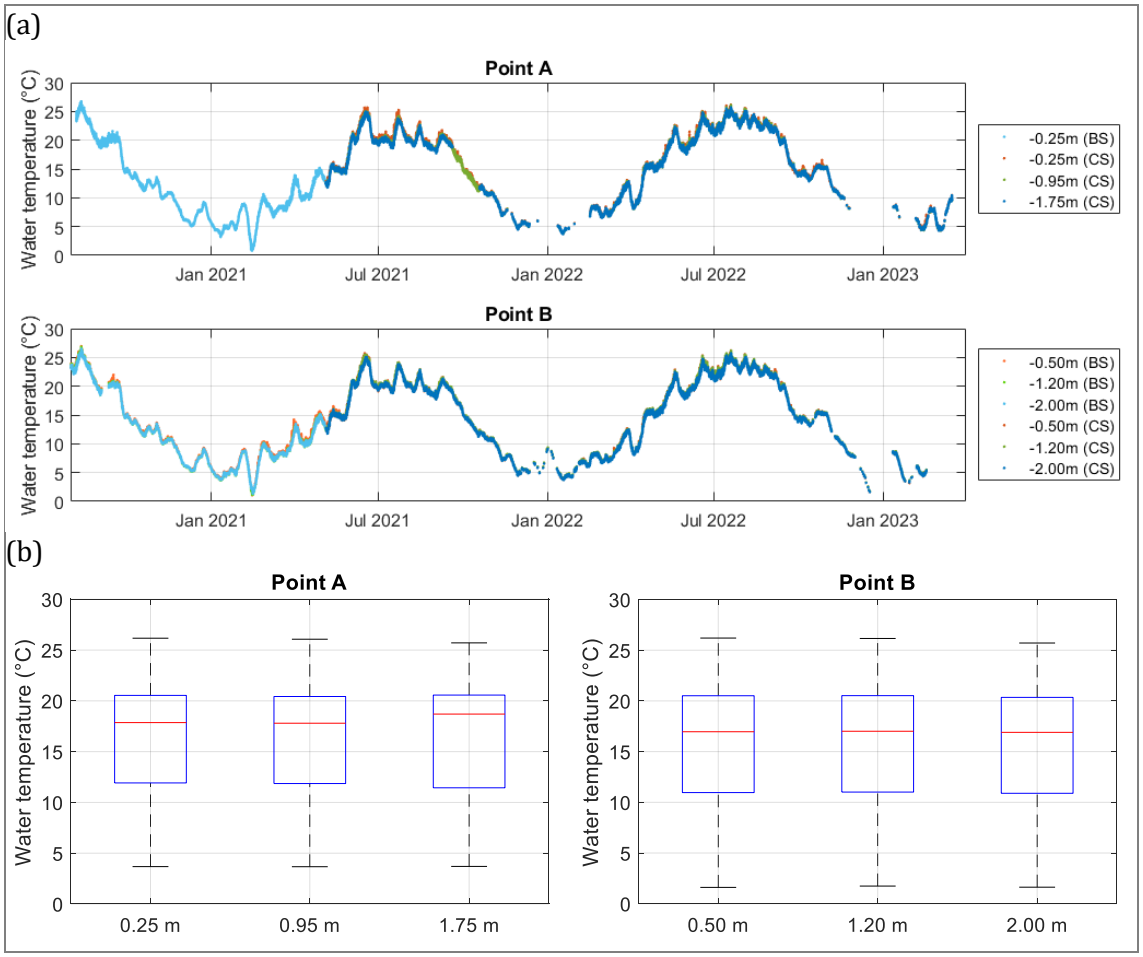


Figure 4.28. (a) Water temperature at three depths at points A (top) and B (bottom), of basic (BS) and complete systems (CS) from August 2020 to March 2023; (b) Boxplot of the data of points A (left) and B (right) from May 2021 to March 2023.

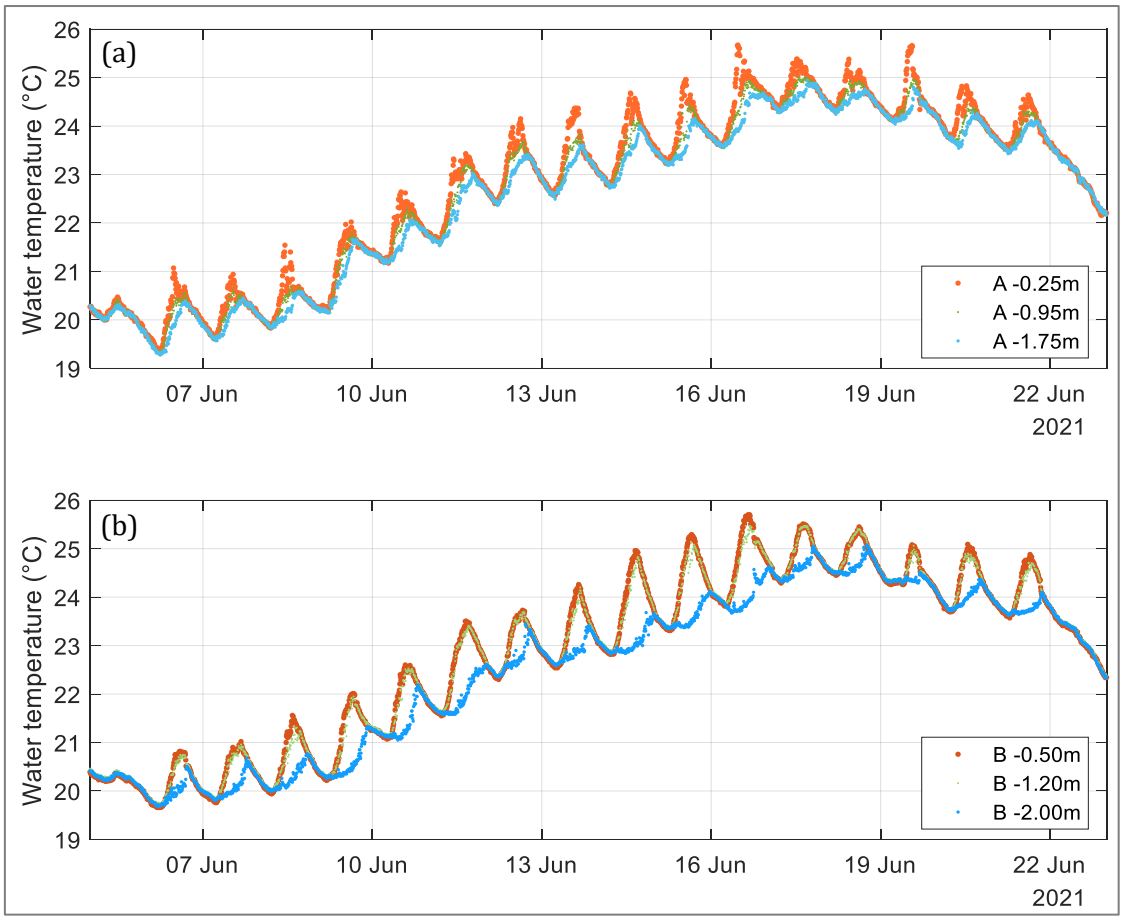


Figure 4.29. Water temperature at three depths at points A (a) and B (b): zoom in a period of higher temperatures, 05 to 23/06/2021.

The amplitude of the thermal stratification was computed as the difference between surface and bottom water temperature (ΔTw, Figure 4.30-a, Figure 4.30-b). Between 19/03/2021 and 05/05/2021, the data recorded by the basic monitoring system were not reliable (Annex IV-e). Therefore, ΔTw could not be computed.

The greatest amplitude at point A (2.51°C) occurred in 04/06/2022 and, at point B (1.93°C), in 14/06/2021. The higher amplitudes at point A than at point B might be due to the position of the surface sensor: at point A, it is 0.25 m closer to the water surface than at point B.

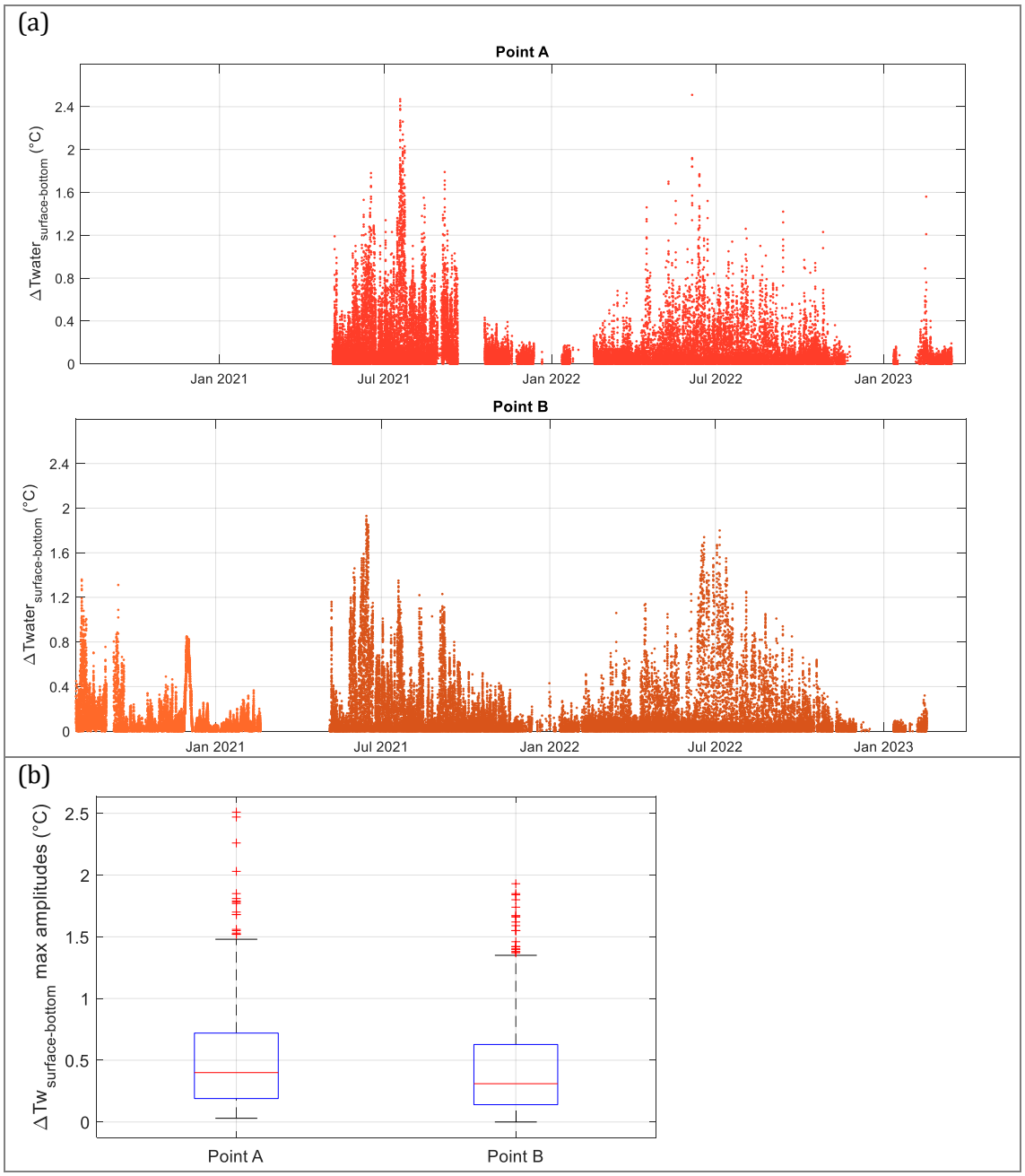


Figure 4.30. (a) Amplitude of thermal stratification at points A (top) and B (bottom), from August 2020 to March 2023; (b) Boxplot of daily maximum positive thermal amplitude at points A and B, from May 2021 to March 2023.

Water temperature monitoring would give information about the long-term thermal regime in a water course. In La Villette basin, it was possible to record the time evolution of the water column thermal stratification.

Urban waterbodies can provide different services, such as cooling energy (Fraîcheur de Paris, 2023). In this case, monitoring the water temperature stratification can be useful to improve the management of the energy system.

4.1.5.3. Conductivity timeseries

The range of conductivity values was similar at both monitoring points A, upstream, and B, downstream. (Figure 4.31). From August 2020 to March 2023, the conductivity range in La Villette basin was [529-890] $\mu S.cm^{-1}$ (Figure 4.32). Higher conductivity was measured during winter, with a gradual decrease until summer. This annual trend was observed in 2021 and 2022.

Conductivity variations were considered as representative of the environment when they were greater than 20 $\mu S.cm^{-1}$. Particularly, a decrease of conductivity would occur after rainy events on the upstream watershed (Bhurtun et al., 2019; Halliday et al., 2015). Rainfall in the upstream watershed of the Ourcq canal would cause the decrease of conductivity in La Villette basin.

The variations at point A were followed by variations at point B some hours later (Figure 4.33). The inflexion point represented the arrival of a different water mass from upstream. Therefore, it was used as reference to find the equivalent water mass of point A in point B. The time-lag between the inflexion points enables a more precise estimation of the mean transfer time between the upstream monitoring (point A) and the bathing area (point B).

Well-defined inflexion points from June to July 2021 were identified and the transfer time between points A and B was estimated (Table 4.6). In this period, the range of the transfer time was [9h30-24h10] (Figure 4.34).

The differences among the time-lags of conductivity evolution between points A and B may be related to variations in water velocities. Moreover, when the water column is stratified, the velocity may be different in the surface and the bottom layers.

The mixing of water masses may interfere on the measured conductivity evolution, flattening the peaks or valleys of the data. For a representative transfer time estimation, the identified inflexion points should be well-defined at both upstream and downstream.

In conclusion, the high-frequency monitoring of conductivity could be used to detect the arrival of new water masses, particularly when heavy rainfall occurs in the upstream watershed. In addition, the conductivity variation between upstream and downstream points enables the estimation of a mean transfer time between these points. However, one limitation is caused by the detection of representative inflexion points at both monitored points.

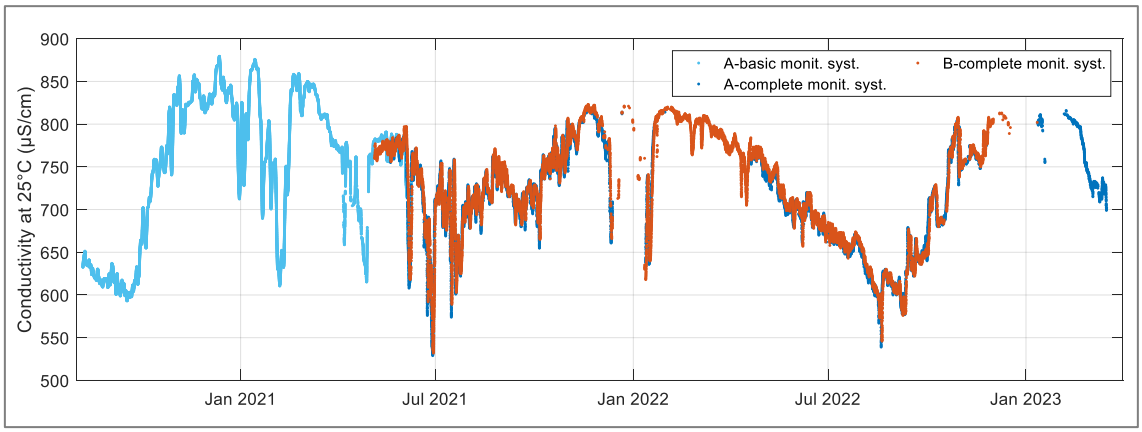


Figure 4.31. Conductivity at points A (basic monitoring system in light blue, complete monitoring system in dark blue) and B (complete monitoring system in orange), from August 2020 to March 2023.

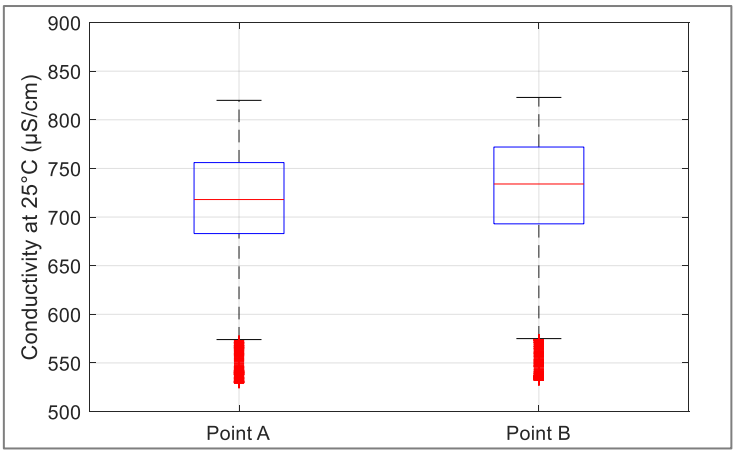


Figure 4.32. Boxplot of conductivity data at points A and B from May 2021 to March 2023.

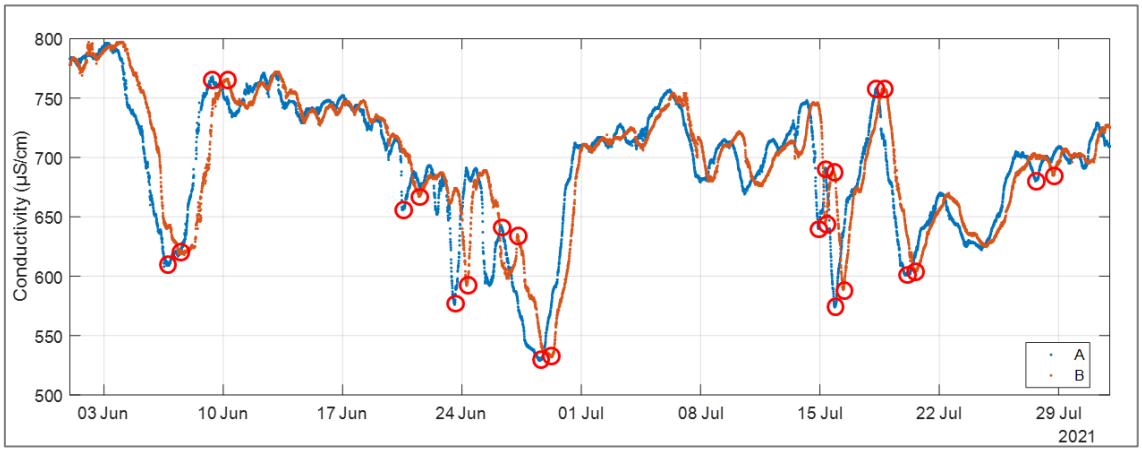


Figure 4.33. Conductivity at points A (blue) and B (orange), with the well-defined points of local maximum and minimum (red circles), from 01/06/2021 to 30/07/2021.

Table 4.6. Estimation of transfer time between points A and B. Data from June to July 2021.

Datetime at point A	Datetime at point B	Transfer time between A and B
06/06/2021 17:40	07/06/2021 09:40	**16:00:00**
09/06/2021 08:50	10/06/2021 06:20	**21:30:00**
20/06/2021 12:40	21/06/2021 12:20	**23:40:00**
23/06/2021 13:40	24/06/2021 06:40	**17:00:00**
26/06/2021 06:40	27/06/2021 06:20	**23:40:00**
28/06/2021 17:00	29/06/2021 08:00	**15:00:00**
14/07/2021 22:30	15/07/2021 08:00	**9:30:00**
15/07/2021 07:50	15/07/2021 18:20	**10:30:00**
15/07/2021 21:20	16/07/2021 09:20	**12:00:00**
18/07/2021 09:00	18/07/2021 21:20	**12:20:00**
19/07/2021 23:00	20/07/2021 11:40	**12:40:00**
27/07/2021 15:30	28/07/2021 15:40	**24:10:00**

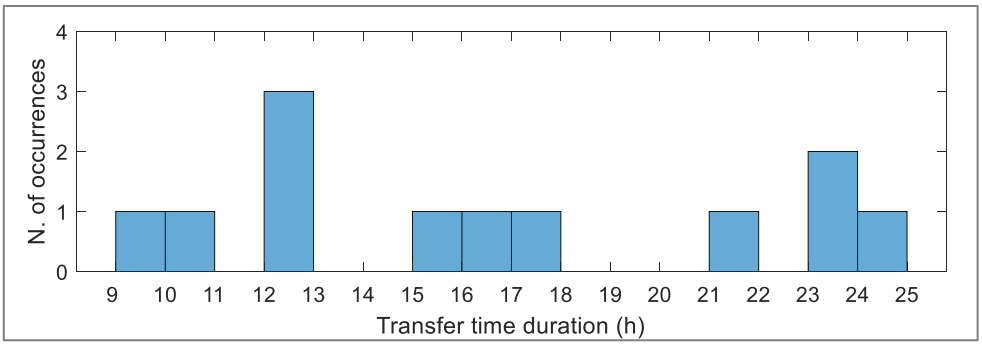


Figure 4.34. Distribution of the estimated transfer time (in hours) from points A to B. Data from June to July 2021.

An autonomous monitoring system was installed in two points, upstream (point A) and downstream (point B), of the urban La Villette canal. The system is of easy implementation. It includes a data transfer platform for high-frequency measurements of 20-min time step. The autonomous system is powered by a photovoltaic panel and battery power system. During winter, the duration of daily sun exposition is limited, and it is not enough to recharge the battery.

Three physicochemical variables were measured, validated, and explored: water level, conductivity, and water temperature at three depths (0.25 m, 0.95 m, 1.75 m at point A, and 0.50 m, 1.20 m, and 2.00 m at point B). The presented data was collected over a 32-month period, from August 2020 to March 2023.

To assure the data reliability, criteria were adopted for data pre-validation:

(i) Verification of battery level and status of sensors;
(ii) Verification of the data range (physically possible and locally realistic);
(iii) Evaluation of sudden variations considering known events such as hydrological conditions;
(iv) Delay since the last sensor maintenance;
(v) Validation with a profiler probe of reference.

The use of reference probes for *in situ* data validation allowed to detect the need of sensor calibration.

Regular sensor cleaning is essential to assure good data quality. The increase of conductivity outliers was found to be an indicator of cleaning need.

High-frequency monitoring enabled to analyse the variations of the physicochemical variables at hourly, daily, and seasonal timescales.

In La Villette basin, from spring to mid-autumn, a daily thermal stratification occurred, breaking up during the night. Large conductivity variations could be used to estimate transfer time between upstream and downstream points.

The observed thermal stratification impacts the transfer time between upstream and downstream points. In the model, which was implemented to compute La Villette hydrodynamics, the high-frequency upstream water temperature data can be used as input values, and the downstream data for the model validation.

The monitoring system was designed to be able to include more sensors. It will improve the surveillance of the long-term evolution of the measured variables. It will help measuring the impact of climate change in this urban system in the next years and defining possible solutions to mitigate climate change impacts. In the short to medium term, the impact of potential sanitation works on the waterbody could be assessed.

4.2. Dissolved organic matter characterization of urban waters from fluorescence analysis

Most results presented in this chapter are part of a paper submitted for publication in *Frontiers in Water*, named *"Fluorescence spectroscopy for tracking microbiological contamination in urban waterbodies"*, presented in Annex V.

To improve the detection of faecal contamination in a water body, a fast-monitoring technique is required. A change in DOM can be related to faecal contamination and the fluorescence spectroscopy can be used to detect microbiological DOM changes (Sorensen et al., 2021).

In this chapter, the results of fluorescence spectroscopy for the characterisation of DOM of different types of urban waters are presented. The variation of fluorescence indices and PARAFAC components was analysed in an EEM dataset of 414 samples. The components from PARAFAC model, specific to our dataset, were associated to the reference fluorophores from Coble (1996), peaks T, M, C and A, for characterization.

Fluorescence data were collected in waterbodies located in Paris metropolitan region, France. The sampling sites are presented in section 3.1.

The presented results focused on:

(i) Identifying the major fluorescence components in different investigated urban waters.
(ii) Examining seasonal variation of FDOM components and the impact of local rainfall episodes in these components.
(iii) Exploring the information provided by FDOM about microbiological water quality for monitoring purposes and identifying a potential signature of local contamination due to SN and WWTP outlets or runoff during rainfall events.

In this section, the results concerning the indices HIX, BIX, FI and the PARAFAC components in La Villette basin, the river Marne, stormwater network outlets (SO-N and SO-S) and WWTP outlet (WWTP-O) are presented, discussed, and compared to other results in similar waterbodies.

The land cover in the river Marne and the seasonal variation of fluorescence components in La Villette basin are also discussed.

4.2.1. Sampling weather conditions in La Villette basin and the river Marne

In La Villette basin, samples were collected over 29 months (October 2020 to March 2023). From the total samples (n=183), 54% (n=98) were collected in wet weather, and 46% (n=85) in dry weather. Wet weather is defined as a rainfall episode higher than 2 $mm.day^{-1}$ within 3 days (definition given in section 3.3.1).

In the river Marne, a total of 34 sampling campaigns were conducted in four months (June to September 2022): 18 in VGA (4 in dry and 14 in wet weather) and 19 in SMV points (11 in dry and 5 in wet weather). The campaigns were predominantly conducted during dry weather (78% of the samples).

The rainfall events associated to wet weather samples have different distribution of intensities in both watercourses, due to the difference in the length of the sampling periods.

The distribution of the cumulated precipitation during 72h before sampling in wet weather in La Villette basin shows 24% of precipitations between 8 and 10 mm. The observed maximum intensity was of 48 mm in 72h (Figure 4.35-a).

In the river Marne, 27% of the wet weather samples were collected after rainfall between 18 and 20 mm in 72h. The observed maximum intensity was of 26 mm in 72h (Figure 4.35-b).

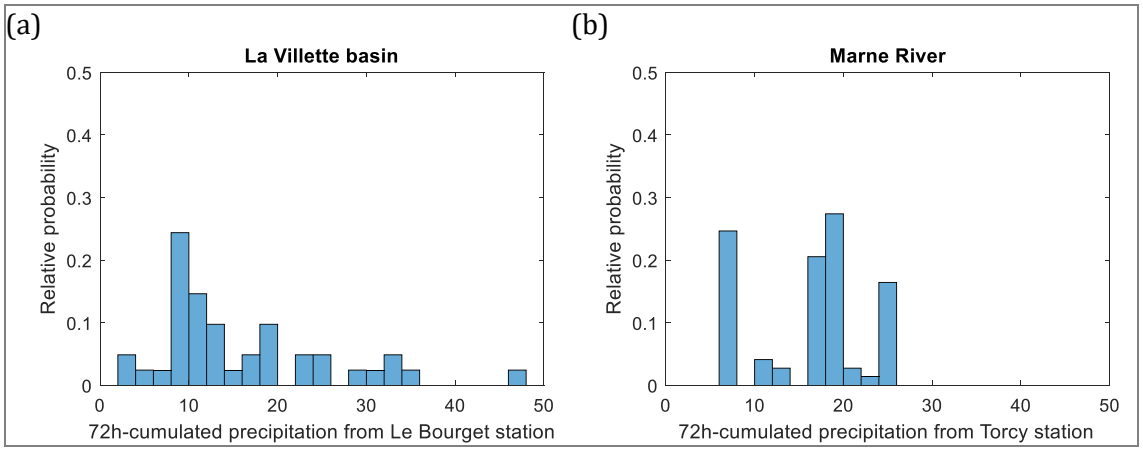


Figure 4.35. Distribution of relative probability of 72h-cumulated precipitation before sampling in La Villette basin (a) and the river Marne (b).

4.2.2. Fluorescence indices for DOM characterization

4.2.2.1. Comparison of HIX values

Huguet et al. (2009) associated DOM characteristics with a range of values calculated according to Zsolnay et al. (1999). To find the correspondent range values for Ohno's expression (2002), which provides values in the range [0 1], better adapted to comparison between different water bodies, HIX was calculated with both expressions for all samples (Equation 4.2, Figure 4.36). The proposed conversion between the different HIX scales allows us to compare the values in a unified way.

The HIX ranges from Huguet et al. (2009) and their DOM characteristics were then associated with new HIX ranges from Ohno (2002) (Table 4.7). For comparative purposes, DOM characteristics for HIX and BIX ranges defined by Huguet et al. (2009) are included in Table 4.7 and presented in Figure 4.37.

$$HIX_{Ohno} = \frac{HIX_{Zsolnay}}{HIX_{Zsolnay} + 1} \qquad 4.2$$

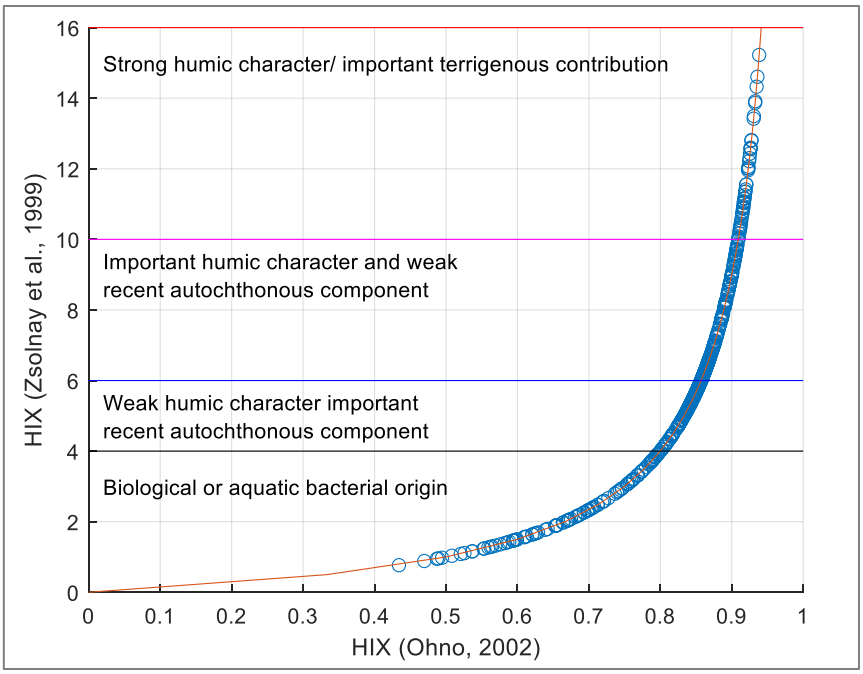


Figure 4.36. HIX values calculated according to Ohno (2002) and Zsolnay et al. (1999). The dataset included samples from La Villette basin, the river Marne, and stormwater networks and WWTP outlets (SO-N, SO-S, WWTP-O).

Table 4.7. DOM characteristics associated with HIX values calculated according to Zsolnay et al. (1999) and Ohno (2002) expressions. Adapted from Huguet et al. (2009).

$HIX_{Zsolnay}$	HIX_{Ohno}	DOM characteristics from HIX	DOM characteristics from BIX	BIX
10-16	0.91-0.94	Strong humic character/important terrigenous contribution	Low biological activity	0.6-0.7
6-10	0.86-0.91	Important humic character and weak recent autochthonous component	Average / Intermediate biological activity	0.7-0.8
4-6	0.80-0.86	Weak humic character and important recent autochthonous component	High biological activity	0.8-1
<4	<0.80	Biological or aquatic bacterial origin	Biological or aquatic bacterial origin	>1

NOTE: The authors use “biological activity” as equivalent to “autochthonous component”.

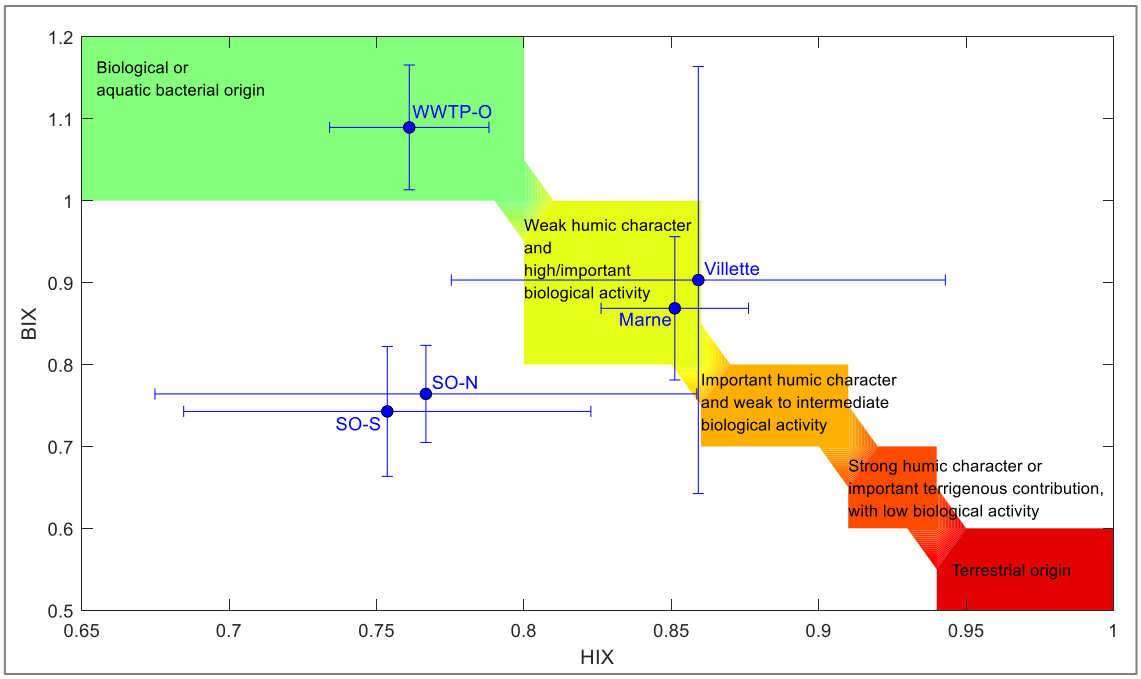


Figure 4.37. DOM characteristics associated with HIX and BIX according to Huguet et al. (2009).

4.2.2.2. DOM character and origin of the water types

The characteristics and origin of DOM can be estimated with the fluorescence indices HIX, BIX (Figure 4.38, Table 4.8).

FI was introduced by McKnight et al. (2001) for the source assessment of humic substances, specifically fulvic acids. FI variability is related to the terrestrial DOM signal (when FI~1.3) (Marcé et al., 2021).

All the samples had FI≥1.8, indicating a predominant microbial or algae contribution on DOM composition. FI did not provide additional discriminating information about DOM composition. Therefore, FI was not part of further analysis. Huguet et al. (2009) already underlined this limitation of FI, affirming that FI is less effective on determining the DOM predominant source in surface natural waters, where organic material originates from a variety of sources that can contain significant amounts of non-humic substances.

In La Villette basin (n=183), HIX (0.86±0.08) and BIX (0.90±0.26) evolution in time presented values, with lower HIX and higher BIX than the average indicating increases of microbial activity.

In the river Marne (n=207), HIX (0.87±0.09), and BIX (0.87±0.09) values were very similar along the stretch, with less variation than in La Villette basin.

The mean HIX values from La Villette basin (0.86) and the river Marne (0.85), were similar to the ones from the streams (0.86) and rivers (0.85) in Berlin area (Romero González-Quijano et al., 2022). BIX values were greater in La Villette (0.90) and the Marne (0.87), than in streams (0.73) and rivers (0.79) of Berlin area. This suggested higher biological activity in La Villette and the Marne than in Berlin watercourses.

In WWTP-O (n=5), a low variation of indices was observed. HIX values were close to 0.80 and BIX values were higher or equal to 1, indicating a mainly biological origin of DOM.

In SO-N (n=9) and SO-S (n=8), a higher variation of HIX and BIX values were observed, with similar ranges. HIX values were grouped in dry and wet weather samples, with lower values

in dry weather, suggesting lower humic character. BIX values were under 1.0, with mean around 0.7 at both sampling points.

In overall, in La Villette basin and the river Marne, HIX values indicated an important humic character, and BIX indicated an intermediate biological character. In both SN and in WWTP outlets, HIX values indicated a very low humic character. In SN outlets, mean BIX indicated average biological character and in WWTP, high biological character (Figure 4.37).

BIX is associated to the biological activity from first stages (around 2 days) of decomposition of organic matter. BIX and HIX values are related to the DOM characteristics: HIX and BIX are complementary and expected to be inversely proportional to each other. Samples with high bacterial activity should have high BIX and present low humic character, associated to low HIX, and vice-versa.

In our samples, however, not all HIX and BIX values were compatible with that assumption. In SN samples, both mean values of HIX and BIX were in the lower range (<0.80 and [0.7-0.8], respectively). According to Huguet et al. (2009) (Table 4.7), HIX<0.80 indicates predominant biological activity, and BIX between [0.7-0.8] and average biological activity.

This limitation indicates that using only one of the indices cannot give correct DOM characteristics. It is necessary to use both indices, HIX and BIX, to estimate DOM characteristics, with HIX primarily indicating the strength of the humic character and BIX the strength of recent autochthonous component or biological activity. The definition of BIX makes it only a partial indicator of biological activity.

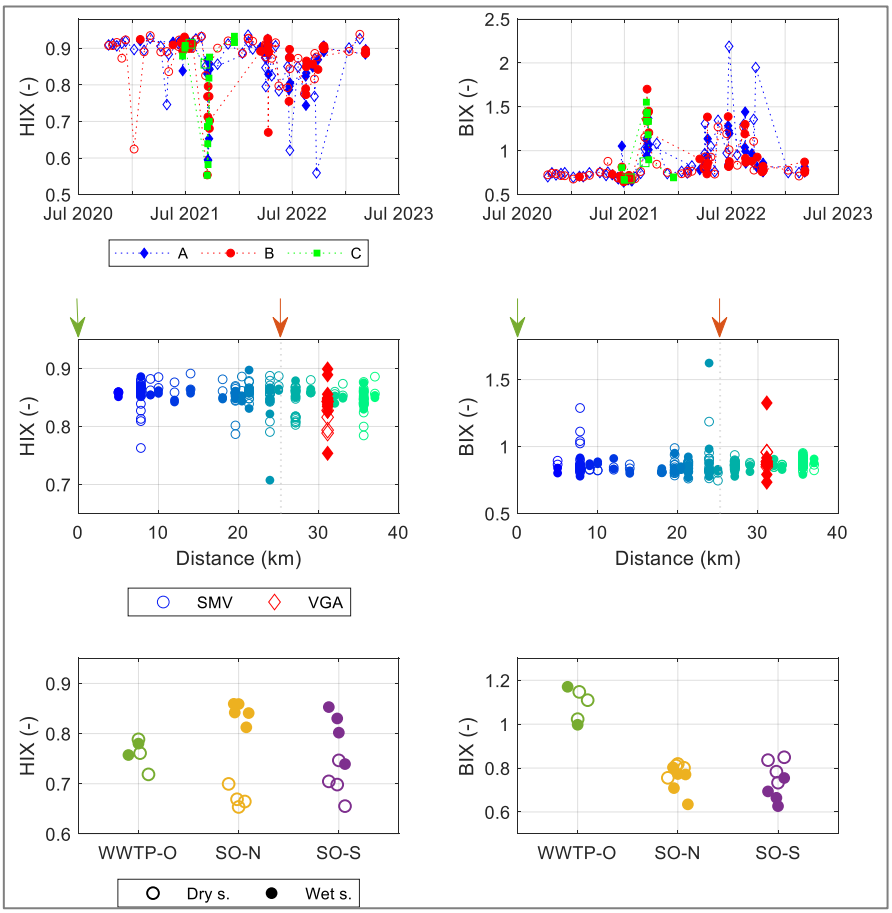


Figure 4.38. HIX and BIX: time evolution in La Villette basin at points A (blue), B (red) and C (green), from October 2020 to March 2023 (top); distance evolution in the river Marne SMV, from upstream to downstream of the stretch, with VGA in red. The green arrow indicates the position of WWTP-O and the orange arrow the outlet of WWTP-PMA (middle); HIX and BIX values in SO-N, SO-S and WWTP-O (bottom). Dry weather samples are open markers and wet weather samples are filled markers.

Table 4.8. Fluorescence indices (mean ± SD and range) of the different sampling sites and the total number of samples (N) per site.

Water type	HIX (-)		BIX (-)		FI (-)	
	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range
Villette (n=183)	0.86 ± 0.08	[0.55-0.94]	0.90 ± 0.26	[0.64-2.19]	1.97 ± 0.07	[1.76-2.28]
Marne (n=204)	0.85 ± 0.02	[0.71-0.90]	0.87 ± 0.09	[0.73-1.62]	2.13 ± 0.08	[1.90-2.40]
WWTP-O (n=5)	0.76 ± 0.03	[0.72-0.79]	1.09 ± 0.08	[1.00-1.17]	2.41 ± 0.09	[2.29-2.53]
SO-N (n=9)	0.77 ± 0.09	[0.65-0.86]	0.76 ± 0.06	[0.63-0.82]	2.25 ± 0.32	[1.93-2.79]
SO-S (n=8)	0.75 ± 0.07	[0.66-0.85]	0.74 ± 0.08	[0.63-0.85]	2.56 ± 0.15	[2.30-2.79]

4.2.3. PARAFAC model results

4.2.3.1. PARAFAC function

The creation of a robust PARAFAC model requires a large number of EEMs with great diversity of DOM sources of diverse chemical quality. However, once the DOM fluorescence signal is decomposed, new EEMs from the same or similar study site, should not interfere in the obtained PARAFAC model. Therefore, these new EEMs should fit well in the existing model (Fellman et al., 2009; Murphy et al., 2011).

In order to obtain the PARAFAC values for future samples collected in the studied waterbodies, we established a function which can be applied to the previously identified PARAFAC components. This function provides the new PARAFAC scores, and therefore Cf, on future samples from the same sampling sites.

With the emission (b_{jf}) and excitation (c_{kf}) modes obtained from the original PARAFAC model, it was possible to obtain the score a_{if} and compute Cf values of new EEMs (Fellman et al., 2009). To this end, the values of a_{if} to obtain the minimum squared error ε_i in the sample *i* are computed by Equation 4.3.

$$\varepsilon_i = \sum_{j=1}^{J} \sum_{k=1}^{K} \left(x_{ijk} - \sum_{f=1}^{F} a_{if}\, b_{jf}\, c_{kf} \right)^2 \qquad 4.3$$

4.2.3.2. PARAFAC components

A model with 7 components was validated by PARAFAC analysis (C1-C7, Figure 4.39). Each component was associated with similar components from OpenFluor Database (Table 4.9). The similarity score (Tucker congruence coefficient) of all the components was >0.95 for emission and excitation spectra loading, with global score >0.90.

C1 to C5 components were identified as humic-like components. C1, C2, C3 and C5 were identified as terrestrial humic-like components. The maximum loadings of C1, C2 and C5 were associated to peak A. C2 displayed a greater emission wavelength (red-shifted). The maximum loading of C3 was similar to peak C.

C4 was associated to humic-like substances produced *in-situ* by phytoplankton and microbial activity. The maximum loading of C4 was in lower emission wavelengths (blue-shifted) than peak C and in greater excitation wavelengths than peak M.

C6 and C7 were identified as protein-like components, close to tryptophan-like fluorescence (peak T). The maximum loading of C7 has greater excitation wavelength than peak T. Compared to peak T, C6 corresponds to a larger window of excitation-emission wavelengths. As the location of the fluorescence peaks can be slightly shifted according to the water matrix, this facilitates the detection of the fluorescence peak.

C6 was similar to an identified component as the main protein-like fluorophore in WWTP effluents. In previous studies (Cohen et al., 2014; Tedetti et al., 2012), C6 displays a high correlation with biogeochemical parameters, such as organic carbon, phosphates and faecal bacteria.

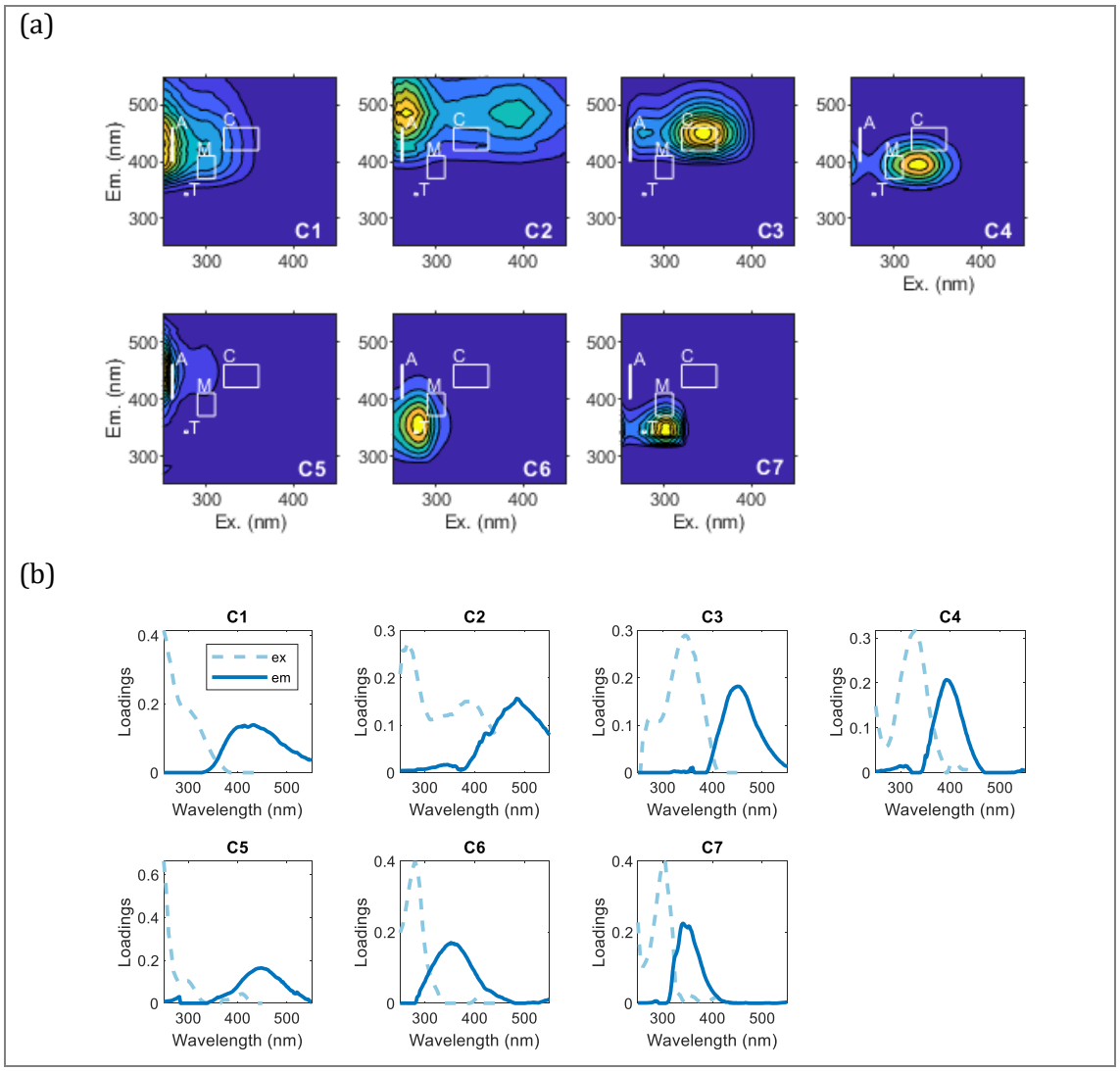


Figure 4.39. EEMs with the loadings of the 7-component PARAFAC model (a) and spectral loadings results (b).

Table 4.9. Designation, excitation (Ex) and emission (Em) wavelengths peaks of PARAFAC components, number of references with matching components from OpenFluor Database (checked on 30/06/2023), with minimum similarity score (Tucker congruence) of 0.95 for excitation and emission wavelength loads (Murphy et al., 2014).

PARAFAC component	Ex (nm)	Em (nm)	%C range	Matches in OpenFluor (0.95)	Description and references
C1	250	432	0.0-44.5	79	Terrestrial humic-like, similar to peak A. Frequently found in lentic freshwater (C4) (Marcé et al., 2021); widespread in freshwaters (C1) (Lambert et al., 2017).
C2	265	484	3.1-26.8	89	Terrestrial humic-like, associated with high aromaticity and molecular weight (C1) (Lambert et al., 2016); plant-derived material (C2) (Yan et al., 2020); associated mainly with peak A.
C3	345	450	6.0-25.1	14	Terrestrial humic-like, similar to peak C, widespread in freshwaters, (C2) (Lambert et al., 2017); associated with aromatic molecules of high molecular weight (C3) (Lambert et al., 2016).
C4	330	392	8.6-26.5	5	Humic-like, low molecular weight, close to peaks C and M; potentially related to agricultural catchments (C6) (Graeber et al., 2012); associated to humic-like substances produced in-situ by phytoplankton and microbial activity (C4) (Galletti et al., 2019).
C5	250	448	0.0-27.6	9	Terrestrial humic-like, estuarine streams of agricultural catchment (C6) (Asmala et al., 2018); associated with peak A (C3) (Li et al., 2015).
C6	280	352	2.7-36.9	44	Protein-like, similar to peak T. Indicative of recent autochthonous production, commonly related to anthropogenic occupation (C7) (Lambert et al., 2017) and WWTP effluents (C1) (Cohen et al., 2014); aliphatic (non-aromatic) with low molecular weight (C6) (Lambert et al., 2016).
C7	305	340	0.0-41.8	21	Protein-like, close to peak T; might be associated with sediment-derived or autochthonous tyrosine, and may be a potential indicator of eutrophication (C3) (Ren et al., 2021); indicator of DOM bioavailability (C4) (Ryan et al., 2022).

The components C1-C7 were represented by Cf, the fluorescence intensity at the maximum loading of excitation and emission modes for each component f.

The components were grouped in terrestrial humic-like (C1, C2, C3 and C5), "low" humic-like (C4), with lower molecular weight, and protein-like (C6 and C7).

The comparison of PARAFAC results among different studies is usually limited by the correspondence of the obtained components, represented by the Tucker coefficient. Similar

water types may not present the same characteristic components, as we observed when comparing our results with similar study sites. This limitation may be overcome when using the relative abundance of protein-like components and humic-like components.

The mean relative abundance of each component *f* (%Cf) was calculated for the five water bodies (Table 4.10). The relative abundance of each component contributes to the signatures of SN and WWTP outlets and of the watercourses La Villette basin and the river Marne.

In La Villette basin, terrestrial humic-like composition was predominant in average (66.0%). The predominant terrestrial humic-like component and with higher variability was C1 (22.4% ± 6.6%), followed by C5 (15.0% ± 6.5%). Protein-like components represent 18.5%. The predominant protein-like component was C7 (11.7% ± 9.6%). It is more abundant and more variable than C6 (6.8% ± 4.0%). The highest variations of C1 and C7 occurred mostly during wet weather (Figure 4.40).

In the river Marne, terrestrial humic-like composition was predominant (65.1%). The predominant humic-like component was C5 (20.8% ± 1.6%). C1 has the highest variability (17.8% ± 2.7%). The protein-like components represent 20.0%. The predominant protein-like component was C6 (11.8% ± 2.3%). The higher variability was associated to C7 (8.2% ± 4.4%) (Figure 4.40).

In WWTP-O (Table 4.10, Figure 4.40), terrestrial humic-like components were the most abundant (52.3%). Protein-like components represented 25.6%. The predominant humic-like components were C5 (23.2% ± 4.2%), terrestrial humic-like, and C4 (21.9% ± 2.0%), "low" humic-like. The predominant protein-like component was C6 (16.5% ± 2.1%). The variability of the components was very low, as evidenced by their standard deviations (ranging from 0.5 to 4.2%) and narrow %Cf ranges ([0.0-26.5%]).

In both SO-N and SO-S (Table 4.10, Figure 4.40), terrestrial humic-like components were the most abundant (62.5% and 60.1%, respectively). Protein-like components represented 25.7% and 25.8%, respectively. The predominant terrestrial humic-like component in SO-N was C1 (24.9% ± 17.8%), and in SO-S, C3 (22.5% ± 1.6%). The dominant protein-like component was C6 in both SO-N (23.4% ± 11.0%) and SO-S (23.6% ± 8.0%). These values are the highest measured in all samples. The components with the highest variability were C1 and C6. Higher values of %C1 and lower %C6 were obtained in wet weather, suggesting that they could be potentially used as signature of local precipitation.

In La Villette basin and the river Marne, the average of relative abundance of all terrestrial humic-like components was around 66%, of "low" humic-like component was 15%, and of protein-like components, 19%. In SN outlets, terrestrial humic-like components corresponded to 61%, "low" humic-like components to 13%, and protein-like components to 26%. In WWTP-O, terrestrial humic-like components represented 52% of components abundance, "low" humic-like component, 22%, and protein-like component, 26% (Figure 4.41).

In SN and WWTP outlets, protein-like components are more abundant than in La Villette and the river Marne. In WWTP-O "low" humic-like component was more abundant than in SN outlets.

Among 9 rivers in Berlin area (Romero González-Quijano et al., 2022), two were considered highly organically polluted due to the presence of WWTP effluents. As in our PARAFAC model, 7 components were obtained, including two protein-like. Similar to La Villette basin and the river Marne, the average of protein-like components was between 10% and 20%. In the two Berlin area polluted rivers, the protein-like components were slightly above 20%, and the terrestrial humic-like components were around 50%. This was similar to the

relative abundance of protein-like and terrestrial humic-like components found in WWTP-O (25.6% and 52.3%, respectively). From this comparison we can deduce that La Villette basin and the studied the river Marne stretch are not organically polluted.

Among our study sites, C6 was high in SN and WWTP outlets, corroborating with a previous study (Cohen et al., 2014). Contrary of what is expected in a stormwater network, a wastewater contamination probably occurs in the studied SN during dry weather, as indicated by the higher presence of C6 in dry than in wet weather samples (Figure 4.40).

In lakes, C6 can also be considered as a potential indicator of eutrophication. In Chinese lakes, a component close to peak T and our C6 was found (Ren et al., 2021). It was associated to pollutant-degrading activity of microalgae, including cyanobacteria. Therefore, in future studies, it will be interesting to explore the sensitivity of C6 to eutrophication indices.

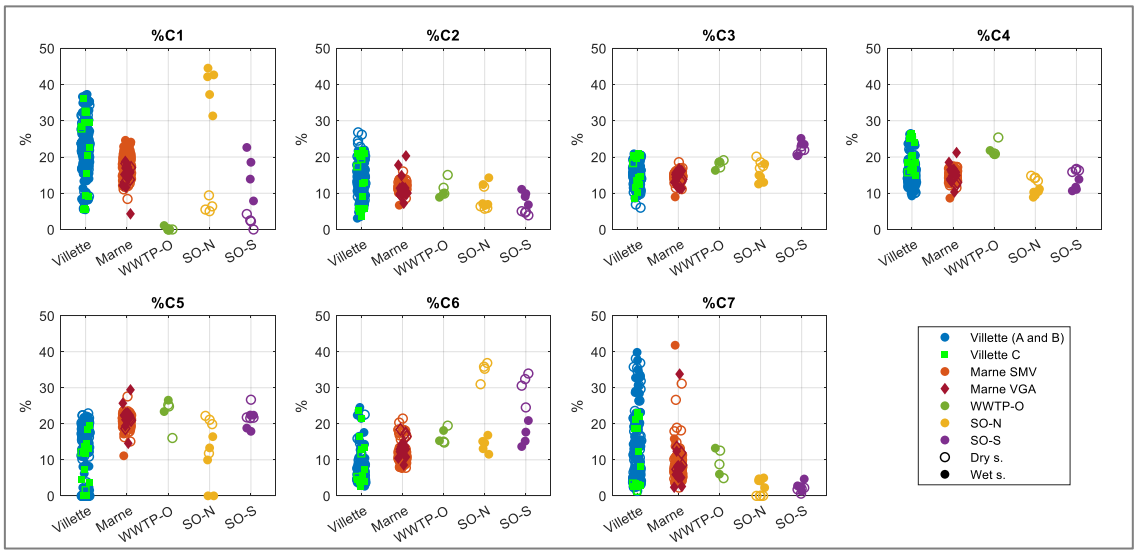


Figure 4.40. Relative abundance of PARAFAC components (%) in La Villette basin points A (blue), B (blue) and C (bright green), in the river Marne points SMV (orange) and VGA (dark red), in the outlet of WWTO-O (dark green) and the outlets from stormwater networks SO-N.

Table 4.10. Relative abundance of PARAFAC components (mean ± SD) of the sampling sites (n=number of samples).

Water type	%C1 (%)	%C2 (%)	%C3 (%)	%C4 (%)	%C5 (%)	%C6 (%)	%C7 (%)
Villette (n = 183)	22.4 ± 6.6	13.9 ± 4.1	14.7 ± 2.9	15.6 ± 3.2	15.0 ± 6.5	6.8 ± 4.0	11.7 ± 9.6
Marne (n = 204)	17.8 ± 2.7	11.5 ± 1.1	15.0 ± 1.0	14.9 ± 1.3	20.8 ± 1.6	11.8 ± 2.3	8.2 ± 4.4
SO-N (n = 9)	24.9 ± 17.8	8.6 ± 3.3	16.3 ± 2.6	11.9 ± 2.3	12.7 ± 8.4	23.4 ± 11.0	2.3 ± 2.3
SO-S (n = 8)	9.0 ± 8.4	6.9 ± 2.7	22.5 ± 1.6	14.1 ± 2.6	21.7 ± 2.6	23.6 ± 8.0	2.2 ± 1.2
WWTP-O (n = 5)	0.2 ± 0.5	11.0 ± 2.4	17.9 ± 1.2	21.9 ± 2.0	23.2 ± 4.2	16.5 ± 2.1	9.1 ± 3.7

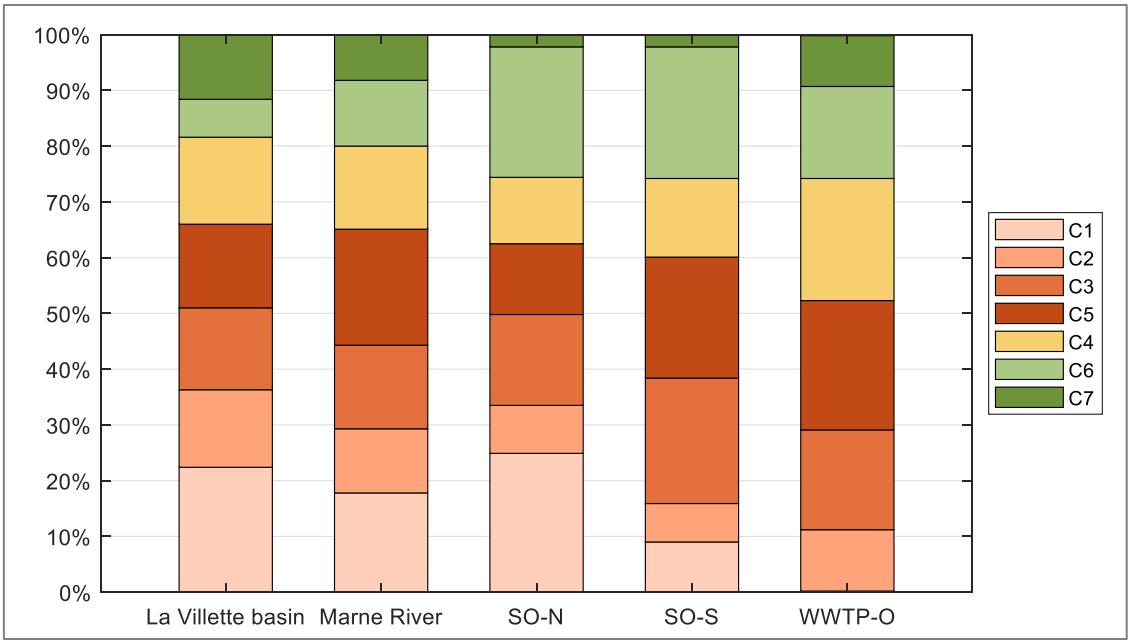


Figure 4.41. Relative abundance of the PARAFAC components. Components were gathered in three groups: terrestrial humic-like (orange), "low" humic-like (yellow) and protein-like (green).

4.2.4. DOM composition among water types

The distribution of the relative abundance of PARAFAC components among the different water types was compared through Kruskal-Wallis test. The average ranks of the median of each component for each water type was computed. The statistical similarity of each component among the water types is presented in Figure 4.42.

Overall, the contribution of all PARAFAC components were very similar in SN and WWTP outlets. The main difference between these groups were the low humic-like component C4 and the protein-like component C7. Both %C4 and %C7 were higher in WWTP-O and significantly different from the SN outlets. In WWTP-O, %C4 was significantly higher from the other water types and %C7 was similar to the watercourses.

The highest %C6 values were measured in SN and WWTP outlets. In contrast, %C6 in the river Marne was greater than in La Villette basin. This may suggest that the outlets which discharge in the river Marne could be the main source of C6 in the river.

Higher variability in the composition of protein-like components could contribute to the detection of change in microbial DOM. In the watercourses, the protein-like component with higher standard deviation and wider range was C7. In SN and WWTP outlets, it was C6 (Table 4.10). This suggests that %C7 could be an indicator of microbial DOM change in La Villette basin and the river Marne.

The median HIX of La Villette basin (0.89) was greater than in the river Marne (0.86). The median BIX of La Villette basin (0.80) was lower than in the river Marne (0.86). La Villette basin and the river Marne HIX medians were significantly different (Figure 4.42), indicating a potentially higher humification level in La Villette basin.

However, HIX and BIX median values between both watercourses were close enough to be part of the same category, indicating a DOM composition with an important humic character and an intermediate biological activity.

Among the samples from SN and WWTP outlets, HIX was very similar. WWTP-O samples presented the highest mean BIX values, being significantly similar from the river Marne samples.

The differences between the variances in La Villette basin and the river Marne was most probably due to the different characteristics of the sampling periods of each site. In La Villette basin, samples were collected for more than two years (October 2020 to March 2023), covering all seasons, and including 38 rainfall events of intensities from [2.6-48.2mm] in 72h. In the river Marne, samples were collected only during Summer 2022, in scheduled dates regardless the weather conditions.

A PCA analysis was performed to analyse the patterns of PARAFAC components and fluorescence indices HIX and BIX among the different water types. It explained 76.4% of the total variance (Figure 4.43). The first axis, PC1, explained 49.2% of total variance. It captured the variation between water types. PC1 was defined positively by C1, C3, and by C6. PC2 explained 27.2% of total variation and was defined positively by C7 and BIX.

The variability of La Villette basin was driven positively by C1, C3 and C6 in axis PC1, and C7 and BIX in axis PC2. In the river Marne, the variability was mostly defined by axis PC2. The variability of SO-N and SO-S were defined by similar drivers, C2 and C6. In WWTP-O, the variability was driven by PC1.

In the plan defined by axis PC1 and PC2, the samples from the watercourses were well separated from SN and WWTP outlets. The dispersion of La Villette basin samples was higher than in the river Marne. It can be due to the longer studied periods (two years vs. one summer). However, both were mostly defined by PC2 and driven by the same components, with an overlapping of the Marne in La Villette basin values. The variability of SN and WWTP outlets was mostly defined by PC1.

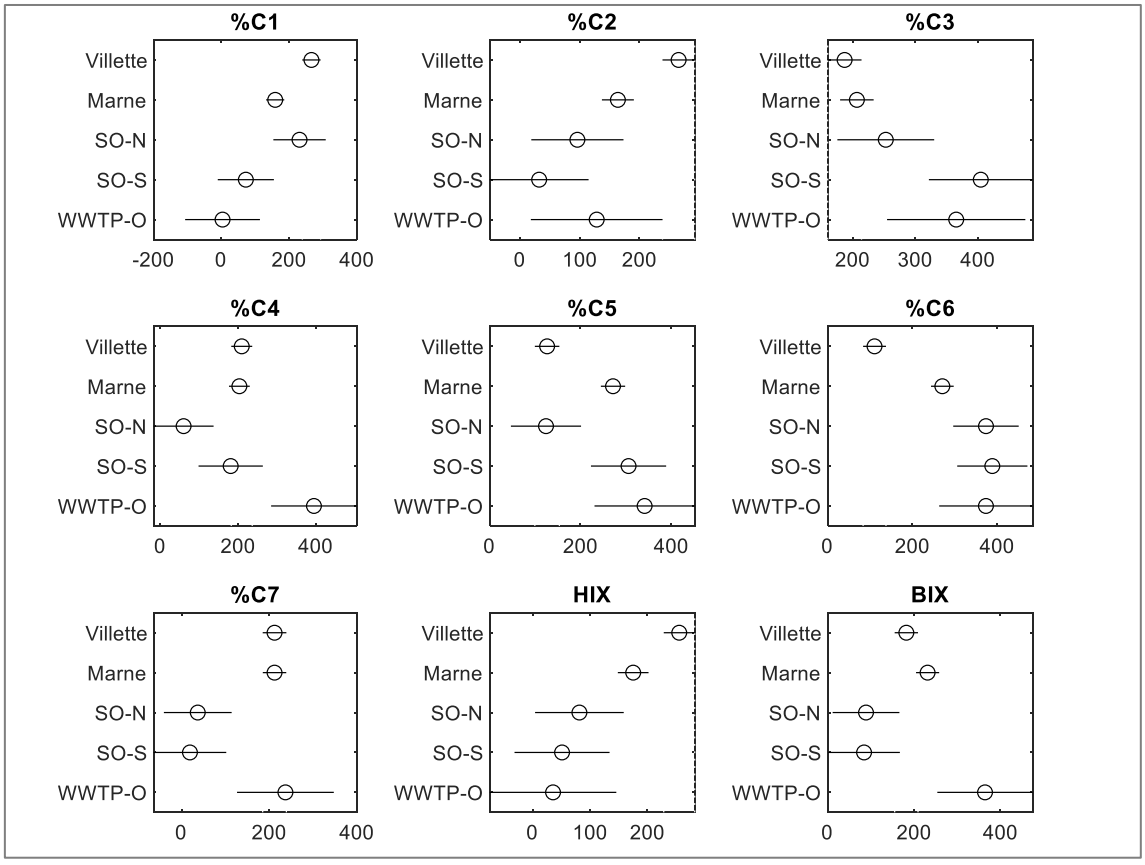


Figure 4.42. Kruskal-Wallis mean ranks of median relative abundance of PARAFAC components (in %) and fluorescence indices HIX and BIX of all water types.

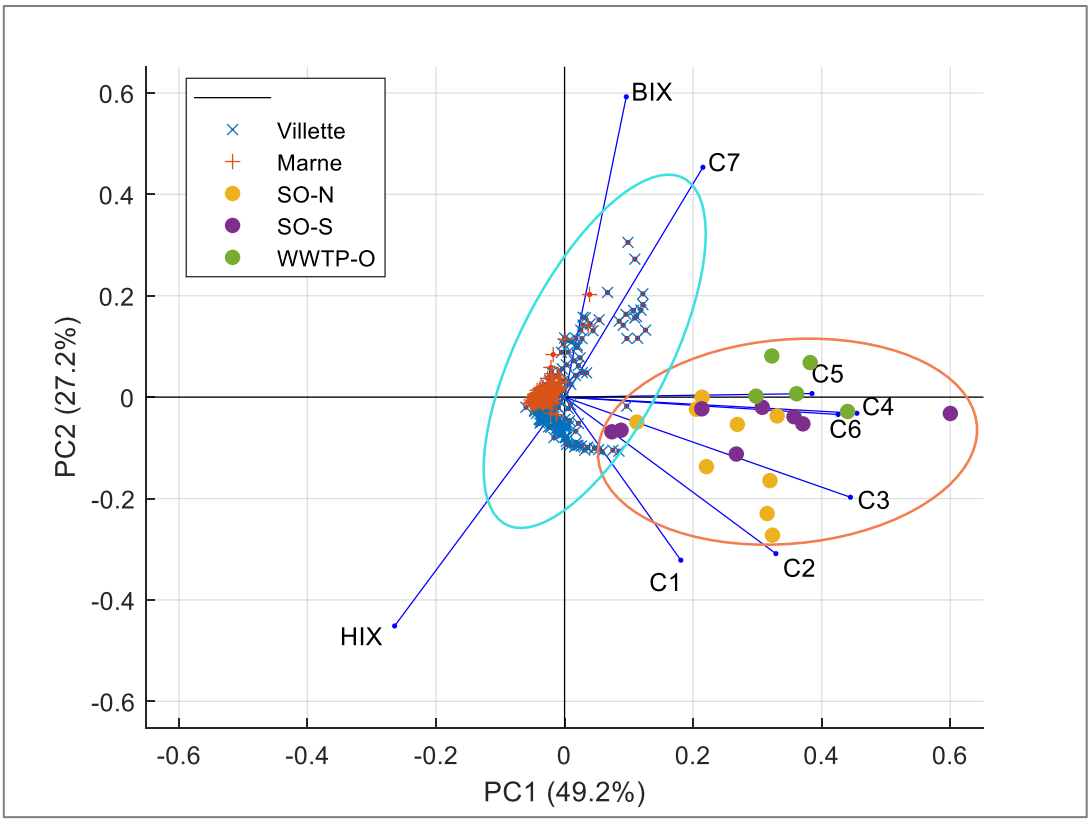


Figure 4.43. PCA of PARAFAC components, HIX and BIX, with dataset of La Villette basin (blue, n=183), the Marne (orange, n=207), SO-N (yellow, n=9), SO-S (purple, n=8), and WWTP-O (green, n=5). Blue ellipse groups watercourses samples (La Villette basin and the river Marne) and orange ellipse the SN and WWTP outlets.

4.2.5. Impact of land-use in the river Marne

In the river Marne samples, the influence of land-use on fluorescence components and their variation along the river stretch, upstream to downstream, were investigated. Samples collected at the same day were compared. A total of 15 sampling campaigns were conducted in the main 6 SMV points (SMV1, SMV6B, SMV7C, SMV8B, SMV10 and SMV14B) (Figure 4.44).

Two WWTP outlets are located in the studied stretch. The most upstream, WWTP-O, is located 7.8 km upstream SMV1. The second, WWTP-PMA, is 1.8 km upstream SMV10. The distance between the sampling points and their nearest WWTP outlet at upstream was considered (DW): the distance between WWTP-O and SMV1, SMV6B, SMV7C, SMV8B; and the distance between WWTP-PMA and SMV10, SMV14B (Table 4.11).

The mean values of the relative presence of PARAFAC components and the fluorescence indices were very close among all the sampling points.

Among the terrestrial humic-like components, the distribution of %C1 was not significantly different among all 6 SMV points, from upstream to downstream.

Regarding the statistical similarity of the protein-like components, some differences appeared.

%C6 was higher at SMV6B than SMV1 (Figure 4.45). The distance between SMV1 and SMV6B is the longest among the sampling points (11.8 km). It includes 26 stormwater network outlets, including SO-N. The higher %C6 level at SMV6B could be due to larger outlet loading.

%C6 was also slightly higher at SMV10 than SMV8B (Figure 4.45). Between SMV8B and SMV10 is located the outlet of WWTP-PMA, suggesting an impact of the WWTP on the protein-like DOM composition.

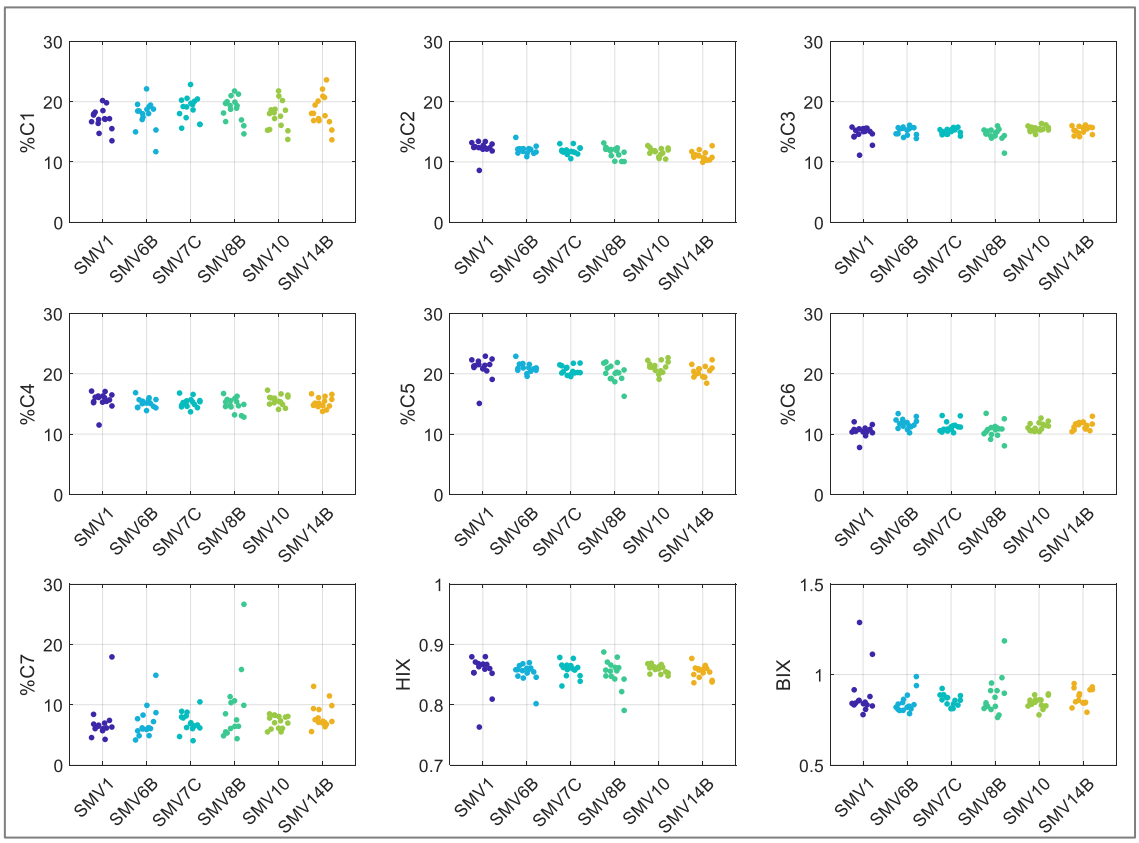


Figure 4.44. Swarm chart of %C1 to %C7 of the main group of 6 SMV sampling points along the river Marne.

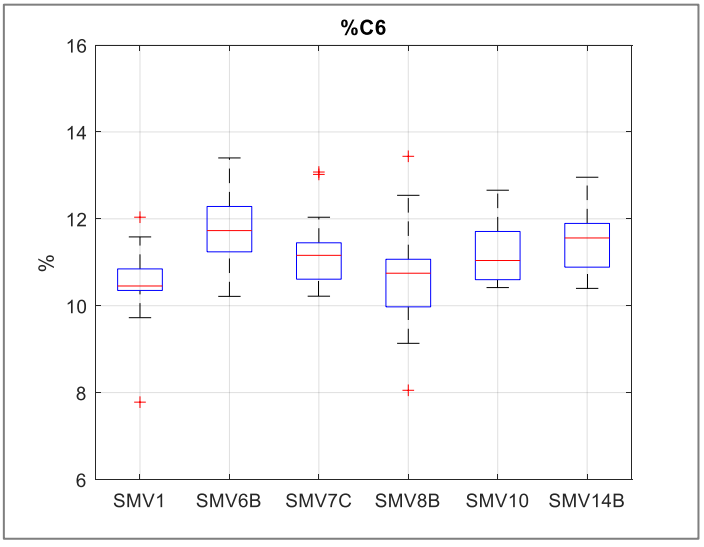


Figure 4.45. Boxplot of %C6 of the main group of 6 SMV sampling points along the river Marne.

Table 4.11. Fluorescence indices (mean ± SD) in the main group of 6 points along the river Marne. Distance from the nearest WWTP outlet upstream of each sampling point (DW).

Sampling point (n=15/ point)	DW (km)	%C1	%C2	%C3	%C4	%C5	%C6	%C7	HIX	BIX
SMV1	7.8	17.2 ±1.8	12.3 ±1.1	14.7 ±1.2	15.6 ±1.3	21.0 ±1.9	10.5 ±0.9	8.7 ± 7.0	0.85 ±0.03	0.89 ±0.13
SMV6B	19.6	17.8 ±2.4	12.0 ±0.7	15.1 ±0.7	15.2 ±0.8	21.0 ±0.8	11.7 ±0.8	7.1 ± 2.7	0.85 ±0.02	0.84 ±0.06
SMV7C	21.3	18.9 ±2.0	11.9 ±0.6	15.1 ±0.5	15.2 ±0.8	20.6 ±0.7	11.2 ±0.9	7.1 ± 1.7	0.86 ±0.01	0.86 ±0.03
SMV8B	23.9	18.8 ±2.0	11.6 ±0.9	14.6 ±1.0	14.9 ±1.2	20.1 ±1.5	10.7 ±1.3	9.3 ± 5.7	0.85 ±0.02	0.88 ±0.11
SMV10	1.8	17.7 ±2.3	11.7 ±0.7	15.5 ±0.5	15.6 ±0.9	21.1 ±1.0	11.2 ±0.7	7.1 ± 1.1	0.86 ±0.01	0.85 ±0.03
SMV14B	10.3	18.5 ±2.6	11.0 ±0.7	15.3 ±0.6	15.2 ±0.9	20.3 ±1.0	11.4 ±0.7	8.2 ± 2.0	0.85 ±0.01	0.87 ±0.05

A Principal Component Analysis (PCA) was computed including all seven PARAFAC components, the distance from the closest upstream WWTP outlet (DW) and the land cover in the 1km^2 upstream each sampling point, grouped in “urban green areas”, “natural green areas” and “impervious areas” (Figure 4.46).

The first two principal axis explained, together, 53.8% of total variance of the dataset. The first axis, PC1, explained 34.5% of the total variance. It was predominantly defined by the positive loadings of the PARAFAC components 1 to 6, and the negative loadings of distance from the nearest upstream WWTP outlet (DW).

PC1 captured the variation of fluorescence components in sampling sites and the distance from a WWTP outlet.

The second axis, PC2, explained 19.3% of total variance. It positively defined by BIX, C7 and the impervious areas land-use, and negatively defined by urban green areas and HIX.

Regarding the fluorescence components, C1 to C6 were closely distributed and orthogonally to C7. C1 to C6 were inversely related to the distance from a WWTP outlet upstream, suggesting that the lower the distance, the higher the loadings of these components.

The influence of the land use only appeared slightly. It could be suggested that impervious areas, more related to BIX and C7, may have higher microbial loading, and urban green areas, more related to HIX, may have higher DOM humification level.

These results are consistent with previous studies. More urbanized areas have a more microbial fluvial organic matter composition, resulting in higher BIX values, and less plant-soil-derived character, with lower HIX (Lambert et al., 2017).

The relationship between microbial loading and impervious areas may origin from surface runoff, composed by different sources of pollution, and from WWTP effluents (Romero González-Quijano et al., 2022).

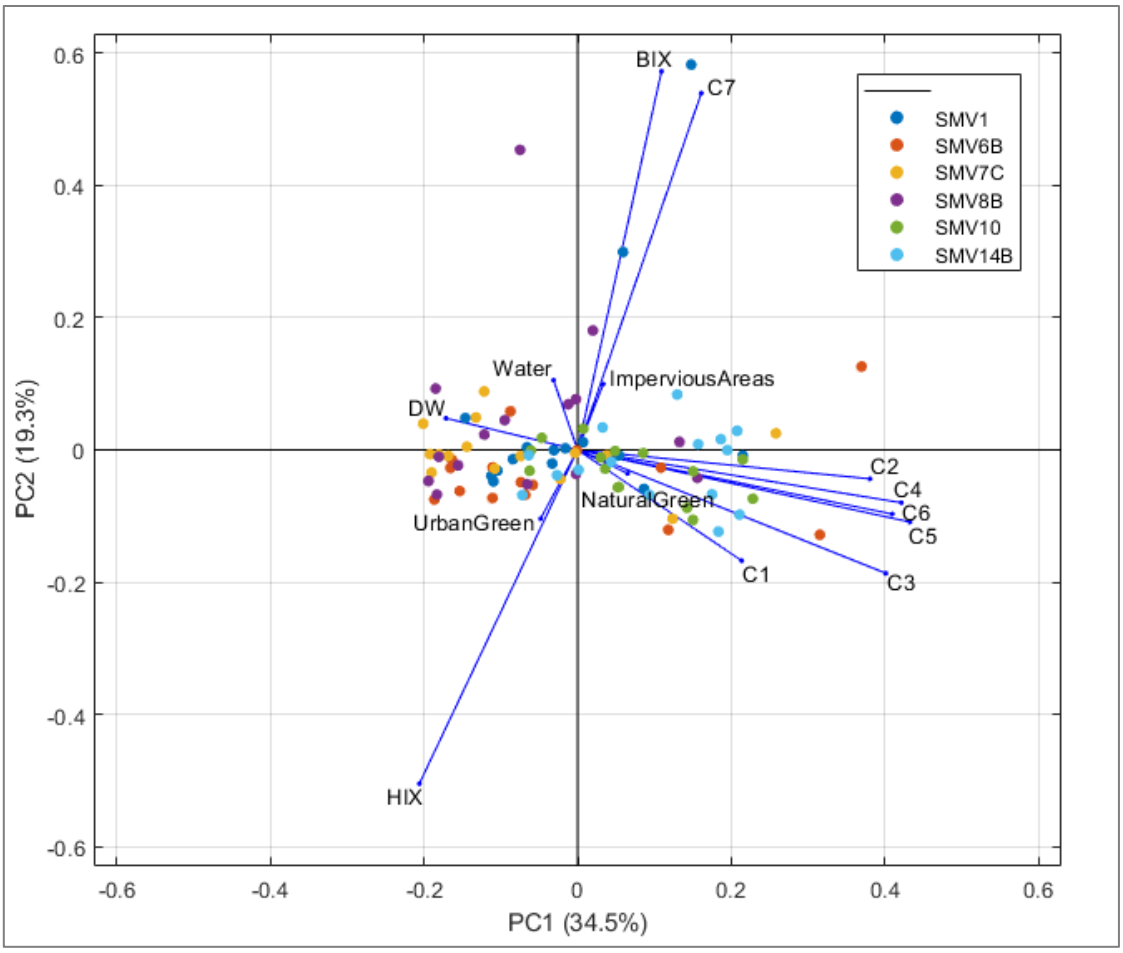


Figure 4.46. PCA of the main group of 6 SMV Marne River points with C1-C7, land-use and distance from the nearest WWTP outlet upstream (DW).

4.2.6. Rainfall impact on DOM composition in La Villette

The seasonal variation of fluorescent DOM was explored during dry and wet events in La Villette basin. Kruskal-Wallis test was performed for assessing differences among samples collected over 2.5 years (October 2020 to March 2023), 85 samples during dry weather and 98 during wet weather (n=183).

The relative contribution of the PARAFAC components, HIX and BIX were compared to identify a potential signature of local contamination due to stormwater network or WWTP outlets or runoff during rainfall events. No significant differences were observed between dry and wet weather for %C3, %C5, %C7 and BIX (Figure 4.47).

Significant differences were identified in %C1, %C2, %C4, %C6 and HIX. In wet weather, the humic-like components %C1, %C2 and HIX were lower than in dry weather. The values of %C4, a component related to microbial activity (Galletti et al., 2019), and %C6, a protein-like component, were significantly higher. The increase of the components C6 and C4 suggests an increase of microbial loadings during rainy events.

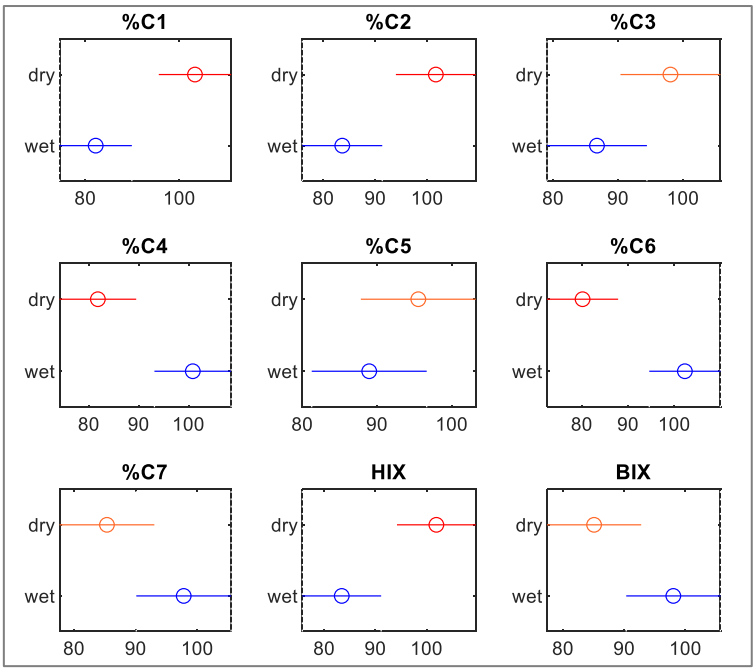


Figure 4.47. Mean ranks of Kruskal-Wallis test of relative PARAFAC components, HIX and BIX, for comparison between dry (n=85) and wet (n=98) weather samples.

In the period of bathing activities, from the 1st of June to 30th of September, a comparison between fluorescence components in dry and wet weather samples was conducted. The results showed a non-significant difference in the protein-like fluorescence components in dry and wet weather (Figure 4.48).

The results suggested that: either the protein-like components C6 and C7 were not very sensitive to local runoff during rainfall episodes; or that local runoff does not affect the microbial loadings in La Villette basin.

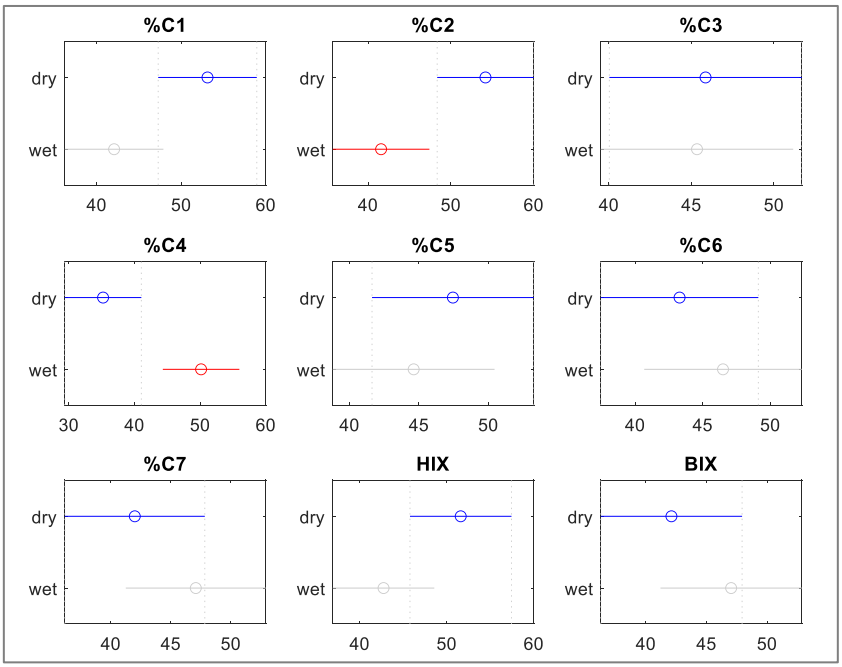


Figure 4.48. Mean ranks of Kruskal-Wallis test of relative PARAFAC components, HIX and BIX, for comparison between dry (n=28) and wet (n=62) weather samples in bathing periods of 2021 and 2022 (from 01-June to 30-September). The significantly different components are in red.

Generally, humic-like fluorescence is strongly controlled by hydrological processes, whereas protein-like fluorescence is more tightly linked with biological processes,

depending of water temperature, sunlight intensity, water turbidity, etc. (Fellman et al., 2010; Meng et al., 2013; Parlanti et al., 2000). However, in urban catchments, local rainfall may have a significant impact not only in humic composition but also in the protein-like components.

In rivers, the results are contrasted. In Oubangui River (Central African Republic), the humic character was compared between dry season, or low-water season, and wet season (Lambert et al., 2016). Lower humic character was found during wet season.

Higher humic character was found during high-water periods in the catchments of the Seine River (France) (Nguyen, 2014), the Congo River (Democratic Republic of Congo) (Lambert et al., 2016) and in Florida Bay (USA) (Fellman et al., 2010).

Regarding the protein-like components, in Zhujiang River (China), they were lower during the rainy season. This lower level was probably due to a dilution effect of higher discharge, combined to intermittent strong sunlight irradiation causing photodegradation (Meng et al., 2013).

The impact of local rainy events on DOM composition was measured in karstic aquifers (Frank et al., 2018b). An increase of all components was observed.

In La Villette basin, the wet weather samples presented lower humic character (lower HIX and lower presence of terrestrial humic-like components) and higher microbial loading (higher BIX and higher presence of protein-like components) than the dry weather samples.

The increase of the protein-like components C6 and C7 in wet weather samples was more important than the increase of humic-like components, resulting in the increase of the relative contribution of protein-like components.

It must be noticed that La Villette basin is a highly controlled hydraulic system, with many sluices upstream and one sluice downstream. The variation of discharge is very limited, even during heavy rainfall episodes. The impact of wet weather on humic-like components is therefore reduced. The loading of protein-like components from direct runoff and close stormwater networks is dominant. The increase of the protein-like components and of BIX suggested that these FDOM values, representing microbial loadings in La Villette basin, were able to detect the impact of local rainy events.

To have an improved assessment of rain impact on FDOM in La Villette basin, it would be required to sample before and after rainy events every at least 5h, for four days. Rainfall characteristics (intensity, duration, return time…) would be required in order to be associated to C6 increase (potential contamination) or decrease (dilution effect). Turbidity and COD may also be measured to provide additional information to improve the relationship between FIB and FDOM.

The origin (terrestrial or microbial) of the main components of the studied urban waters, at each sampled site, was identified through the analysis of the fluorescence of dissolved organic matter (FDOM). The validated 7-component PARAFAC model was effective in characterizing the different types of sampled water.

Components C6 and C7, as well as BIX, were identified as indicative of microbial loadings. These protein-like components may serve as valuable markers for the assessment of microbial contamination in urban water samples.

In La Villette basin, the impact of local rainfall on water microbiological quality was highlighted. The FDOM values demonstrated a significant increase in microbial loading following rainy periods. FDOM analysis can provide a sensitive technique for detecting changes in water quality after rainfall events.

In SN and WWTP outlets, component C6 exhibited high values. In agreement with previous studies, protein-like fluorescence components, such as C6, could be used as effective indicators for monitoring microbiological water quality. An increase of C6 level would suggest a microbial contamination and deserves further attention for managing water resources.

Finally, our results illustrated the value of the PARAFAC model in characterizing urban water types. They highlighted the importance of FDOM analysis for assessing microbial loading and water quality. The identification of protein-like fluorescence components, especially C6, provided valuable insights for microbiological water quality monitoring. It may help in the development of effective strategies for microbiological water quality monitoring.

4.3. Estimation of faecal indicator bacteria

This chapter presents part of the results of a paper submitted for publication in *Frontiers in Water*, named *"Fluorescence spectroscopy for tracking microbiological contamination in urban waterbodies"*, presented in Annex V.

4.3.1. Introduction

The estimation of faecal bacteria in waterbodies is of significance importance. Faecal indicator bacteria (FIB), such as *E. coli* and intestinal enterococci (IE), are indicators of potential waterborne pathogens, harmful for humans. Concerning bathing activities, a high level of FIB in water increases the risk of waterborne diseases. To avoid them, water quality standards and regulations often include limits for faecal bacteria concentrations in recreational waters (EU, 2006).

In urban waters, the increase of faecal pollution can be associated with heavy rainfall and floods, when runoff pollution is washed into the watercourses. In addition, stormwater network outlets may present problems like infiltration or incorrect connection of wastewater. Therefore, large volumes of stormwater can increase the discharge of combined sewage systems outlets and corrupted stormwater network outlets, discharging diluted wastewater into the watercourse (EEA, 2018).

Predictive models help regulatory agencies and bathing area managers make informed decisions based on real-time (nowcasted) and forecasted FIB levels. Preventing measures, such as frequent testing, targeted beach closures, or improved waste management practices, would contribute to sanitary risk management of the bathing area.

To that mean, two approaches were explored for FIB estimation: (1) nowcasting through fluorescence components and (2) forecasting through hydrometeorological variables.

The first approach focused on a specific type of component potentially related to faecal contamination, the protein-like fluorescence components. The results of the relationship between *E. coli* and the PARAFAC protein-like fluorescence components of section 4.2 are presented.

The second approach focused on the increase of FIB during and after heavy rainfall events. The relationship between FIB and the hydrometeorological variables flow rate, precipitation and duration of dry weather were explored.

In this section, the information provided by FDOM about microbiological water quality for monitoring purposes is explored. Also, a potential signature of local contamination due to stormwater network (SN) and WWTP outlets or runoff during rainfall events is identified. A link between protein-like components and FIB concentrations from the dataset of La Villette basin (point C), the river Marne (VGA), SN outlets (SO-N and SO-S) and WWTP outlet (WWTP-O) is presented.

Finally, the relationship between FIB and hydrometeorological aspects to predict FIB concentration in La Villette basin ss presented. To that mean, FIB data from La Villette basin at point C was related to daily mean values of flow rate, precipitation and duration of dry weather.

4.3.2. FIB data

The enumeration of FIB, *E. coli* and IE, was performed in samples from all water types. The sampling points were La Villette basin point C (n=10), the river Marne VGA (n=18), WWTP-O (n=5), SO-N (n=9) and SO-S (n=8). The same sampling protocol was applied to all samples (n=50). The log_{10}-transformed values are presented in Figure 4.49.

As described in section 3.3, dry and wet weather samples were collected in all sampling points. In La Villette basin, wet weather samples were collected during the first day of a rainy event and during the two following days. In the river Marne, wet weather samples were collected during the rainy event and the three following days. In WWTP and SN outlets, the wet weather samples were collected during the first day of the rainy events.

The order of magnitude of the FIB values varied among the sampling sites. For *E. coli*, in La Villette basin, it varied by a factor between 2 and 3, in the river Marne VGA, between 2 and 4, in WWTP-O, between 4 and 5, in SO-N, between 5 and 6, and in SO-S, between 5 and 7.

The order of magnitude of IE, in La Villette basin, it varied by a factor between 0 and 2, in the river Marne VGA, between 0 and 3, in WWTP-O, between 3 and 4, in SO-N, between 4 and 5, in SO-S, between 4 and 6. In WWTP-O, higher FIB values were observed in wet than in dry weather (Table 4.12).

In SO-N and SO-S, however, higher *E. coli* values were observed in dry weather than in wet weather, suggesting a constant contamination on the network (sewage infiltration), with dilution in wet weather. In both sites, the difference between IE values in dry and wet weather samples was less marked.

Among all samples, *E. coli* and IE had a strong relationship (R^2=0.88, p-value<0.001) (Figure 4.50). This result reinforced the use of FIB as indicators of faecal contamination.

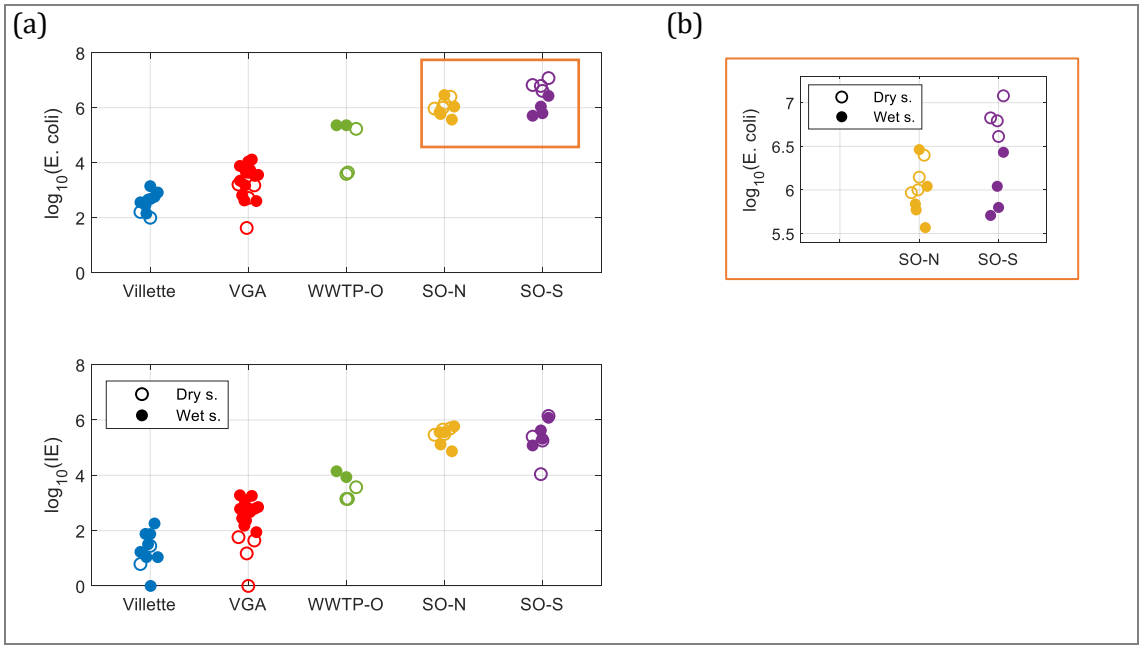


Figure 4.49. (a) log-10 transformed FIB dataset in MPN.100mL^{-1} of the samples from: La Villette basin point C, the Marne River VGA, WWTP-O, SO-N and SO-S; (b) zoom-in in SO-N and SO-S *E. coli* data.

Table 4.12. FIB range, median values for all samples, for dry weather samples and for wet weather samples, and sampling period of samples collected in the five sampling sites.

Sampling site (dry+wet weather s.)	Sampling period	Median *E. coli* (MPN.100mL^{-1})				Median IE (MPN.100mL^{-1})			
		Range	All s.	Dry s.	Wet s.	Range	All s.	Dry s.	Wet s.
La Villette basin point C (n=2+8)	30/06/2021 to 22/09/2021	98 – 1400	405	129	465	0 – 180	21.5	16.1	23.5
Marne River VGA (n=4+14)	15/06/2022 to 20/10/2022	41 – 13 000	2750	1007	3950	0 – 1900	375	28.5	605
WWTP-O (n=3+2)	30/06/2021 to 04/10/2021	3900 – 230 000	170 000	4 400	230 000	1400 – 14 000	3 700	1400	11 300
SO-N (n=4+5)	04/05/2022 to 09/09/2022	370 000 – 2 900 000	1 000 000	1 200 000	690 000	74 000 – 590 000	360 000	380 000	360 000
SO-S (n=4+4)	19/09/2022 to 07/11/2022	510 000 – 12 000 000	3 400 000	6 450 000	865 000	11 000 – 1 400 000	230 000	215 000	315 000

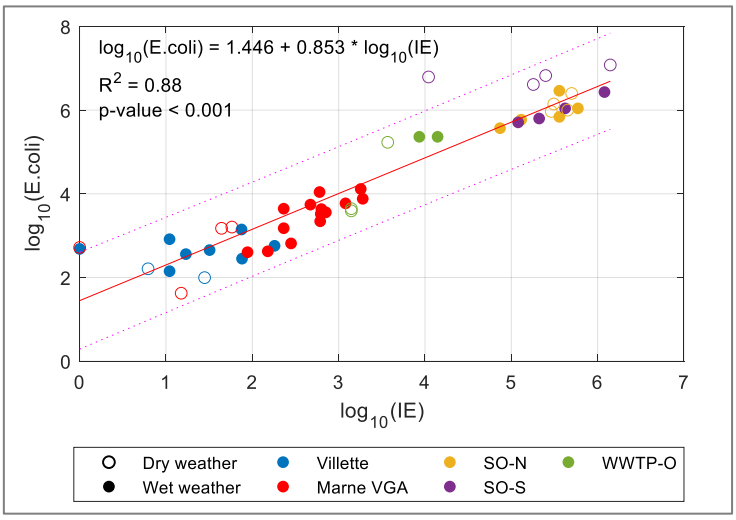


Figure 4.50. Relationship between log10-transformed *E. coli* and IE (n=50).

4.3.3. Fluorescent dissolved organic matter components as indicators of faecal contamination in urban waters

4.3.3.1. Relationship between fluorescent dissolved organic matter and faecal indicator bacteria

The correlation between log10-transformed FIB and PARAFAC components was explored using all FIB data (n=50). We expected to obtain a good relationship with the relative abundance of protein-like components, C6 or C7.

A significant relationship with $\log_{10}$(C6) with both FIB, *E. coli* (R^2=0.72, Figure 4.51-A) and IE (R^2=0.61, Figure 4.51-B), was obtained (p-value<0.001).

According to this relationship, considering the *E. coli* threshold for sufficient bathing water quality (900 MPN.100mL^{-1}), an indicative warning level could be based on C6<0.13 RU.

The relationships between C7 and both FIB were not significant (p-value>0.5). Therefore, they were not considered.

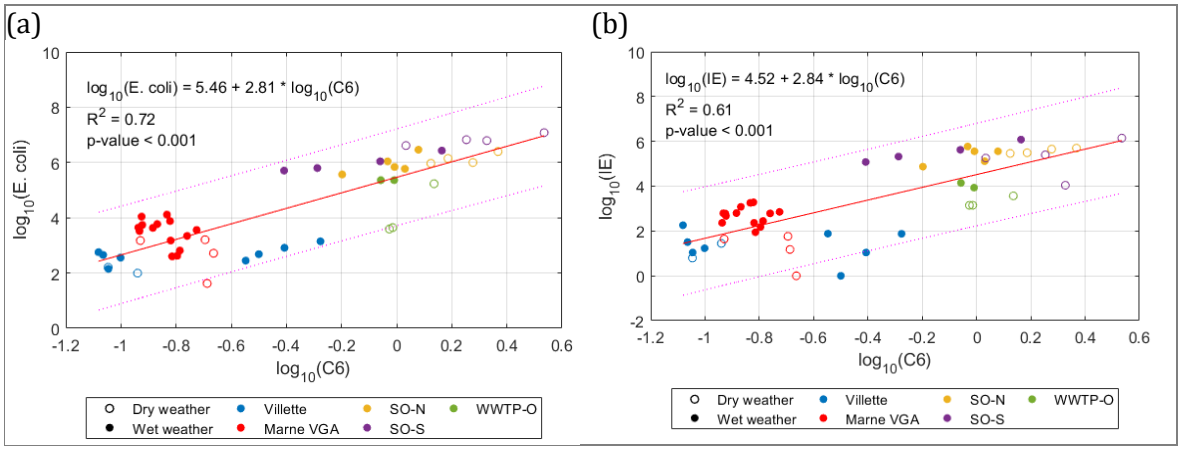


Figure 4.51. Linear relationship between log_{10}(C6) and: (a) log_{10}(*E. coli*) (n=50), (b) log_{10}(IE) (n=50).

In 2023, additional Marne River samples were collected (n=16), close to SMV8B (point IMP). The samples were collected from 25/05-03/07/2023, following the same sampling protocol as in point VGA (24-hour integrated samples during dry and wet weather). C6 values were obtained from the PARAFAC function presented in section 4.2.3.1. The relationship between *E. coli* and C6 was applied for *E. coli* estimation (Figure 4.52-a).

The median of *E. coli* of 2023 samples was 1500 MPN.100mL^{-1}. The estimated *E. coli* median was 1378 MPN.100mL^{-1}, very close to the observed. The distribution of observed and estimated *E. coli* was very similar (Figure 4.52-b). This result contributes to the relationship validation.

The observed *E. coli* values can be used to determine the overpassing of the bathing threshold (900 MPN.100mL^{-1}). From 25/05 to 03/07/2023, 11 out 16 occurrences (68%) were consistent, either higher or lower than the threshold. Of the five incorrect estimations, three overestimations were found and two underestimations.

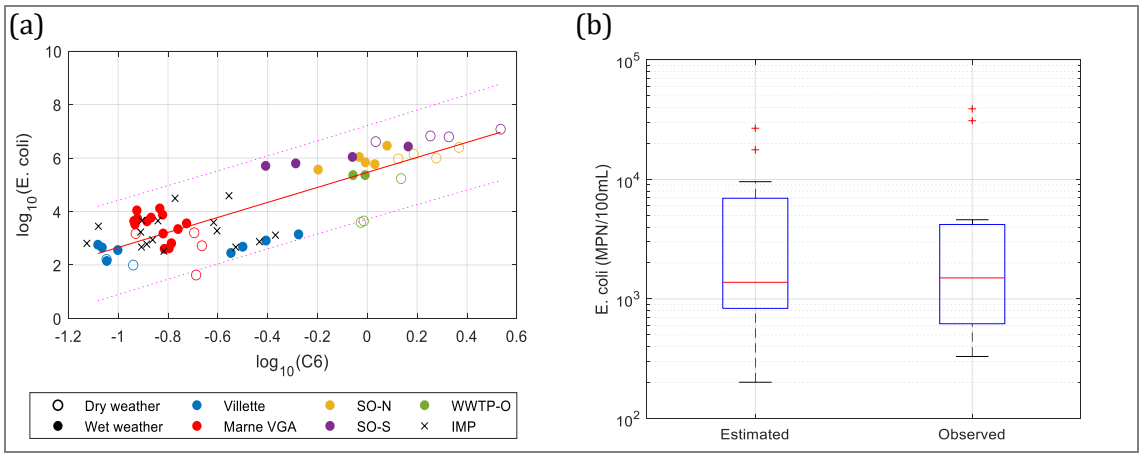


Figure 4.52. (a) *E.* coli and C6 relationship with 2023 additional values superimposed (black x); (b) Estimated *E. coli* from C6 values and observed from additional samples (n=16).

4.3.3.2. Estimation of FIB concentrations from fluorescence intensity

The relationship between FIB concentration and the protein-like PARAFAC component C6 enabled a rapid estimation of *E. coli*. As an application example, the temporal evolution of *E. coli* in La Villette canal was estimated using fluorescence values obtained during regular

field campaigns from June to September 2021 and 2022 (Figure 4.53). The values were compared with the range of Fluidion observations, measured 28/06 to 21/08 in 2021 and 2022.

In summer 2021, 22 samples were collected between 07/06 and 22/09/2021. In this period, 5 of the estimated *E. coli* overpassed the threshold of 900 MPN.100mL^{-1}, representing 22.7% of the total samples. During the period of Fluidion daily measurements, 8 samples were collected. In all of them, the estimate *E. coli* was within the range of observed *E. coli* from Fluidion before and after fluorescence sampling. Particularly, on the 16/07/2021, a threshold exceeding was detected by both estimated and observed *E. coli*.

In summer 2022, 17 samples were collected between 15/06 and 27/09/2022. In this period, 7 of the estimated *E.* coli overpassed 900 MPN.100mL^{-1}, representing 41.2% of the total samples. During the period of Fluidion measurements, 6 samples were collected. The *E. coli* from the relationship were overestimated in relation to the observed Fluidion values. In three samples, the upper limit of estimated *E. coli* exceeded 900 MPN.100mL^{-1}, while the three daily Fluidion values before and after the sampling were lower.

In overall, the variability of the daily observed values from Fluidion were covered by the upper and lower values of estimated *E. coli*. It is important to highlight that estimated and observed *E. coli* were obtained from samples taken in different times. The time between samplings limits the comparison of the obtained *E. coli*.

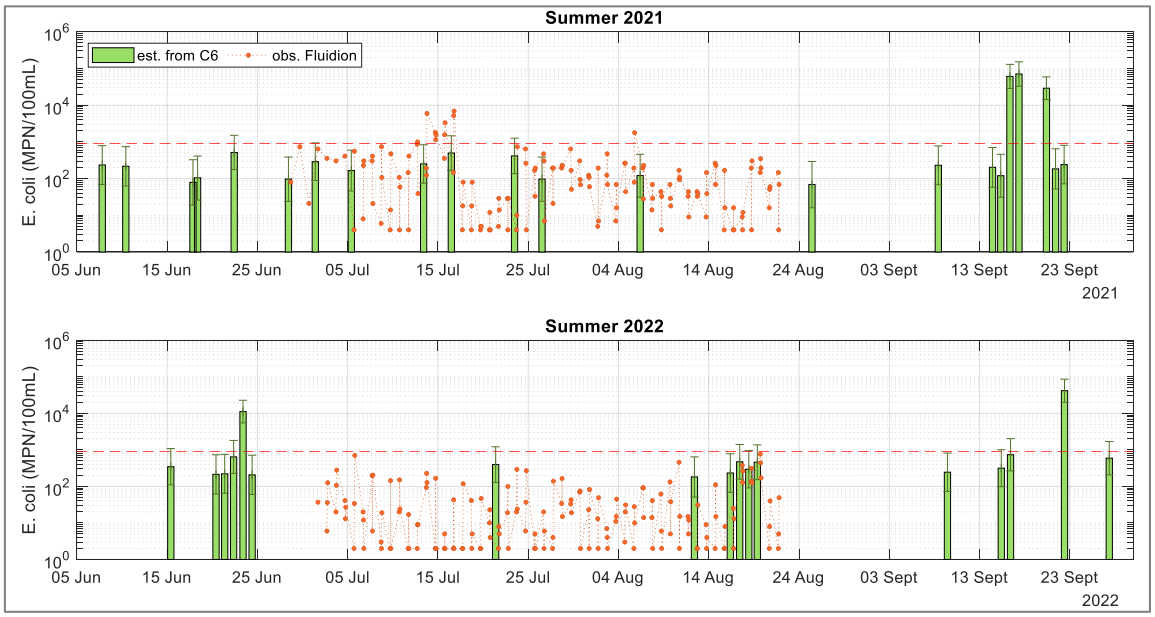


Figure 4.53. La Villette canal: evolution of estimated *E. coli* from protein-like PARAFAC component C6 (green) and observed *E. coli* from Fluidion sensor (orange) at point A (upstream), in summer 2021 (top) and summer 2022 (bottom).

4.3.3.3. Discussion

The link between *E. coli* and protein-like components identified as peak T has been explored for several years. The type and accuracy of the obtained relationships were very variable according to the study sites (Table 4.13).

The matrix composition of water in each sampling site may change among different catchment areas. Since humic-like fluorescence may mask peak T fluorescence intensity at low levels (Ward et al., 2021, 2020), the use of peak T is more suitable to assess moderate to high levels of faecal contamination. In addition, other compounds can emit peak T

fluorescence, increasing the uncertainty of peak T as a proxy of bacterial concentrations (Baker et al., 2007; Bedell et al., 2022; Sorensen et al., 2016).

In six rivers in South Africa, Baker et al. (2015) obtained a relationship between *E. coli* and peak T, expressed in ppb tryptophan (n=136, r=0.75). In two urban rivers, one major river and one sewage treatment effluent in UK, Cumberland et al. (2012) obtained a relationship between *E. coli* and peak T, expressed in Arbitrary Units (AU) (n=46, r=0.87).

The use of PARAFAC analysis to find the characteristic components to our study sites allowed to identify a protein-like component very close to peak T, the component C6. It was possible to obtain a local relationship between C6 and *E. coli*, associated to our study sites (n=50, r=0.85). More data of FIB and C6, with different FIB levels, will allow to have more statistically reliable relationship, particularly for higher levels of contamination.

This relationship was applied to additional samples from the river Marne, collected in 2023. The distribution of estimated and observed *E. coli* were close (Figure 4.52). This result opens the way to the application of relationship between C6 and *E. coli* to determine, from FDOM measurement, the overpassing of the *E. coli* threshold (900 MPN.100mL^{-1}) for bathing. It would be useful as a complementary method to provide additional data for the development of bathing water profiles, for example.

Table 4.13. Summary of previous studies with relationships between *E. coli* and tryptophan-like fluorescence (peak T), including the present study. The relationship type "binary classification" is established for the overpassing of a defined threshold.

Water type	n	Relationship type	Correlation factor/ accuracy	Reference
River and sewage treatment effluent (West Midlands, UK)	46	Linear	r=0.87	(Cumberland et al., 2012a)
River (KwaZulu Natal, South Africa)	136	Linear	r=0.75	Baker et al., (2015)
Karst aquifer springs (Vorarlberg, Austria)	n_1=96 n_2=50 n_3=45	Rank correlation	$r_{Spearman1}$= 0.58 $r_{Spearman2}$=0.08 $r_{Spearman3}$=0.16	(Frank et al., 2018b)
Water points (wells and borehole, Kwale County, Kenya)	162	Rank correlation	$r_{Kendall}$=0.59	(Nowicki et al., 2019)
River (Colorado, USA)	298	Binary classification for *E. coli* >10 CFU*.100mL^{-1}	83% accuracy (95% CI: 78% - 87%)	(Bedell et al., 2022)
River, SN and WWTP outlets (Ile-de-France, France)	50	Linear	r=0.85	Present study

*CFU: colony forming units, comparable to most probable number (MPN).

4.3.4. Faecal indicator bacteria estimation with hydro-meteorological variables

Runoff and storm water overflows are known to be one of the main sources of contamination in urban rivers. In this section, the relationship between hydro-meteorological variables, such as rainfall and flow rate, were compared with the *E. coli* dataset of La Villette basin at point C in order to predict faecal contamination.

4.3.4.1. Hydro-meteorological variables from La Villette basin

Forecasting FIB concentrations for the following day to three days after the present day is of interest of this study. Therefore, a relationship between *E. coli* and predictable variables, such as rainfall, were explored.

In a first moment, a linear regression model was used to find a relationship between *E. coli*, the flow rate of Chouy station and precipitation from Le Bourget meteorological station, located in around 9 km in the northeast from La Villette basin (Figure 4.54).

E. coli dataset was from point C of La Villette basin (n=10), with the values $\log_{10}$-transformed. Precipitation was also $\log_{10}$-transformed. A comparison between daily values and daily mean values of two to three days were compared to find the most suitable variable.

La Villette canal is one of the downstream branches of Ourcq canal. Its flow rate is highly regulated by navigation locks over many kilometres upstream. The closest flow rate station where the river is still under influence of rainfall variations due to watershed catchment is at Chouy.

Chouy station is located at around 100 km upstream La Villette system (Figure 4.54). Assuming a mean velocity of 0.20 $m.s^{-1}$ on the canal (Vanhalst, 2018), the transfer time from Chouy station to La Villette system is on the order of 6 days. Therefore, the flow rates adopted for the further analysis include the daily mean measures at Chouy station from 4 to 9 days before the sampling at La Villette basin (named Q_4 and Q_9, respectively).

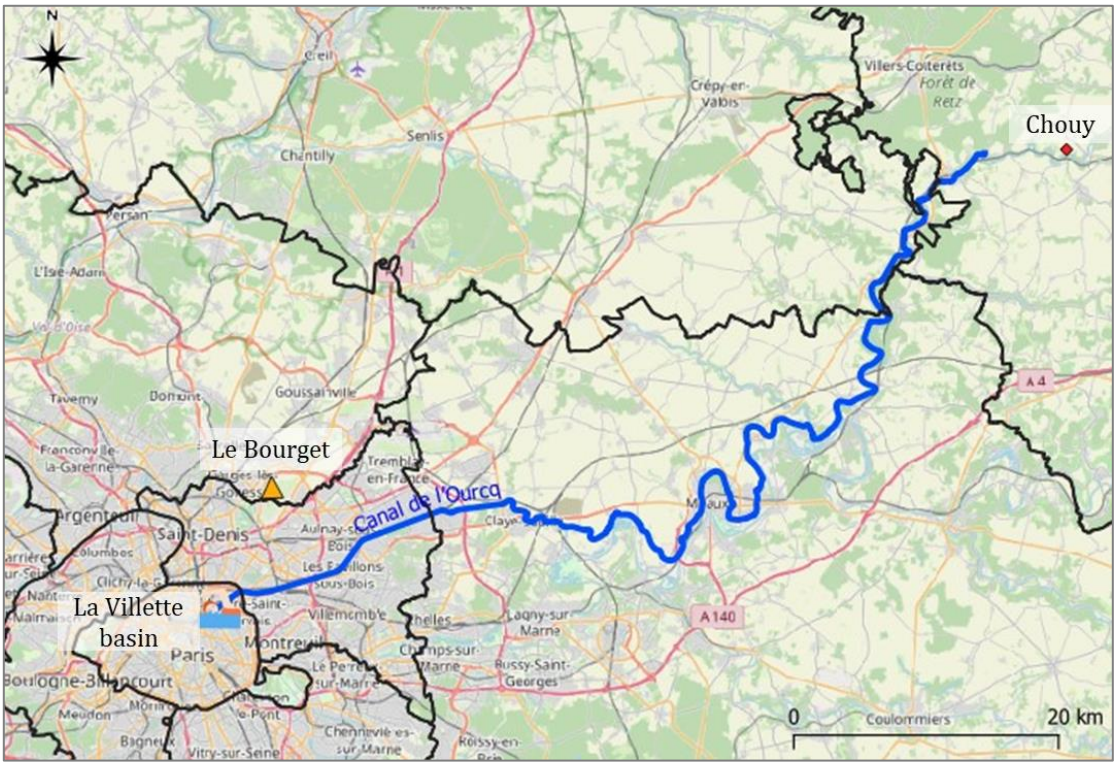


Figure 4.54. Location of La Villette basin, Chouy and Le Bourget station (base map: QGIS)

The dataset from 10/06/2021 to 25/09/2021 was used (Figure 4.55). Comparisons between the individual relationships between *E. col*i and daily precipitation and *E. coli* and mean daily flow rate were made. The most promising variables were then chosen to be part of the linear model relationship with *E. coli*. The obtained relations were visually compared, as well as their R^2.

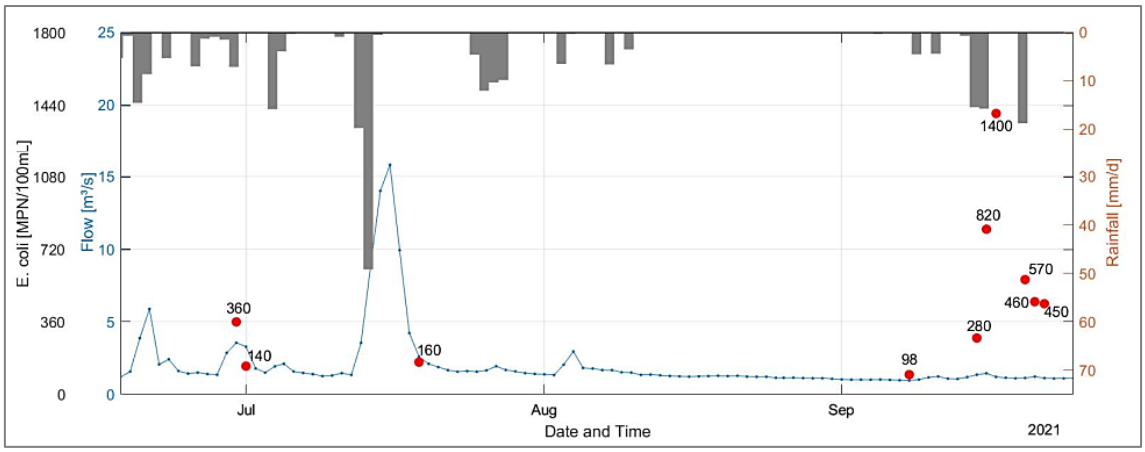


Figure 4.55. Dataset of *E. coli* at La Villette basin (point C, n=10), rainfall at Le Bourget and flow rate at Chouy, from 10/06/2021 to 25/09/2021).

4.3.4.2. Relationship between *E. coli* and hydro-meteorological parameters

Linear relationships between $\log_{10}$-transformed *E. coli* and $\log_{10}$-transformed of different precipitations were compared. Herein, "P_i" and "P_{i-j}" correspond, respectively, to the $\log_{10}$-transformed of the daily precipitation "*i*" days after sampling at La Villette, and the mean daily precipitation from *"i"* to *"j"* days before sampling (in $mm.d^{-1}$).

The precipitations from day 1 to day 4 before the sampling were considered, as well as the combinations of daily means of two and three days of duration (Table 4.14). The best relation was with the mean precipitation of 1 to 3 days before the sampling (P_{1-3}), with $R^2 = 0.74$ and p-value < 0.01 (Figure 4.56).

Table 4.14. Results for linear relationship between $\log_{10}$(*E. coli*) and $\log_{10}$-transformed precipitations.

Precipitation $\log_{10}$-transformed	R^2	p-value
P_1	0.11	0.35
P_2	0.27	0.12
P_3	0.24	0.15
P_4	0	0.86
P_{0-1}	0.13	0.31
P_{0-2}	0.25	0.15
P_{0-3}	0.58	0.01
P_{1-2}	0.42	0.04
P_{2-3}	0.39	0.05
P_{3-4}	0.20	0.20
P_{1-3}	**0.74**	**<0.01**

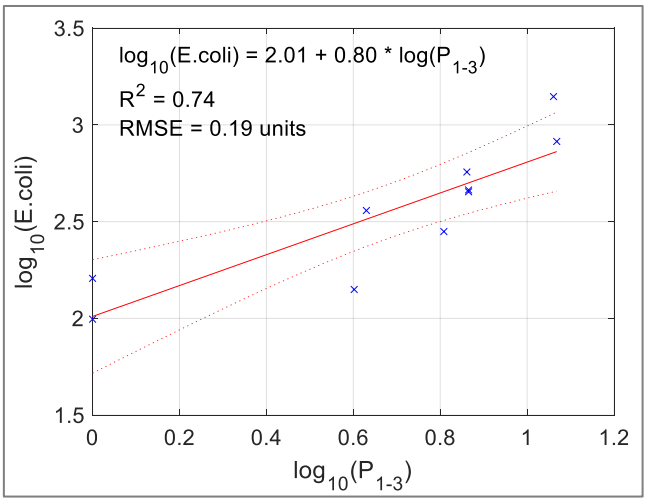


Figure 4.56. Linear relationship between log_{10}(*E. coli*) and $log_{10}(P_{1-3})$.

The same method was used with flow rates from day 5 to day 9 before sampling (Table 4.15). The best relationship was with Q_8, with R^2 = 0.19 and *p-value* = 0.21 (Figure 4.57). It is therefore not significative.

Table 4.15. Results for linear relationship between log_{10}(*E. coli*) and mean daily flow rates, in m^3s^{-1}.

Flow rate	R^2	p-value	Flow rate	R^2	p-value
Q_4	0.12	0.33	Q_{4-5}	0.12	0.33
Q_5	0.12	0.34	Q_{5-6}	0.11	0.35
Q_6	0.09	0.40	Q_{3-6}	0.11	0.34
Q_7	0.02	0.72	Q_{6-7}	0.09	0.41
Q_8	**0.19**	**0.21**	Q_{7-8}	0.10	0.38
Q_9	0.01	0.77	Q_{6-8}	0.15	0.27
Q_{10}	0.14	0.28	Q_{7-9}	0.04	0.60

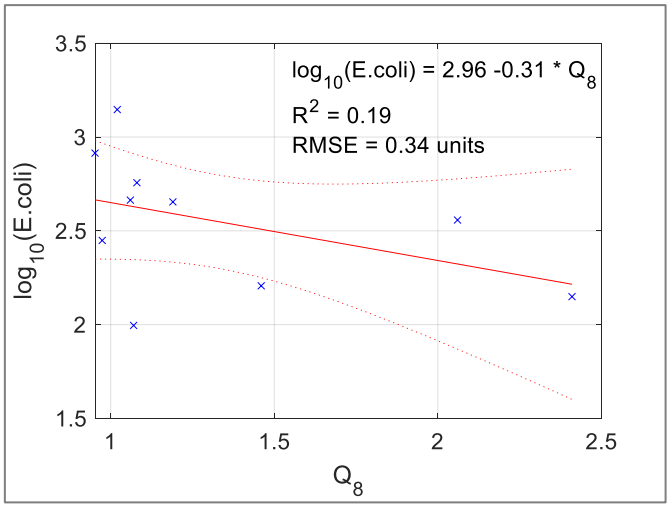


Figure 4.57. Linear relationship between log_{10}(*E. coli*) and flow rate Q_8.

A multivariate analysis was performed to explore the relationships between *E. coli*, flow rate and rainfall. The initial relationship followed the Equation 4.4.

$$log_{10}(E.\ coli) = \beta_0 + \beta_1 \cdot Q_8 + \beta_2 \cdot log_{10}(P_{1-3}) + \beta_3 \cdot DDW_{1-3} \qquad 4.4$$

Where:

- Q_8 is the mean flow rate at Chouy 8 days before *E. coli* measurement (m^3s^{-1})
- P_{1-3} is the log10-transformed daily mean precipitation from 1 to 3 days before *E. coli* measurement ($mm.d^{-1}$)
- DDW_{1-3} = duration of dry weather before precipitation P_{1-3} (days)

The *p-value* of each variable was calculated. A new relationship was established excluding one of the variables from the equation. The R^2 and the global p-value, and the Bayesian Information Criteria (BIC) relative to only one variable were calculated (Table 4.16).

Table 4.16. Statistical results for *E. coli* linear relationship between flow rate, precipitation, and duration of dry weather.

Variables	R^2	*BIC*	*p-value*	
Q_8 P_{1-3} DDW_{1-3}	0.81	-2.92	Q = 0.20 P = 0.05 DDW = 0.48	Global = 0.014
Q_8 P_{1-3}	0.79	-4.33	Q = 0.25 P = 0.003	Global = 0.004
P_{1-3}	*0.75*	*-4.59*	*P = 0.0013*	*Global = 0.001*

The best relationship, *i.e.*, with lower BIC, was the one with only the rainfall as the dependent variable and presented satisfactory results, with R^2 = 0.75 and *p-value* = 0.001 (Equation 4.5).

$$\boldsymbol{log_{10}(E.\ coli) = 2.01 + 0.80 \cdot log_{10}(P_{1-3})} \quad 4.5$$

This relationship represents the increase of *E. coli* only after rainfall events. Therefore:

- it does not consider the immediate *E. coli* increase during a local rainfall and
- it is valid only for wet weather, defined as a rainfall episode with precipitation height higher than 5 $mm.day^{-1}$ within 3 days. When this condition is not achieved, $\log_{10}(P_{1-3}) = 0$.

This expression is mostly useful to predict the exceeding of the *E. coli* threshold for bathing activities, and less precise when predicting *E. coli* values of the bathing area.

In further studies, we recommend including water temperature and solar radiation on the forecasting model to study their relevance on the *E. coli*. In addition, in order to have more representative values rainfall that impact the upstream watershed from La Villette canal, data of more meteorological upstream stations should be considered.

4.3.4.3. Estimation of *E. coli* concentration from hydro-meteorological variables

The estimation of faecal bacteria from hydro-meteorological variables enables to forecast the level of faecal contamination for the within 1-3 days. This information would guide the decision-making on the risks of opening the bathing area after a rainy event. To that mean, the relationship between *E. coli* and precipitation P_{1-3} (Equation 4.5) can be used to estimate *E. coli* and identify when the bathing threshold of 900 $MPN.100mL^{-1}$ would be exceeded.

In this section, *E. coli* values obtained from Ville de Paris and estimated with the Equation 4.5 were compared. Two types of observed values were used. The first was the bacteriological activity measured with ColiMinder sensor at La Villette basin every hour, during summer 2019. The measurements took place at point C, from 02/07/2019 to 03/10/2019, with a time step of 30 minutes on the first day and 1 hour on the rest of the

period. On the first part of this section, the daily average of ColiMinder activity was compared to the estimated *E. coli*.

The second type was the *E. coli* values measured with Fluidion sensor. The sensor was installed at the inlet of La Villette canal (point A) during summer, from 2019 to 2022. The measurements took place three times per day at 19h, 20h and 21h (Paris local time, CEST), in the following periods:

- 23/06 to 10/09/2019
- 15/06 to 30/08/2020
- 28/06 to 21/08/2021
- 01/07 to 21/08/2022

The daily average of the three measurements is used as guidance on the decision-making of opening the bathing area the following day. Therefore, the daily averages were compared with the estimated *E. coli* on the second part of this section.

a) Comparison with ColiMinder observations

In the period of ColiMinder measurements in summer 2019, 10 rainfall events were observed between 02/07 and 03/10/2019 (Table 4.17, Figure 4.58).

In the rainy events from 18/07 to 30/07/2019, the increase of ColiMinder activity occurred earlier than the increase of the estimated *E. coli*. A peak of ColiMinder activity occurred on the 26/07, on the same day of a 3 $mm.day^{-1}$ precipitation. As previously described, the Equation 4.5 for *E. coli* estimation does not consider an increase at the same day of rainfall event, being not valid at this scenario.

After the peak of 26/07, ColiMinder activity decreased on the 27-28/07 and increased again from 29-31/07/2019. The estimated *E. coli* peak, however, occurred on the 28-29/07/2019, detecting only the second increase of ColiMinder activity one day before it occurs.

Nevertheless, in the rainy events between 06/08 to 29/09/2019, an increase of both estimated *E. coli* and ColiMinder activity were observed. The ColiMinder activity peaks of 11/08 and 18/08, and the activity increase of 23-30/09/2019 were detected by the estimated *E. coli*.

The 01/10/2019 was classified as a dry weather day, and the estimated *E. coli* decreased on that day. However, it increased on the 02/10, recovering a similar evolution of the ColiMinder activity in time. The continuous increase of ColiMinder activity on the 01/10 may be due to local impacts of the rainfall of the 01/10.

A relationship between estimated *E. coli* and ColiMinder activity was explored (Figure 4.59). When all dataset is considered (n=89), a weak relationship could be established (R^2=0.29, p-value<0.001). However, all dataset includes weather conditions that do not correspond to the validity of *E. coli* estimation, so only wet weather conditions should be considered. In this case, when considering only values in wet weather conditions (n=26), no significant relationship was obtained (R^2=0.01, p-value=0.56).

Table 4.17. Rainy events during ColiMinder measurements in summer 2019.

Rainy event period	Duration (days)	Total precipitation (mm)	Activity increase after rainfall	
			ColiMinder	Estimated *E. coli*
18-20/07/2019	3	8.5	Not clear	No
27/07/2019	1	20.6	No	Yes
06/08/2019	1	10.9	Yes	Yes
09/08/2019	1	20.7	Yes	Yes

Rainy event period	Duration (days)	Total precipitation (mm)	Activity increase after rainfall	
			ColiMinder	Estimated *E. coli*
12/08/2019	1	18.5	No	No
17-18/08/2019	2	19.7	Yes	Yes
22/09/2019	1	9.5	Yes	Yes
24/09/2019	1	10.4	Yes	Yes
27/09/2019	1	12.7	Yes	Yes
01/10/2019	1	5.3	Yes	Yes

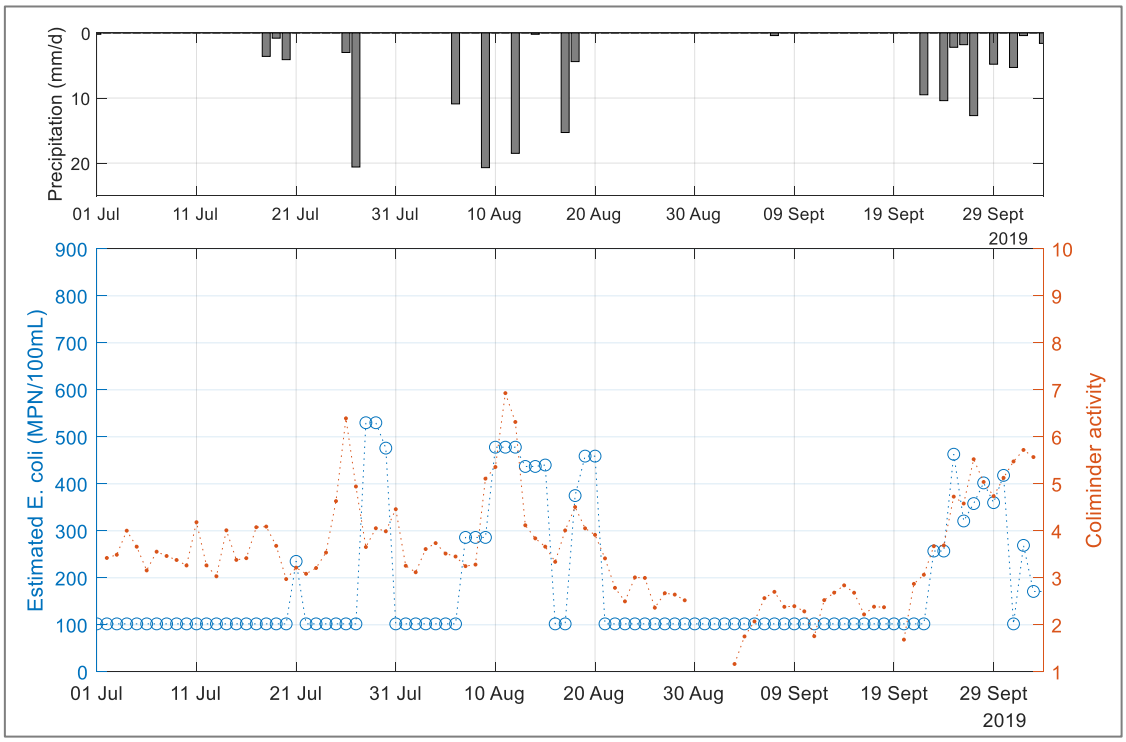


Figure 4.58. Evolution of estimated *E. coli* from precipitation relationship and ColiMinder activity, with daily precipitation, from 01/07/2019 to 04/10/2019.

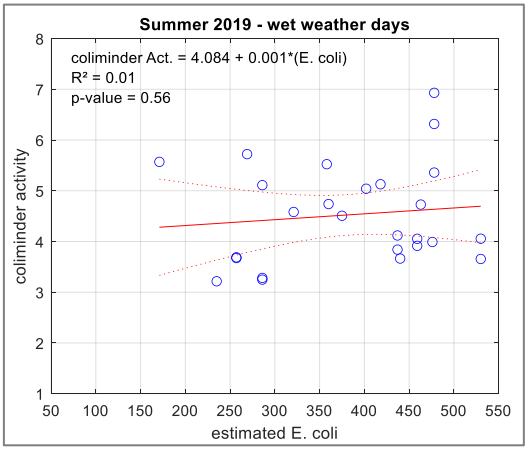


Figure 4.59. Relationship between *E. coli* estimated from precipitation $P_{1\text{-}3}$ and ColiMinder activity considering only wet weather values (n=26).

b) Comparison with Fluidion observations

During summer, Fluidion sensor is installed for the daily assessment of microbiological water quality in La Villette basin. The average of three measurements at the same day is used to define whether the bathing area in La Villette should be opened or not. In the following paragraphs, the observations from 2019 to 2022 are presented and compared with the *E. coli* estimation from the relationship with precipitation $P_{1\text{-}3}$ (Equation 4.5).

In 2019, 6 rainy episodes were observed between 23/06 and 11/09/2019 (Table 4.18, Figure 4.60-a).

In 2020, 3 rainy episodes were observed between 15/06 and 01/09/2020 (Table 4.18, Figure 4.60-b). The intensity of all episodes was low, with daily precipitation between 5.6 and 6.5 mm. In this period, both observed and estimated *E. coli* were lower than the threshold.

In 2021, 5 rainy episodes were observed between 28/06 and 21/08/2021 (Table 4.18, Figure 4.60-c). During this period, in the rainy episode of 12-14/07/2021, both observed and estimated *E. coli* overpassed the threshold. Fluidion observations exceeded the threshold from the 14 to 17/07. The estimated *E. coli* exceeded the threshold from the 14-16/07/2021, but underestimated *E. coli* on the 17/07/2021, where a second increase of *E. coli* was observed four days after the most intense rainfall 13/07/2021. This rainfall, of 42.7 $mm.day^{-1}$, has a return period of 50 years (Brigode et al., 2020), and was considered as an extreme event (Infoclimat, 2021). The second peak of the 17/07/2021 may be related to a late arrival of an upstream water contaminated due to this extreme event.

In 2022, two rainy episodes were observed between 01/07 and 21/08/2022 (Table 4.18, Figure 4.60-d). The daily precipitation of the episodes was 6.8 mm on the 20/07 and 9.9 mm on the 16/08/2022. In this period, both observed and estimated *E. coli* were lower than the threshold of 900 $MPN.100mL^{-1}$.

From 2019 to 2022, a relationship between estimated *E. coli* from precipitation and measured with Fluidion sensor during wet weather was explored. The considered wet weather periods during Fluidion measurements started at the day after the first rainfall of the episode. On the 17/07/2021, the *E. coli* increase measured with fluidion sensors was most probably related to the extreme rainfall event 4 days before and was not considered in the relationship. A significant relationship between observed and estimated *E. coli* was obtained (Figure 4.61, n=51, R^2=0.58, p-value<0.001).

In the daily assessment of La Villette basin, when *E. coli* threshold is exceeded, the bathing area should be closed. The occurrences of closures according to Fluidion observations and to the *E. coli* estimation with precipitation $P_{1\text{-}3}$ (Equation 4.5) were compared (Figure 4.60).

In the studied period, 5 closures were identified: one in 2019, and four in 2021. In 2019, the closure occurred during dry weather. In 2021, the four closures occurred after the rainfall event of 12-14/07/2021. These four closure events could be compared.

The *E. coli* estimation with precipitation was able to correctly predict the first three closures of the bathing area, from 14-16/07/2021. On the 17/07/2021, however, the relationship with $P_{1\text{-}3}$ underestimated *E. coli*. This difference between estimated and observed *E. coli* is probably due to the extreme rainfall event of 13/07/2021.

For further studies, it would be recommended to:

- Collect integrated daily mean samples for *E. coli* measurements from the day before the first day of precipitation to five days after the end of the raining event.
- Monitor *E. coli* not only rainfall events during summer but also include spring and autumn, when precipitation can be higher.
- Include solar radiation and water temperature for a multivariable analysis with *E. coli*.
- Consider the mean precipitation of the studied zone upstream the bathing area (using RADAR, for example).

These measures may improve the detection of closure events related to the consequences of precipitation upstream the bathing area, such as runoff and stormwater network overflow, for a larger rainfall range.

Table 4.18. Rainy events during Fluidion measurements from 2019 to 2022.

Year period	Rainy event period	Duration (days)	Total precipitation (mm)	*E. coli threshold overpassing during the episode*	
				Fluidion	Estimated
Summer 2019	18-20/07/2019	3	8.5	No	No
	27/07/2019	1	20.6	No	No
	06/08/2019	1	10.9	No	No
	09/08/2019	1	20.7	No	No
	12/08/2019	1	18.5	No	No
	17-18/08/2019	2	19.7	No	No
Summer 2020	12/06/2020	1	6.5	No	No
	27/06/2020	1	5.6	No	No
	28/08/2020	1	5.8	No	No
Summer 2021	04/07/2021	1	19.1	No	No
	12-14/07/2021	3	69.8	Yes	Yes
	26-27/07/2021	2	16.7	No	No
	03/08/2021	1	6	No	No
	07/08/2021	1	6.7	No	No
Summer 2022	20/07/2022	1	6.8	No	No
	16/08/2022	1	9.9	No	No

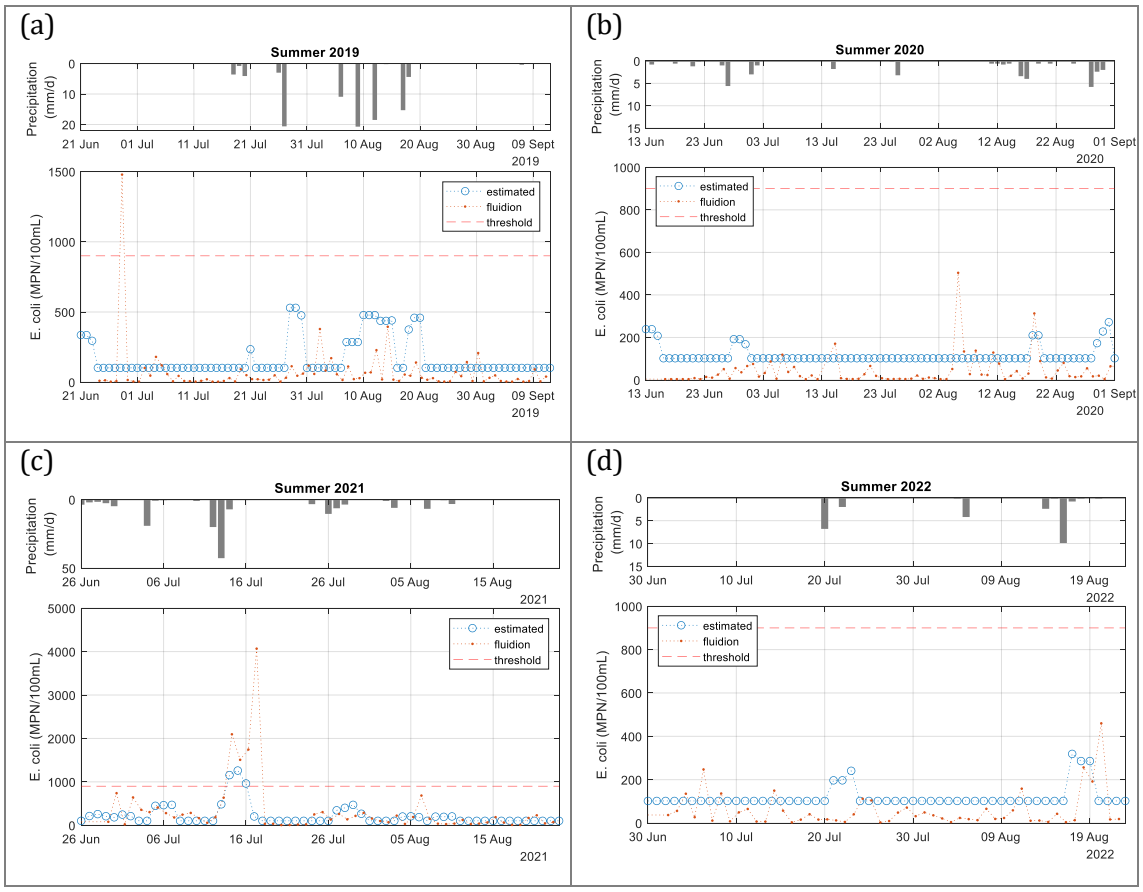


Figure 4.60. Evolution of *E. coli* observed with Fluidion sensor and estimated from relationship with precipitation P_{1-3} during (a) summer 2019, (b) summer 2020, (c) summer 2021, and (d) summer 2022.

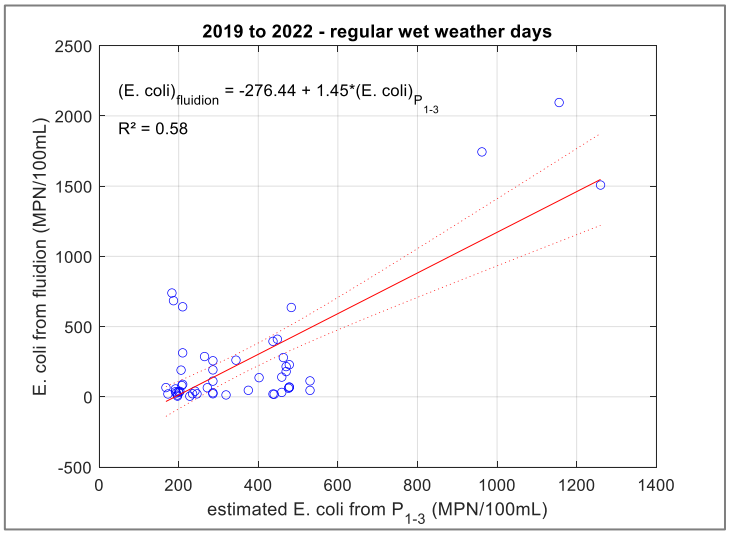


Figure 4.61. Relationship between *E. coli* estimated from precipitation P_{1-3} and observed from Fluidion, considering only values in wet weather conditions and without value from 17/07/2021 (n=51).

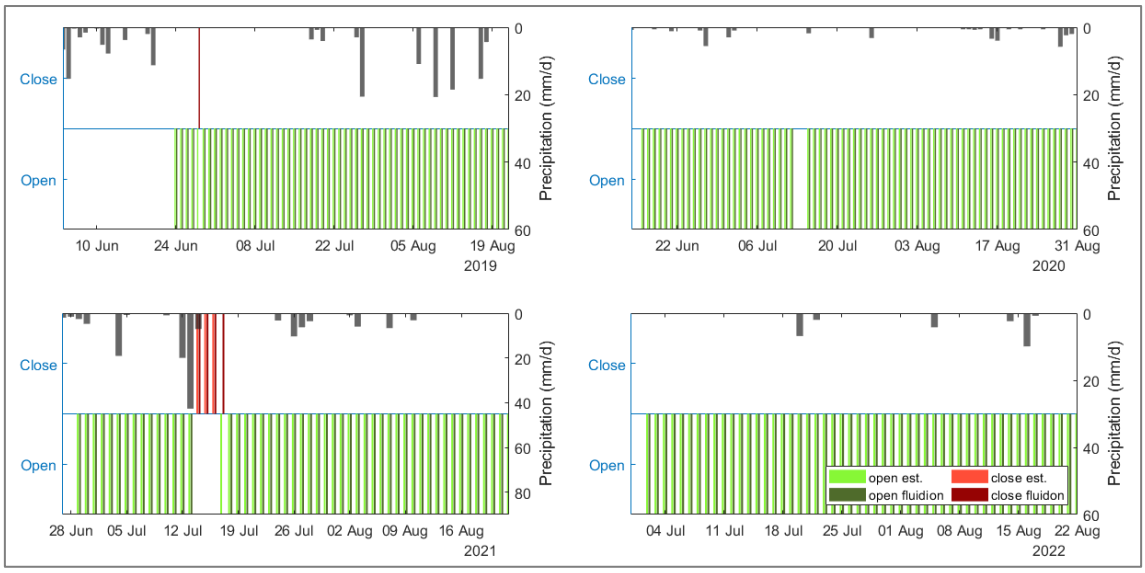


Figure 4.62. Daily rainfall evolution (in mm.day^{-1}) and indication of closure (red) and opening (green) of the bathing area according to *E. coli* estimation (light) and *E. coli* Fluidion observation (dark).

4.3.4.4. Discussion

A relationship between faecal contamination and hydro-meteorological variables could be useful on FIB forecasting (Seis et al., 2018). Flow rate and precipitation is a known factor that may affect microbiological water quality (EEA, 2018; Herrig et al., 2019). However, some obstacles may limit the obtention of a significant correlation with FIB. These limitations may include low sampling frequency, low precipitation intensities and low occurrences of high precipitation (Ordulj et al., 2022).

An increased flow rate can be transporting the contamination related to an important precipitation upstream (EEA, 2018) or can resuspend sediments (Wilkes et al., 2011), which may increase faecal bacteria levels.

Seis et al. (2018) tested nine model equations for implementing early warning systems based on statistical modelling. The variables of the best relationship included the average flow rate of one to two days prior to sampling for *E. coli* measurements ($Q_{1\text{-}2}$); the log-transformed average daily precipitation from one to four days prior to sampling ($P_{1\text{-}4}$); and WWTP daily sum discharges ($WWTP_{1\text{-}4}$) one to four days prior to sampling.

The results in La Villette basin corroborate with Seis et al. (2018). The differences between both relationships can be explained by the particularities of La Villette basin, where flow rate variation is not significant and the closest WWTP discharge is 20 km upstream, also being affected by the navigation locks. In addition, Seis et al. (2018) used spatial averages of 15 rain gauges, considering a wider watershed. This would explain the longer period of the precipitation average (four days, $P_{1\text{-}4}$) than in La Villette basin (three days, $P_{1\text{-}3}$).

In the bathing water plan of La Villette basin (SAFEGE, 2017b), the relationship between FIB and daily precipitation from two meteorological stations were explored. The first station was Roissy Charles-de-Gaulle (49.015278 N, 2.534444 E), 18 km northeast upstream of the bathing area. The second, Changis (48.964722 N, 3.011944 E), 47 km east-northeast from the bathing area, upstream the watershed.

With precipitation at Roissy, no relationship was found. The relationship between the precipitation at Changis and *E. coli* was strong (n=53, $r_{Pearson}$=0.86). However, the rainfall with the strongest correlation was at the same day as the sampling in La Villette basin.

This result is contradictory because the meteorological station is far upstream the bathing area. The canal is 75 km long from Changis to La Villette basin. The mean water velocity of the canal is 0.20 $m.s^{-1}$ (SAFEGE, 2017b), resulting on a transport time of at least 4 days. Therefore, the best correlation was expected to be with rainfall from around 4 days before sampling.

Finally, to improve the reliability of the relationship between *E. coli* and P_{1-3} in La Villette basin, a more robust dataset should be added. Nevertheless, its application, even if limited to wet weather conditions, can contribute on reinforcing the microbiological water quality assessment of the bathing area.

In La Villette basin, the estimation of FIB was explored in two approaches: nowcasting through protein-like fluorescence components and forecasting through hydrometeorological variables.

The FIB estimation was possible through the PARAFAC component C6. It was the protein-like component with the strongest relationship with both *E. coli* and IE. C6 could be used as effective indicator for monitoring microbiological water quality in both dry and wet weather.

The relationship between flow rate and *E. coli* was not significant. The reason may be that, in La Villette basin, the variation of the flow rate is very limited. The flow rate is strongly regulated by a series of navigation locks upstream.

The forecasting of FIB was possible through the mean rainfall, measured in a meteorological station located 9km northeast of La Villette basin, from one to three days before sampling (P_{1-3}). It was the hydrometeorological variable with the strongest relationship with *E. coli*.

The *E. coli* estimation from P_{1-3} presents some limitations. It does not consider immediate *E. coli* increase related to local runoff. It cannot predict a further increase after storm overflow due to heavy rainfall in a more distant catchment. Nevertheless, it indicated that the impact of a heavy precipitation may last at least for three days.

Finally, *E. coli* relationship with C6 may help effective strategies for microbiological water quality monitoring. *E. coli* relationship with P_{1-3} could be used for microbiological forecast.

4.4. Hydro-microbiological modelling

In this chapter, the results of the implementation of a 3D hydrodynamic model in La Villette study site are presented.

Part of the results were published in the proceedings of Telemac User Conference (TUC) of 2021 and 2022 (Angelotti et al., 2022, 2021), and of Novatech International Conference of 2023 (Angelotti et al., 2023).

This chapter presents the results on hydro-microbiological modelling in our study site, the La Villette basin. A three-dimensional hydrodynamic model, TELEMAC-3D, was used to compute the time and space distribution of water temperature, velocity, and contamination tracer. A thermal module (WAQTEL) was activated to reproduce the thermal stratification on the basin.

The timeseries of water temperature and conductivity measured with in situ sensors installed upstream were used for model input, and measured at the bathing area were used for model validation of thermal stratification and transfer time. Episodes of hot weather with thermal stratification (10-22 June 2021 and 13-23 July 2022) and rainy events with large conductivity variation (01-10 June 2021 and 03-23 September 2022) were simulated.

Then, the model was used to predict spatiotemporal distribution of *E. coli* from the upstream measurement point to the bathing area. *E. coli* values were inferred from the two relationships presented in chapter 4.3 established for La Villette basin. To have a few-hour forecast of the sanitary risk at the bathing area, real-time fluorescence data from the upstream point, the component C6, was used as an input (12-24 September 2021 and 20-30 June 2022). To have a few-days forecast, daily precipitation from an upstream meteorological station was used as an input (10-22 June 2021 and 10-20 July 2021).

4.4.1. Model implementation

4.4.1.1. Input variables

The simulations of TELEMAC-3D require two steering files. The first one is the main steering file ("t3d") and corresponds to the 3D hydrodynamic computation. The second is the WAQTEL steering file ("waq") and corresponds to the computation of the thermal processes in the study site.

Two input files are also required: the geometry file and the boundary conditions file (detailed in section 3.5.2). The geometry file contains the bathymetry and the mesh grid of La Villette system in Lambert projection coordinates. The boundary conditions file contains the definition of all boundaries of the simulation domain. The two liquid boundaries are also defined in this file:the inlet as Boundary 1 and the outlet as Boundary 2. These files remain the same among simulations and do not require any modification.

In the main steering file, the declaration of three other input files and four additional variables are required in every simulation (Table 4.19). The input files are the PREVIOUS COMPUTATION FILE, the LIQUID BOUNDARIES FILE, and the ATMOSPHERIC DATA FILE.

- The PREVIOUS COMPUTATION FILE corresponds to a simulation of three days, ending immediately before the beginning of the period of interest. This previous simulation aims to set the velocity field of the studied domain without considering the thermal exchanges with the atmosphere.
- The LIQUID BOUNDARIES FILE contains the inlet values of water surface level, water temperature and tracers (p.e. conductivity or FIB), and outlet values of flow rate. This file includes data from the previous simulation and simulated periods in an hourly timestep, with hourly mean values.
- The ATMOSPHERIC DATA FILE contains the following meteorological conditions of the simulated period: wind magnitude and direction, air temperature, atmospheric pressure, nebulosity, rainfall, relative humidity, and visibility. It is necessary on the computation of thermal exchanges with the atmosphere. The data is obtained from the nearest upstream meteorological station with all required data: Roissy airport station, at an hourly timestep.

The input variables are the ORIGINAL DATE OF TIME, TIME STEP, NUMBER OF TIME STEPS and INITIAL VALUES OF TRACERS.

- The ORIGINAL DATE OF TIME indicates the year, month, and day of the beginning of the pre-simulated period, three days before the period of interest.
- The TIME STEP indicates the duration of the computation time step, in seconds. In all simulations, the time step was 20 s.
- The NUMBER OF TIME STEPS indicates the quantity of time steps will be computed. It determinates the simulated period. It is calculated as the ratio between the total duration of the simulated period and the time step, both in seconds.
- The INITIAL VALUES OF TRACERS indicates the initial values of the tracers (water temperature, conductivity, and *E. coli*). It corresponds to the values of all the computation domain at the beginning of the simulation.

In the WAQTEL steering file, the declaration of one additional variable is required in every simulation: the Secchi depth (Table 4.19).

Table 4.19. Input files and variables of TELEMAC-3D steering files

Steering file	Input	Filename	Description
Main (t3d)	Geometry file	Villette_lambert.slf	File with bathymetry and mesh grid of La Villette system in Lambert projection coordinates. Same file for all simulated periods.
	Boundary conditions file	Villette_BC_ HinQout.cli	File with the types of each boundary (liquid or solid). Same file for all simulated periods.
	Previous computation file	r3d_presim_ YYYYMMDD_3d.slf	3D result file from previous simulation of 3 days before the beginning of the present simulation. In this pre-simulation, the thermic process is not considered.
	Liquid boundaries file	qsl_ YYYYMMDD_Xd.qsl	File with the surface level and tracers values at inlet and flow rate values at outlet boundary in the simulated period, at an hourly timestep.

Steering file	Input	Filename	Description
	ASCII atmospheric data file	meteo_7157_ YYYYMMDD_Xd.txt	Meteorological data from Roissy meteorological station in the simulated period, at an hourly timestep.
	Original date of time	-	Date corresponding to the beginning of the pre-simulation period. Input format: YYYY;MM;DD.
	Time step	-	Time step of the computation, in seconds. Same value for all simulated periods.
	Number of time steps	-	Number of time steps to be computed. Indicates the total duration of the simulation.
	Initial values of tracers	-	Initial values of the tracers: water temperature, conductivity, and, when available, *E. coli*.
	Water quality module	waq_11.cas	WAQTEL steering file with water quality parameters for thermic process computation. Same file for all simulated periods.
WAQTEL	Geometry file	Same as from main steering file.	
	Boundary conditions file	Same as from main steering file.	
	Secchi depth	-	Mean value of Secchi transparency field measurements during the simulated period. If not available, mean value of equivalent period in a different year.

4.4.1.2. Calculation of flow rate

The hourly average flow rate, in $m^3.s^{-1}$, is one of the inputs of the hydrodynamic model. In La Villette basin, no flow rate is measured *in situ*. Two methods of flow rate calculation are presented: from a balance of the daily outlet volumes and from Manning-Strickler equation. The inflow rate is assumed to be equal to the total outlet flow rate. A validation between the two methods is presented.

4.4.1.2.1. Flow rate estimation from daily outlet volumes

The daily outlet volumes (in $10^3.m^3.day^{-1}$) were obtained from *Service des Canaux* (*Ville de Paris*), from January 2020 to December 2022. The two outlets of La Villette basin are the navigation locks to St. Martin canal (V_{nav}), and the pumping station of non-potable water from *Eau de Paris* (V_{ps}).

From January 2020 to December 2022, the daily outlet volumes from the pumping station of the following months were not available: August, October, November and December 2020, March 2021, and June and July 2022 (Figure 4.63). When both V_{nav} and V_{ps} were available, volumes from the pumping station (V_{ps}) were greater than the volumes from the navigation locks (V_{nav}), with V_{ps} representing, in average, 94% of the total volume.

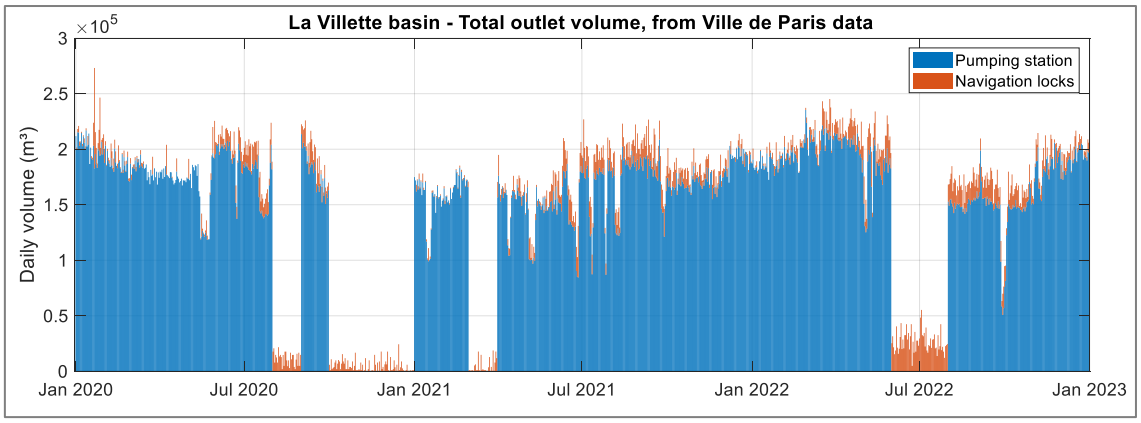


Figure 4.63. Daily outlet volume of La Villette basin. Volumes from pumping station (blue) and from navigation locks (orange), from January 2020 to December 2022.

The total outflow rate in La Villette basin (Q_V) is the sum of the daily mean flowrate of the navigation locks (Q_{nav}) and of the pumping station (Q_{ps}) (Equation 4.6).

$$Q_V = Q_{nav} + Q_{ps} \quad 4.6$$

With $Q_{nav} = V_{nav}/86400$ and $Q_{ps} = V_{ps}/86400$.

From January 2020 to December 2022, the mean flow rate of La Villette basin was of 2.12 ± 0.30 $m^3.s^{-1}$, in the range [0.73-3.16] $m^3.s^{-1}$ (Figure 4.64).

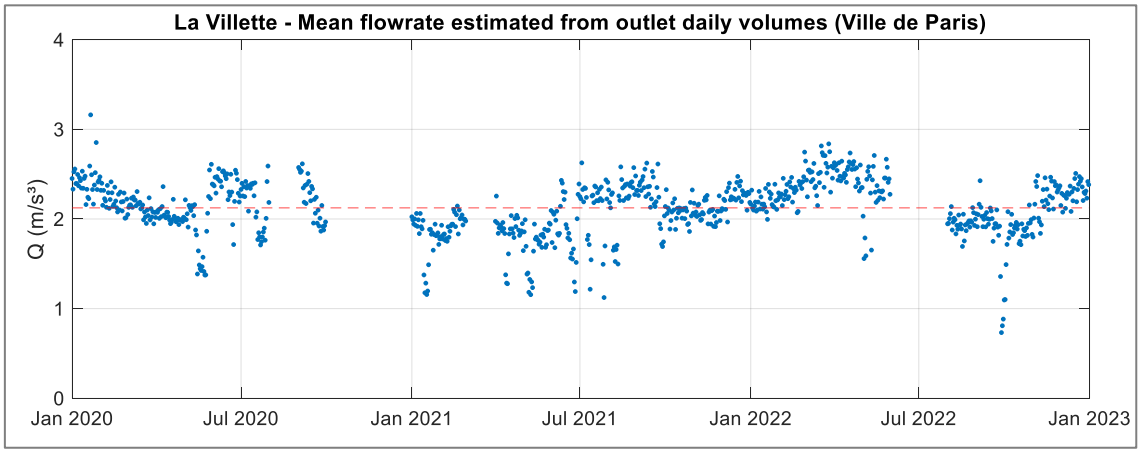


Figure 4.64. Daily mean flow rate of La Villette basin computed from daily outflow volumes, from January 2020 to December 2022.

4.4.1.2.2. Manning-Strickler equation

Since May 2021, the water level was measured upstream at point A. The hourly flow rate was calculated from the upstream water level (h). The Manning-Strickler equation was applied to estimate the average flow rate on a free surface canal at the same time step than field measurements (Equation 4.7).

$$Q(h) = K_s.S_{tot}.R_H^{\frac{2}{3}}.\sqrt{i} \quad 4.7$$

Where K_S is the coefficient of Stricker, S_{tot} is the cross-sectional area of the flow, R_H is the hydraulic radius and i the hydraulic slope. K_S, S_{tot} and R_H can be expressed as a function of the water level.

The Strickler coefficient, K_S, is the inverse of Manning coefficient and represents the channel roughness. For build-up channels with vegetal lining, K_S can be between 2 and 33 $m^{1/3}.s^{-1}$. With unfinished concrete, K_S can be between 50 and 71 $m^{1/3}.s^{-1}$ (Chow, 1959).

The bed vegetation in La Villette basin can be higher than 2 m, as observed during field campaigns. Also, the walls of the canal are very weathered. Therefore, K_{Sb} = 2 $m^{1/3}s^{-1}$ was adopted for the bottom of the canal and K_{Sw} = 50 $m^{1/3}.s^{-1}$ for the walls of the canal.

To find the equivalent Strickler coefficient K_S, the composite roughness coefficient for two different Strickler coefficients was calculated (Equation 4.8).

$$\frac{P}{{K_S}^{3/2}} = \frac{P_b}{{K_{S_b}}^{3/2}} + \frac{P_w}{{K_{S_w}}^{3/2}} \quad 4.8$$

Where P is the total wetted perimeter, P_b is the wetted perimeter of the bottom and P_w the wetted perimeter of the walls. The three wetted perimeters can be expressed in function of the water level h:

- $P = 2h + L$
- $P_b = L$
- $P_w = 2h$

When replacing P, P_b and P_w in the Equation 4.8, the expression of $K_S(h)$ is obtained (Equation 4.9).

$$K_S(h) = \left(\Big/ \frac{L}{K_f^{\frac{3}{2}}} + \frac{2h}{K_b^{\frac{3}{2}}} \right)^{2/3} \quad 4.9$$

The total cross-sectional area of the flow, S_{tot}, is the sum of the rectangular cross-sectional area with the canal dimensions and the cross-sectional area at the bottom of the canal, S_{bottom}. The dimension of the rectangular area is L x h, where L is the width of the canal (25 m), and h is the water level (Equation 4.10). S_{bottom} has an irregular geometry and was estimated from the bathymetry of La Villette canal (Figure 4.65).

$$S_{tot}(h) = S_{bottom} + L.h \quad 4.10$$

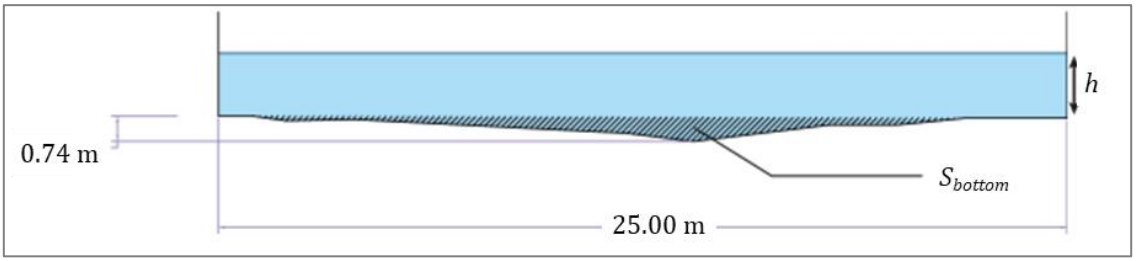


Figure 4.65. Scheme of the cross-sectional area of the flow in La Villette canal

The hydraulic radius, R_H, is the ratio between the total cross-sectional area of the flow and the total wetted perimeter (Equation 4.11).

$$R_H(h) = \frac{S_{bottom} + L.h}{2h + P_{bottom}} \quad 4.11$$

Where S_{bottom} = 6.16 m^2 and P_{bottom} = 25.07 m.

The initial value of the hydraulic slope, i_0, was calculated as the ratio between the difference of surface levels of the canal bed upstream, $Z_{upstream}$, at the inlet of the canal, and downstream, $Z_{downstream}$, at the outlet of the canal, divided by the canal length $L_{\Delta Z}$ (Equation 4.12).

$$i_0 = \frac{Z_{upstream} - Z_{downstream}}{L_{\Delta Z}} \quad 4.12$$

Where $Z_{upstream}$ = 49.00 m, $Z_{downstream}$ = 48.92 m, and $L_{\Delta Z}$ = 800 m, resulting in $i_0 = 10^{-4}$ m.m^{-1}. The surface level of La Villette canal bed were obtained from the bathing area plan of La Villette basin (SAFEGE, 2017a).

The flow rate was computed for each water level measurement. The daily mean was computed and compared to the flow rate estimation from daily outlet volumes, presented in the following section.

The hydraulic slope, *i*, is the only constant variable of Manning-Strickler equation. A calibration of *i* was performed with the available data from May 2021 to November 2022.,

The final value of the hydraulic slope was $i = 1.36 \cdot 10^{-4}$ m.m^{-1}. The comparison between both flow rates is presented in the following section (4.4.1.2.3).

4.4.1.2.3. Validation of flow rate calculation

The daily mean flow rate obtained with the Manning-Strickler equation (Equation 4.7) was validated by comparison to the daily mean flow rate computed from the outlet volumes of La Villette basin (Equation 4.6).

The calibration of the hydraulic slope, *i*, of Manning-Strickler equation was conducted. With $i = 1.36 \cdot 10^{-4}$ m.m^{-1}, a good agreement with both flow rates calculations was observed (Figure 4.66-a, RMSE = 0.33 m^3.s^{-1}, R^2 = 0.58). The average difference between daily flow rates from May 2021 to November 2022 was -0.06 ± 0.32 m^3.s^{-1} (Figure 4.66-b).

Hourly mean values from Manning-Strickler equation were used as input of the model (Figure 4.67). From May 2021 to March 2023, the mean flow rate of La Villette basin was of 2.08 ± 0.09 m^3.s^{-1}, with range [1.78-2.29] m^3.s^{-1}.

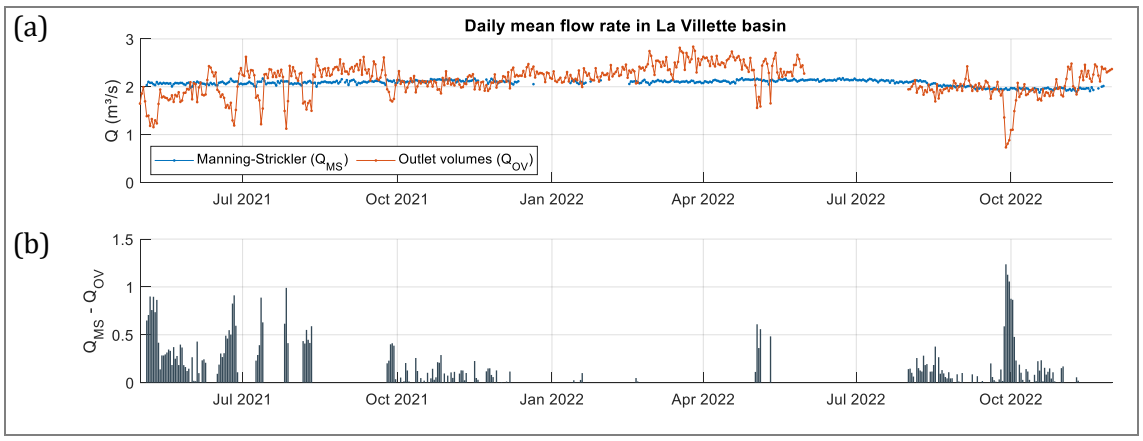


Figure 4.66. (a) Daily mean flow rates computed by Manning-Strickler equation (Q_{MS}, in blue) and from the daily outlet volumes of La Villette basin (Q_{OV}, orange). (b) Difference between Q_{MS} and Q_{OV}. Data from January 2021 to December 2022.

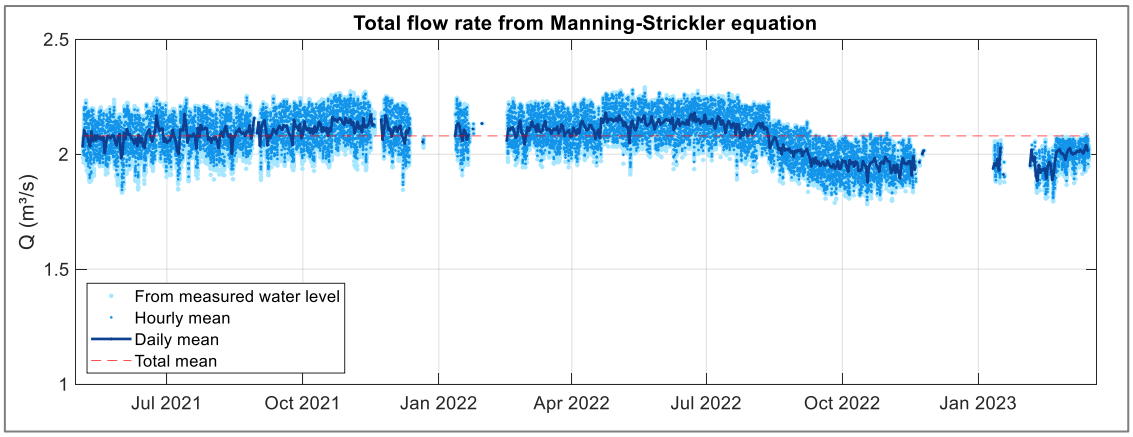


Figure 4.67. Daily mean flow rate of La Villette basin computed from water level measurements, from May 2021 to March 2023.

4.4.2. Model configuration

The hydrodynamics and thermal exchange with the atmosphere of La Villette system were simulated using the hydrodynamic model TELEMAC-3D and the thermal module WAQTEL, with main configuration as detailed in section 3.5.3.1 and summarized in Table 4.20.

The two-dimensional mesh with the bathymetry of the domain of the study site was presented in Figure 3.15. The 3D mesh is composed by 7 horizontal parallel layers of 0.50 m depth. Each layer is composed of 3932 nodes and 7289 horizontal triangular elements of 5 m side.

The vertical turbulence is computed through Nezu and Nakagawa mixing length model. The coefficient for vertical diffusion of tracers is 10^{-4} $m^2.s^{-1}$. The horizontal turbulence uses a constant viscosity coefficient. The coefficient for horizontal and vertical diffusion of velocities is 10^{-6} $m^2.s^{-1}$.

In of La Villette system, the bottom friction follows the Strickler law, with K_S = 2 $m^{1/3}.s^{-1}$, indicated for build-up channels with highly present vegetal lining (Chow, 1959). The friction on the sidewalls follows the Nikuradse law, with k_S = 0.02 m, representing the size of asperities.

The water-atmosphere heat exchange model considers the complete balance of exchanged heat fluxes that affects the energy balance. The required inputs are the timeseries of wind magnitude (in $m.s^{-1}$) and direction (in degrees), air temperature (in °C), atmospheric pressure (in hPa), relative humidity (in %), nebulosity (in octas) and precipitation (in $mm.s^{-1}$), as well as constant values for Secchi depth (in m) and the lightness of the sky.

The meteorological data was obtained from the MétéoFrance meteorological station at Roissy airport (49.015278 N, 2.534444 E), located roughly 18 km from La Villette canal (Figure 3.18). The Secchi depth was obtained from field measurements (presented in section 4.1.5). The lightness of the sky was considered as mean pure sky.

The effect of wind, atmospheric pressure and influence of rain or evaporation are taken into account.

The time-step of the input data from the meteorological station and of the liquid boundaries (surface level, flow rate, water temperature, conductivity and *E. coli*) is of 1 h. The time-step of the model computation is of 20 s. The results are given in time-step of 10 min.

MATLAB® R2022b (MathWorks®, USA) was used for the preparation of input data and processing of simulation results.

Table 4.20. Parameters of TELEMAC-3D model configuration declared in the hydrodynamic steering file (t3d) and the thermal process steering file (WAQTEL)

Parameter	Description	Values
3D mesh	Triangular elements. Elements per layer: Horizontal layers: Depth between layers:	 7289 7 0.50 m
Vertical turbulence model	Nezu and Nakagawa mixing length model.	$D_v = 10^{-6}$ m².s^{-1}
Horizontal turbulence model	Constant viscosity coefficient (default).	$D_v = 10^{-6}$ m².s^{-1}
Bottom roughness	Friction law of Strickler	$K_S = 2$ m$^{1/2}$.s^{-1}
Lateral solid boundaries roughness	Friction law of Nikuradse	$k_S = 0.02$ m
Water-atmosphere heat exchange model inputs	Timeseries of meteorological data: Secchi depth (SD): Lightness of the sky:	from Roissy station. SD from field data. mean pure sky.
Input data time-step	Time-step of the input data: meteorological data and liquid boundaries conditions data.	1 h
Computation time-step	Computation time-step of the simulation.	20 s
Results time-step	Time-step of the simulation results output.	10 min

4.4.3. Model validation

A total of seven simulated periods are presented to illustrate: the validation of the model regarding the heat exchange with the atmosphere and the mean flow velocity, and the spatiotemporal distribution of a faecal contamination from upstream (Table 4.21).

For the validation of the model ability to simulate water temperature and thermal stratification, two hot weather periods were simulated. The first was in June 2021 (10-22/06/2021) and the second on July 2022 (13-23/07/2022).

For the validation of the mean flow velocity, two periods of large variations of electrical conductivity were simulated. The first was in June 2021 (1-10/06/2021) and the second in September 2022 (03-23/09/2022). The transfer time from upstream (point A) to downstream (point B) was estimated through the lag time of conductivity variation between both points.

The position of the water column at point B corresponds to the node 3311 at layer 1, the bottom layer. The simulation results of the correspondent nodes at layers 3, 5 and 6 were compared, respectively, with field data at 2.00 m (bottom), 1.20 m (middle) and 0.50 m (surface).

The model performance was assessed through two indicators, root mean square error (RMSE) and R-squared (R^2) between simulation results and field data at point B.

Table 4.21. Synthesis of the simulated periods.

Short name of period	Simulated period	Output variables of interest	Secchi depth
Early June 2021	01 to 10/062021	Conductivity	1.5 m
Mid-June 2021	10 to 22/06/2021	Water temperature	1.5 m
July 2022	13 to 23/07/2022	Water temperature	0.9 m
September 2022	03 to 23/09/2022	Conductivity	0.9 m

4.4.3.1. Thermal stratification in water column

The validation of the water temperature and thermal stratification is illustrated by two simulated periods: June 2021 (10-22/06/2021) and July 2022 (13-23/07/2022). The difference between the water temperature at surface and bottom layers (ΔTw) is considered an indicator of the stratification intensity.

- *June 2021 (10-22/06/2021)*

From 10/06/2021 to 22/06/2021 (12 days), the air temperature was higher than 15 °C during the evening and raised up to 30 °C during the day between the 14th and the 17th of June (Figure 4.68).

The range of the simulated and measured water temperatures was from 21.1 to 25.5 °C (Figure 4.69). The simulated temperatures are close to the field measurements, with a RMSE of 0.38 °C at the surface, 0.39 °C at the middle and 0.36 °C at the bottom. A good agreement is obtained between model results and field data at the three depths ($R^2 \geq 0.87$, Figure 4.70).

The water column is stratified during the day and mixes at night (Figure 4.71). The thermal stratification begins at the same time for both field data and model results. However, the mixing of the water column in the night occurs earlier in the model, around 2h to 7h before the observations. In addition, the model underestimated the difference between surface and bottom water temperature (ΔTw): field data thermal amplitudes were between 0.8 °C and 1.9 °C, while simulated amplitude was between 0.1 °C and 0.7 °C. The lower ΔTw of the model could be the reason of the lower duration of the stratification: once the temperature at the surface starts to decrease, it takes less time to mix the water column.

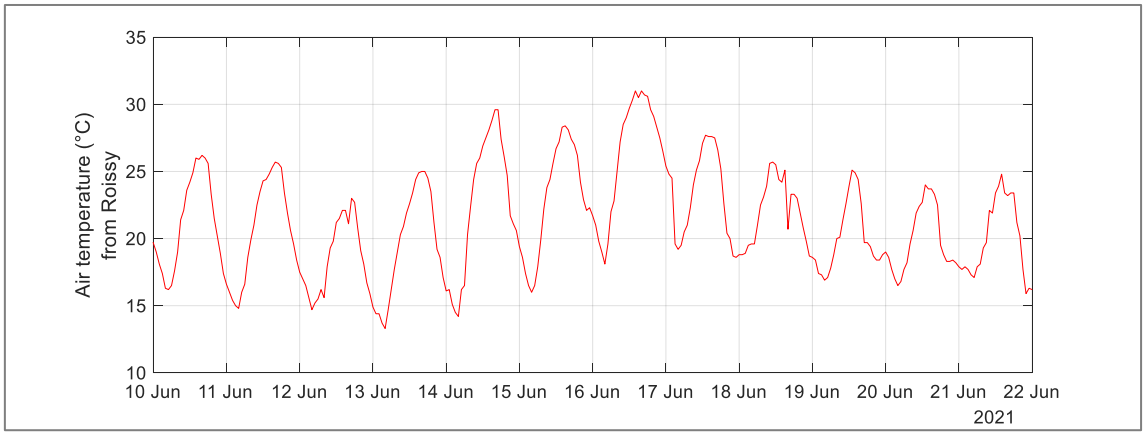


Figure 4.68. Air temperature measured at Roissy meteorological station, used as input on the simulation from 10 to 22/06/2021.

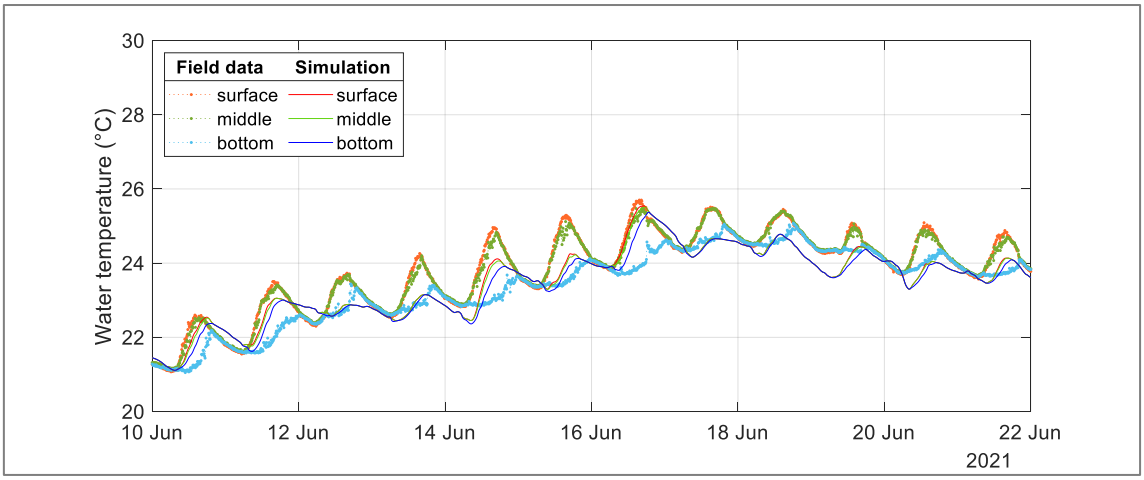


Figure 4.69. June 2021: Measured (dotted) and simulated (line) water temperature in surface (red), middle (green) and bottom (blue) layers. Data from 10 to 22/06/2021.

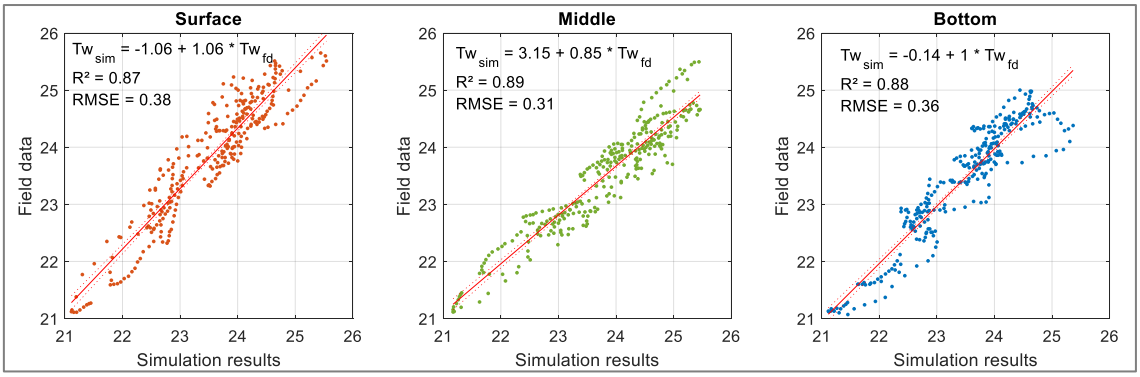


Figure 4.70. June 2021: Relationship between simulated and measured water temperature at surface (orange, left), middle (green, middle) and bottom (blue, right) layers. Data from 10 to 22/06/2021.

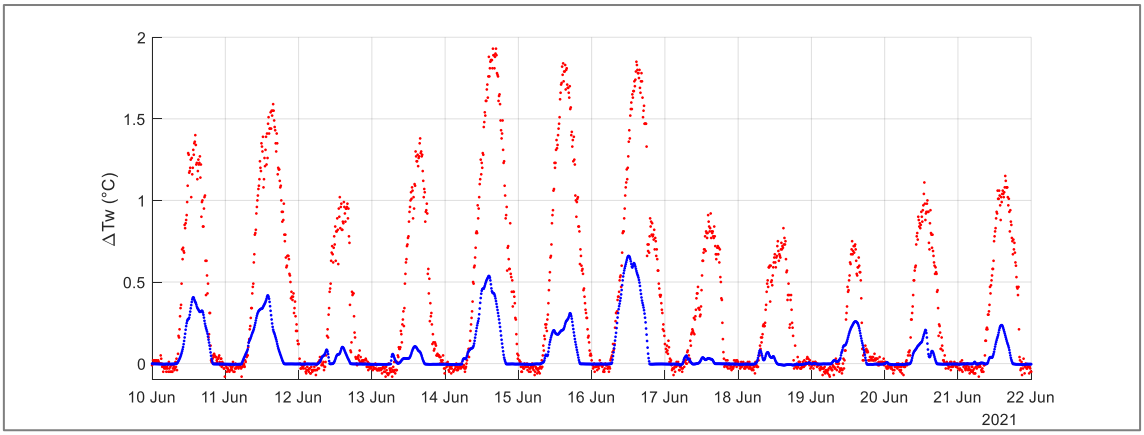


Figure 4.71. June 2021: Difference between surface and bottom temperatures (ΔTw) of field data (red) and simulation results (blue), at point B. Data from 10 to 22/06/2021.

- *July 2022 (13-23/07/2022)*

From 13/07/2022 to 23/07/2022 (10 days), the air temperature was higher than 15°C during the evening and raised up to 40°C during the day between the 17th and the 20th of July (Figure 4.72).

The range of the simulated and measured water temperatures was from 23.1 to 26.2 °C (Figure 4.73). The simulated temperatures are close to the field measurements, with a RMSE of 0.27 °C at the surface, 0.32 °C at the middle and 0.21 °C at the bottom. A good agreement is obtained between model results and field data at the three depths (R^2>0.70, Figure 4.74).

The water column is stratified during the day and mixes at night (Figure 4.75). The thermal stratification of the water column begins and ends at the same time for both field data and model results. However, the model underestimated the difference between surface and bottom water temperature in 6 out of 10 days of simulation. The field data thermal amplitudes were between 0.3 °C and 1.4 °C, while simulated amplitude were between 0.2 °C and 1.0 °C.

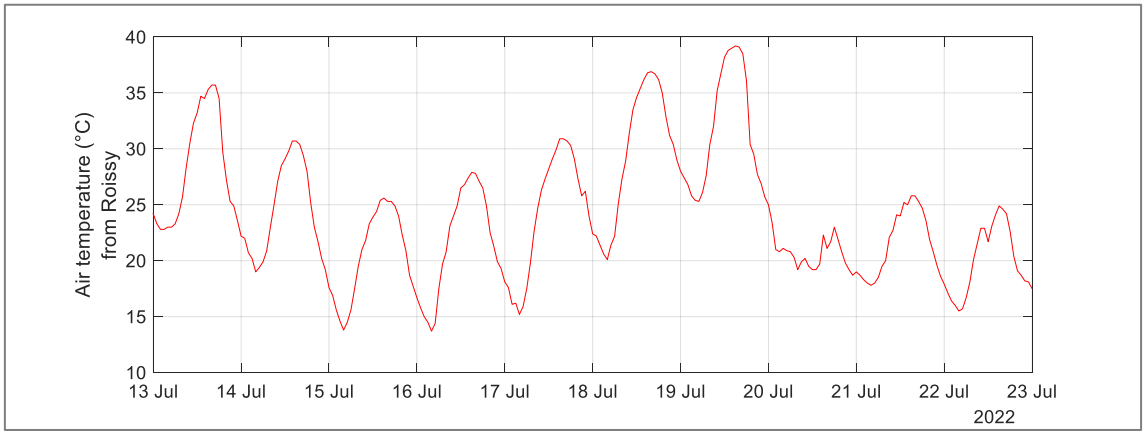


Figure 4.72. Air temperature measured at Roissy meteorological station, used as input on the simulation from 12 to 23/07/2022.

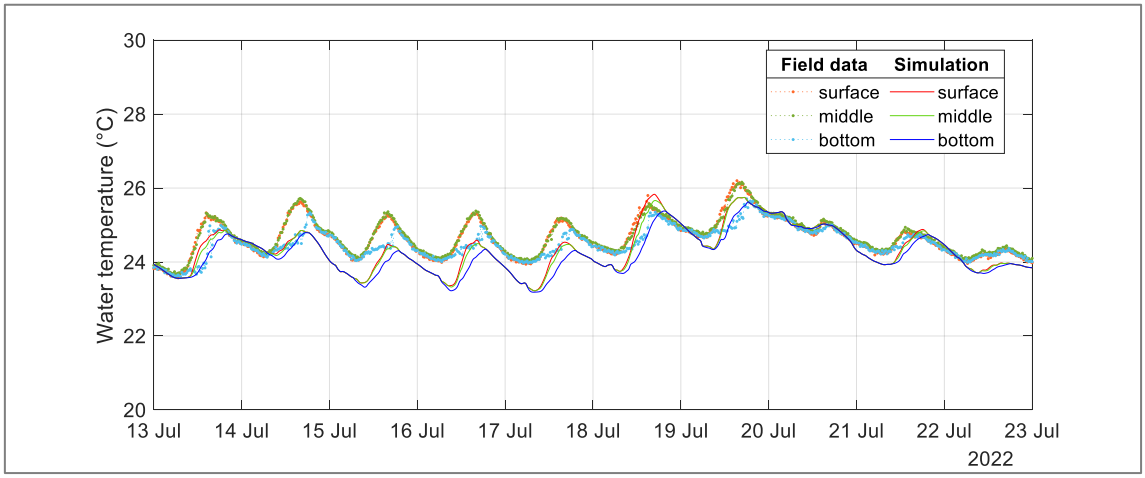


Figure 4.73. July 2022: Measured (dotted) and simulated (line) water temperature in surface (red), middle (green) and bottom (blue) layers. Data from 13 to 23/07/2022.

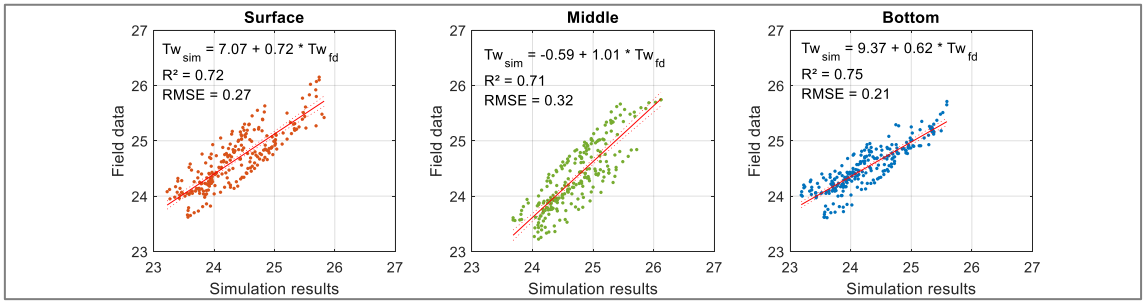


Figure 4.74. July 2022: Relationship between simulated and measured water temperature at surface (orange, left), middle (green, middle) and bottom (blue, right) layers. Data from 13 to 23/07/2022.

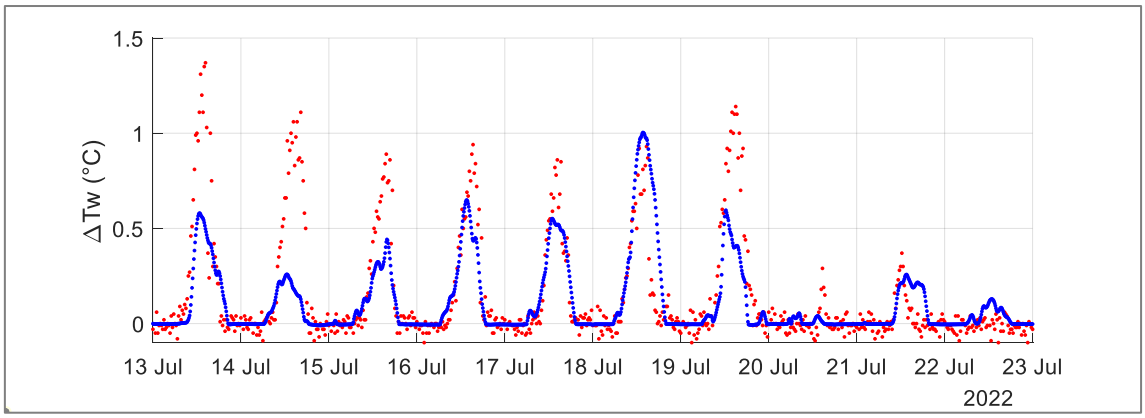


Figure 4.75. July 2022: Difference between surface and bottom temperatures (ΔTw) of field data (red) and simulation results (blue), at point B. Data from 13 to 23/07/2022.

In both hot weather periods of June 2021 and July 2022, the water temperature range was from 21 to 26°C. The water column was stratified during the day, with a difference between surface and bottom temperatures of up to 1 °C. At every night, the water column was mixed.

The evolution of the simulated and measured water temperature at the three depths is very close. A good agreement between field data and simulation results of water temperature was obtained in both periods (RMSE<0.4°C; R^2>0.7).

In June 2021, the duration of the daily thermal stratification of the water column of the model results was shorter than the field measurements. On the other hand, in July 2022, the duration of the thermal stratification in field and simulated data were identical. In both simulated periods, the thermal amplitude between surface and bottom water temperature was underestimated by the model.

These differences between field data and simulation results may be due to the meteorological input data. Roissy station is 18 km northeast the study site, and the meteorological conditions such as air temperature, relative humidity, local rainfall and wind conditions may by significantly different. These differences impact the results of the heat fluxes between water surface and atmosphere. The atmospheric radiation flux, for example, in greatly impacted by the air temperature, a variable raised to power 4, and the nebulosity, raised to power 2. A reliable conclusion would deserve a detailed sensitivity analysis on the meteorological forcing.

Nevertheless, the results are within the incertitude of the field equipment (around 0.2 °C between our different instruments, as presented in section 4.1.3) and can be considered satisfactory.

In order to better characterise the thermal stratification, it may be necessary refine the vertical resolution of the water column, in particular near the probable thermocline zone, between 1.00 and 2.00 m depth. In addition, a calibration of the vertical diffusion coefficient of the tracers may contribute to correctly reproduce the thermocline. The use or the development of an algorithm for coefficient calibration is recommended.

4.4.3.2. Transfer time between upstream and downstream

The validation of the velocity and transfer time between upstream and downstream points is illustrated by two simulated periods: June 2021 (01-10/06/2021) and September 2022 (03-23/09/2022). The results are presented in this section. It is important to highlight that, when the WAQTEL thermal module is not activated, the transfer time is almost constant, around 15h.

- *June 2021 (01-10/06/2021)*

The first simulated period is from 01/06/2021 to 10/06/2021 (9 days), where a rainfall of 37.9 mm in 24 h was observed on the 3rd of June at Le Bourget meteorological station, 9 km northeast the study site (Figure 4.76-a).

The range of the measured and simulated conductivity at the bathing area (point B) is between 600 and 800 µS.cm^{-1} (Figure 4.76-b). The same range was observed at upstream measurements (point A). At the beginning of the simulated period, the conductivity is around 780 µS.cm^{-1}. Following the rainfall event, a continuous decrease was observed, dropping to 610 µS.cm^{-1} in three days. Afterwards, the conductivity increased continuously to 760 µS.cm^{-1} for also three days.

The conductivity variation observed at point B was well reproduced. The simulation results and the field data show a good agreement (RMSE=9 µS.cm^{-1}, R^2=0.99, Figure 4.77).

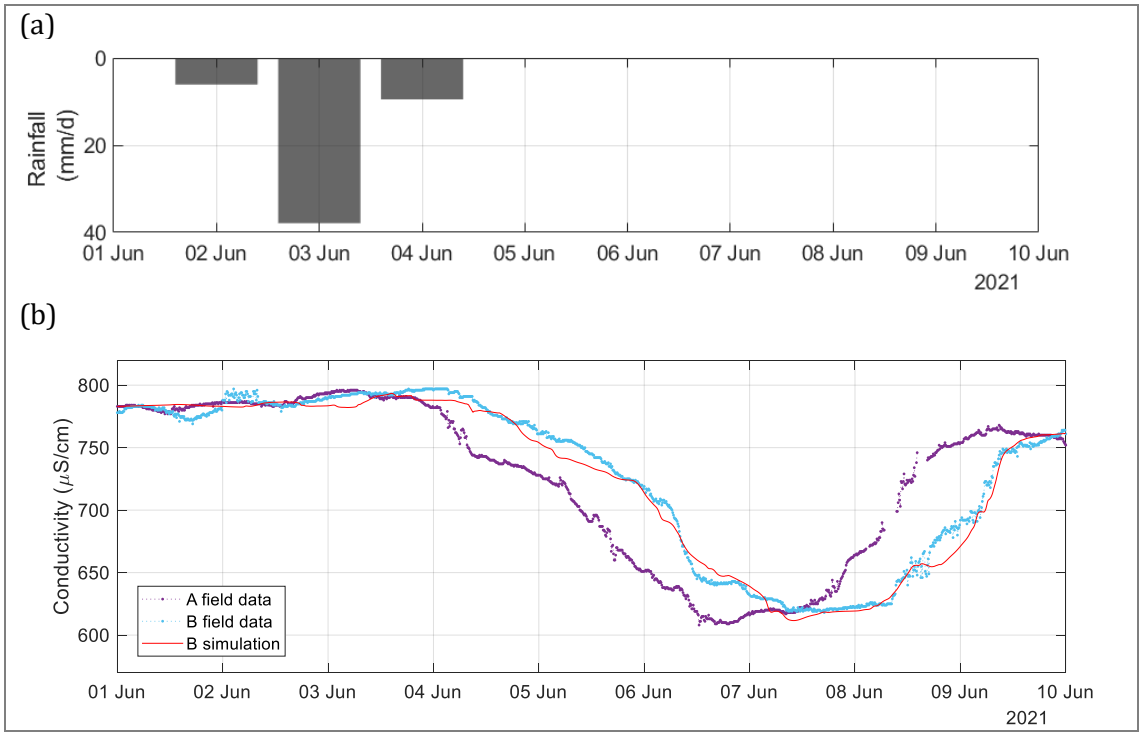


Figure 4.76. June 2021: (a) Daily rainfall from Le Bourget station. (b) Measured conductivity at point A (upstream, purple) and measured (light blue) and simulated (red) conductivity at point B (downstream). Data from 01 to 10/06/2021.

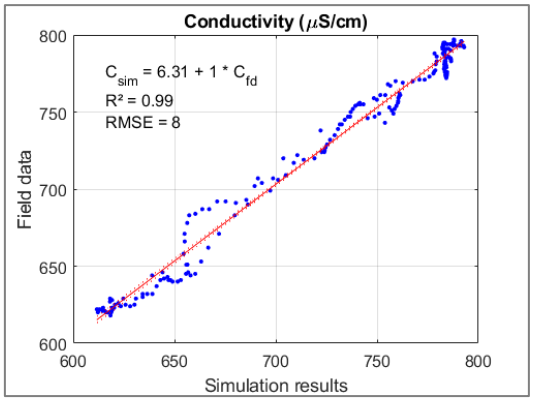


Figure 4.77. June 2021: simulated and measured conductivity at mid-depth. Data from 01 to 10/06/2021.

- *September 2022 (03-23/09/2022)*

The second simulated period is from 03/09/2022 to 23/09/2022 (20 days), where a rainfall of 12.6 mm in 24 h was observed on the 5th of September, followed by a second rainfall episode of 9.3 mm in 48 h between the 7th and the 8th of September (Figure 4.78-a, data from Le Bourget station).

The range of the measured and simulated conductivity at the bathing area (point B) is between 580 and 680 µS.cm^{-1} (Figure 4.78-b). The same range was observed at upstream measurements (point A). During the simulated period, variations greater than 20 µS.cm^{-1} were observed, with marked local peaks and valleys. In the beginning of the simulated period, the conductivity was around 600 µS.cm^{-1} at both points A and B. After the first rainfall event, the conductivity decreased to 580 µS.cm^{-1} for three days, increased for one day and then decreased again after the second rainfall event. Afterwards, it increased for three days and then oscillated between 620 and 670 µS.cm^{-1} until the end of the simulated period.

The conductivity variation observed at point B was well reproduced. The simulation results and the field data show a good agreement (RMSE=3 µS.cm^{-1}, R^2=0.99, Figure 4.79).

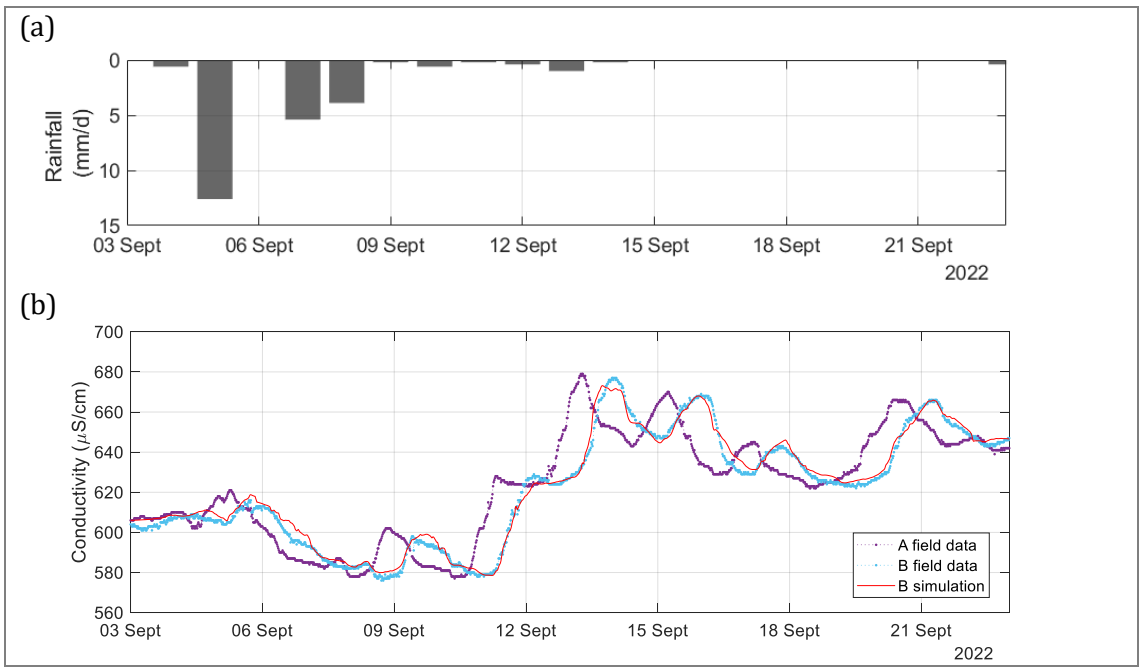


Figure 4.78. September 2022: (a) Daily rainfall from Le Bourget station. (b) Measured conductivity at point A (upstream, purple) and measured (light blue) and simulated (red) conductivity at point B (downstream). Data from 03 to 21/09/2022.

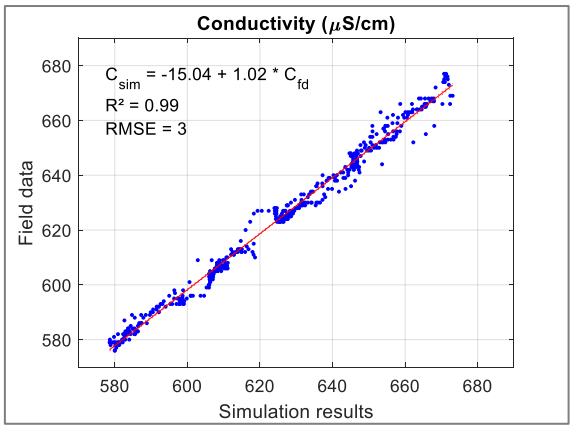


Figure 4.79. September 2022: simulated and measured conductivity at mid-depth. Data from 03 to 23/09/2022.

The conductivity variation at upstream was observed at downstream hours later. The time-lag between nine well-defined local peaks and valleys at both points A and B were used to estimate the transfer time (Figure 4.80-a) and validate the mean velocity field between both points (Figure 4.80-b).

The values estimated from the model results are close to the field data, with similar ranges (Table 4.22). The range of the transfer time from field data was [11h40-20h00], and [11h30-22h20] from the model. The mean velocity range from field data was [1.4-2.6] cm.s^{-1}, and [1.3-2.6] cm.s^{-1} from the model.

The results indicated that the mean velocity field between points A and B was well reproduced. Considering the 9 values corresponding to the inflexion points of the conductivity evolution, a good agreement between field data and model results was observed for both time lag (RMSE=33min, R^2=0.97) and, consequently, mean velocity (RMSE=0.07 cm.s^{-1}, R^2=0.97).

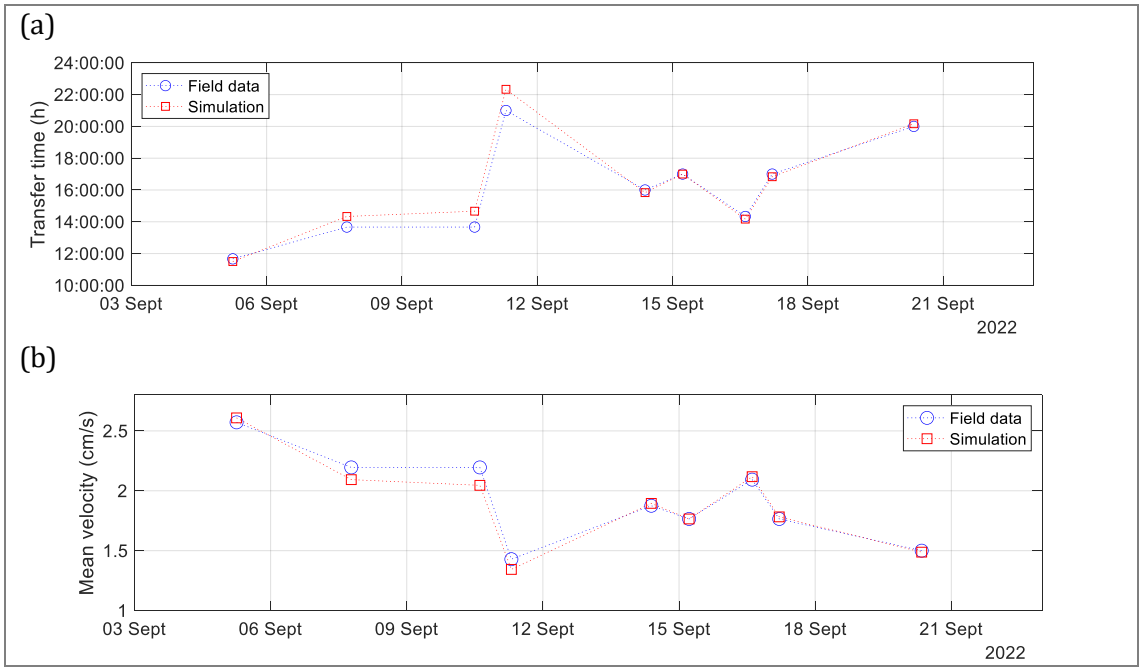


Figure 4.80. September 2022: (a) Mean transfer time and (b) mean velocity obtained from the time-lag of well-defined conductivity peaks at points A and B, from field data (blue) and from simulation results (red). Data from 03 to 21/09/2022.

Table 4.22. Transfer time and mean velocity estimation at mid-depth from inflexion points of conductivity evolution observed in field data and simulation results, in September 2022 (03-23/09/2022).

Datetime of conductivity inflexion			Transfer time (hh:mm)		Mean velocity (cm.s^{-1})	
Point A	Point B	Sim. result	Field data	Sim.	Field data	Sim.
05/09/2022 06:00	05/09/2022 17:40	05/09/2022 17:30	11:40	11:30	2.6	2.6
07/09/2022 18:40	08/09/2022 08:20	08/09/2022 09:00	13:40	14:20	2.2	2.1
10/09/2022 14:40	11/09/2022 04:20	11/09/2022 05:20	13:40	14:40	2.2	2.0
11/09/2022 07:20	12/09/2022 04:20	12/09/2022 05:40	21:00	22:20	1.4	1.3
14/09/2022 09:20	15/09/2022 01:20	15/09/2022 01:10	16:00	15:50	1.9	1.9
15/09/2022 05:20	15/09/2022 22:20	15/09/2022 22:20	17:00	17:00	1.8	1.8
16/09/2022 14:40	17/09/2022 05:00	17/09/2022 04:50	14:20	14:10	2.1	2.1
17/09/2022 05:00	17/09/2022 22:00	17/09/2022 21:50	17:00	16:50	1.8	1.8
20/09/2022 08:20	21/09/2022 04:20	21/09/2022 04:30	20:00	20:10	1.5	1.5

In both simulated periods of June 2021 and September 2022, the conductivity was very well reproduced by the model at downstream. A very good agreement was obtained in both periods (RMSE<8 µS.cm^{-1}, R^2=0.99).

In September 2022, nine local well-defined peaks and valleys at both points A and B were identified. Their time lag from upstream to downstream was used to estimate the mean transfer time and the mean velocity between points A and B. The model was able to correctly simulate the conductivity evolution to track water quality changes after a rainfall episode in the catchment upstream La Villette system.

4.4.4. Distribution of faecal contamination after rainy events

In order to illustrate the aptitude of the model to simulate a faecal contamination in the bathing area after a rainy event, four periods were simulated (Table 4.23). In all simulations, the indicator of faecal contamination used as input was *E. coli*.

The estimation of *E. coli* was obtained from two expressions. The first, from the relationship with FDOM, where *E. coli* can be inferred from measurement of the fluorescence component C6. This relationship (R^2=0.65, *p-value*<0.001) was established through the data set of the 10 samples collected between June and September 2021 in La Villette basin (Equation 4.13). The range of *E. coli* data was from 98 to 1400 MPN.100mL^{-1}. The range of C6 data was from 0.08 to 0.53 RU.

$$[E.coli] = 2177 \cdot C6 \qquad 4.13$$

Where [*E. coli*] is given in MPN.100mL^{-1} and C6 in RU.

The second expression comes from the relationship with precipitation, where *E. coli* can be estimated from the mean precipitation of one to three days before the day of interest P_{1-3} (presented in section 4.3.4).

The estimation of *E. coli* from C6 fluorescence measurements was applied to two simulated periods. The first was in September 2021 (12-22/09/2021), with a total rainfall of 31.3 mm over 48h on September 14th and 15th. The second was in June 2022 (20-30/06/2022), where a total rainfall of 10.5 mm occurred on June 20th and 3.2 mm on June 23rd, with an increase of the fluorescent component C6 on the 23rd. The C6 values were linearly interpolated at an hourly time-step to be used as model input.

The estimation *E. coli* from precipitation P_{1-3} was applied to the other two simulated periods. The first was in June 2021 (01-10/06/2021), with a total rainfall of 47.4 mm over 48h on June 3rd and 4th. The second was in July 2021 (10-20/07/2021), where a total rainfall of 49.2 mm occurred in 24h on the July 13th.

The expression representing bacterial decrease considers a decay rate (Equation 2.3 from section 2.4.2.1). However, to exclusively assess the effect of thermal stratification on the spatiotemporal distribution of faecal contamination, no decay rate (k=0) was considered in the following simulations.

The position of the water column of point B corresponds to the node 3311 at layer 1, the bottom layer with bathymetry information. The position of the water column of point C corresponds to the node 459 at layer 1.

The simulation results of the correspondent nodes at layers 3 (bottom) and 6 (surface) were associated, respectively, with field data at 2.00 m (bottom) and 0.50 m (surface).

Table 4.23. Synthesis of simulated periods with *E. coli* for faecal contamination simulation.

Short name of period	Simulated period	Output variables of interest	Secchi depth
June 2021	01 to 10/062021	*E. coli* from P_{1-3} relationship	1.5 m
July 2021	10 to 20/07/2021	*E. coli* from P_{1-3} relationship	0.9 m
September 2021	12 to 24/09/2021	*E. coli* from C6 relationship	0.9 m
June 2022	20 to 30/06/2022	*E. coli* from C6 relationship	1.2 m

4.4.4.1. Estimation of *E.* coli from FDOM measured at upstream

In this section, fluorescence measurements were used to estimate *E. coli* input on the simulations of faecal contamination in the bathing area after a rainy event.

The *E. coli* input was estimated from the relationship with C6 established for La Villette system (Equation 4.13). To provide the upstream input, C6 data was linearly interpolated at an hourly time step.

Two periods were simulated: September 2021 (12-24/09/2021) and June 2022 (20-30/06/2022). These periods were chosen in function of the availability of daily C6 data from upstream (point A). The simulation results are presented in the following paragraphs.

- *September 2021 (12-24/09/2021)*

From 12/09/2021 to 24/09/2021, a rainfall of 31.3 mm in 48 h was observed between the 14th and 15th of September at Le Bourget meteorological station (Figure 4.81).

The simulated *E. coli* was compared to field data measured at point C (Figure 4.82), for a first validation of the combined use of the model and the expression for *E. coli* estimation.

All the field data were within the 95% confidence interval of the simulation results (RMSE=339 MPN.100mL^{-1}, R^2=0.43, p-value=0.16).

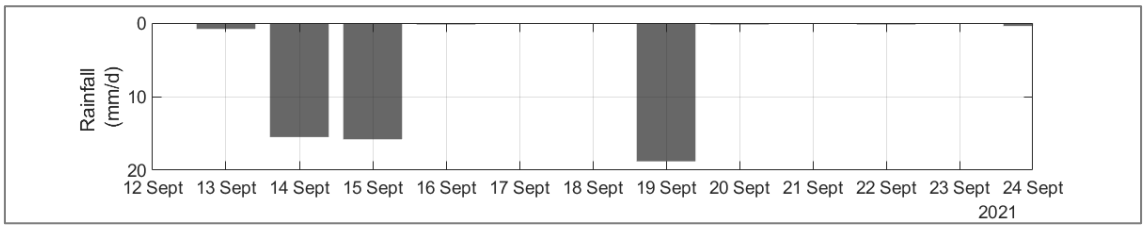


Figure 4.81. September 2021: Daily rainfall at Le Bourget station.

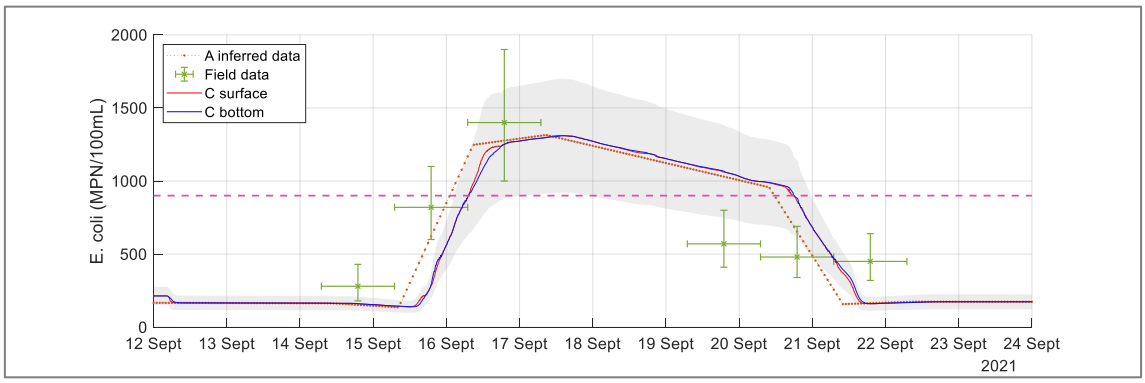


Figure 4.82. September 2021: *E. coli* simulation results at point C (Crimée bridge, outlet of canal) on surface layer (red), bottom layer (blue) and the confidence interval (grey), with the bathing threshold of 900 MPN.100mL^{-1} (dashed line). Data from 12 to 24/09/2021.

At the bathing area, *E. coli* increased from 160 to 1300 MPN.100mL^{-1} between the 16th and 17th of September. Then, it decreased to 920 MPN.100mL^{-1} between the 17th and the 21st and dropped to 180 MPN.100mL^{-1} on the 22nd (Figure 4.83-a).

According to the simulation results, *E. coli* threshold for bathing (900 MPN.100mL^{-1}) was overpassed between 16/09 at 11h00 and 21/09 at 12h00. Thus, the bathing area would have been closed from 16 to 21/09/2021, reopening on the 22/09/2021 (Figure 4.83-b).

At the moment of threshold overpassing, the water column was stratified (Figure 4.83-a, in yellow). In the surface, the threshold was exceeded 3h earlier than in the bottom. When *E. coli* values decreased at point B, the values were uniform on the water column.

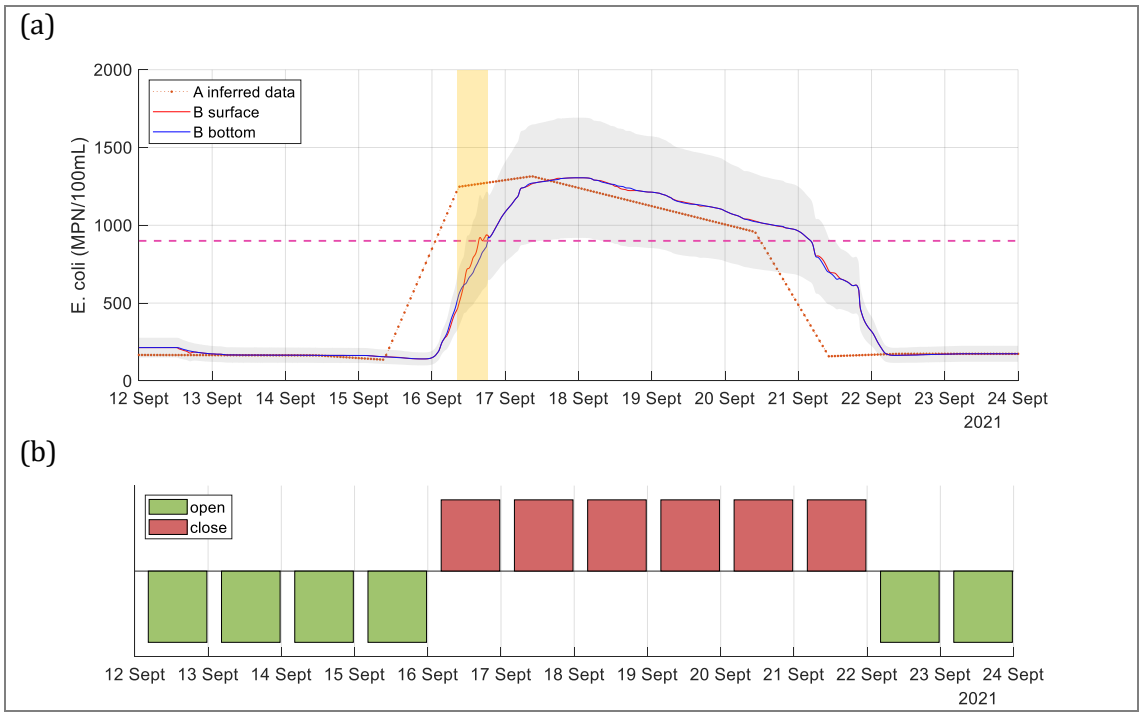


Figure 4.83. September 2021: (a) *E. coli* simulation results at point B (bathing area) on surface layer (red), bottom layer (blue) and the confidence interval (grey). The period of thermal stratification at the overpassing of *E. coli* threshold (dashed line) is highlighted in yellow. (b) Indication of opening (green) and closure (red) days of the bathing area, according to simulation results. Data from 12 to 24/09/2021.

- *June 2022 (20-30/06/2022)*

From 20/06/2022 to 30/06/2022, a rainfall of 12.3 mm in 72 h was observed between the 22nd and the 24th of June at Le Bourget station (Figure 4.84-a).

According to the simulation results at the bathing area (point B), *E. coli* increased from 180 to 630 MPN.100mL^{-1} between the 22/06 and the 24/06. However, considering the confidence interval of *E. coli* estimation, the maximum value was of 1450 MPN.100mL^{-1} on the 24/06 at 2h50, gradually decreasing to 460 MPN.100mL^{-1} on the 25/06 at 6h20 (Figure 4.84-b).

E. coli threshold for bathing (900 MPN.100mL^{-1}) was overpassed between 23/06 at 7h00 and 24/06 at 17h00. Thus, the bathing area should have been closed on the 23/06 and 24/06, reopening on the 25/06/2022 (Figure 4.84-c).

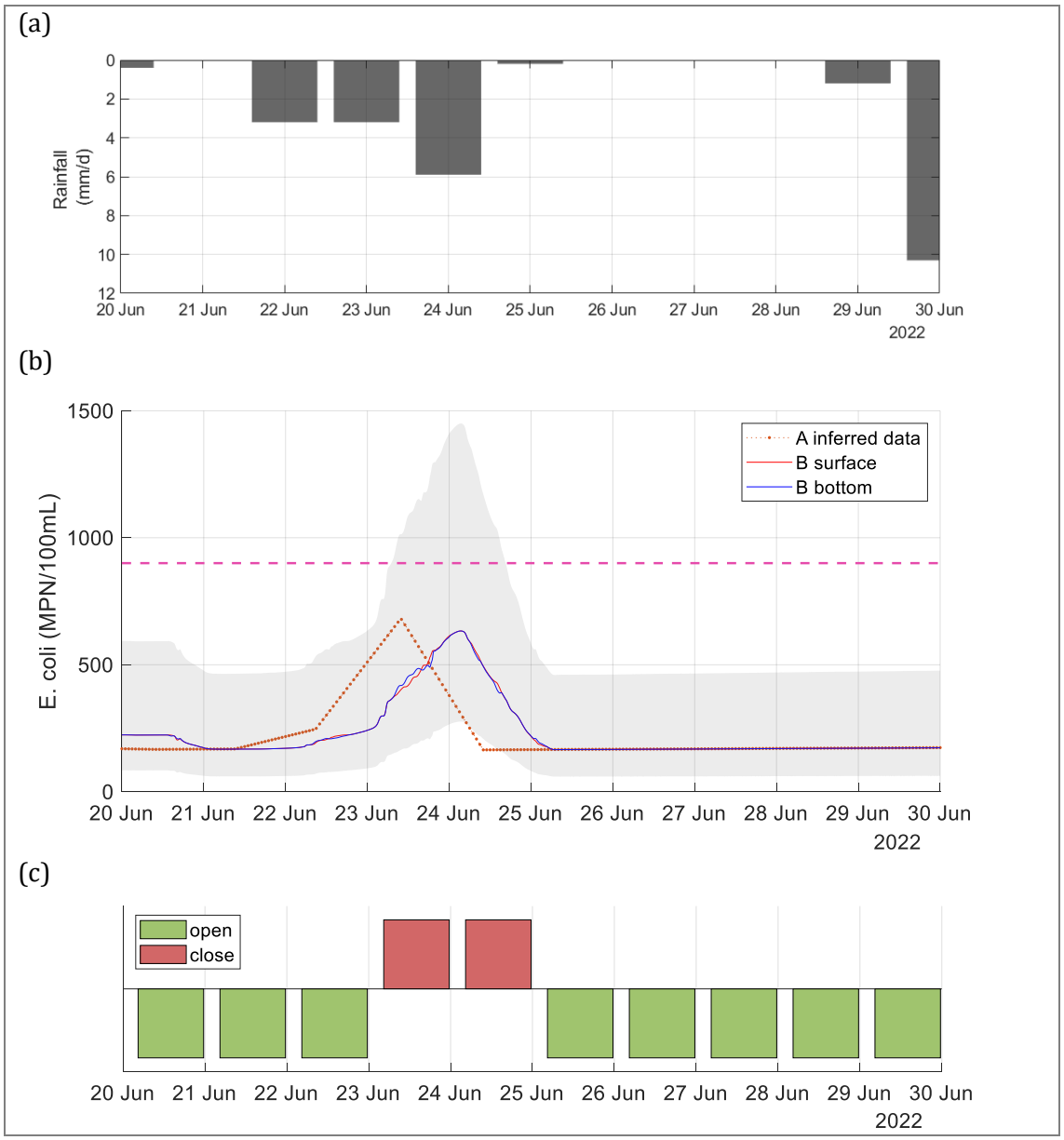


Figure 4.84. June 2022: (a) Daily rainfall at Le Bourget station. (b) *E. coli* simulation results at point B (bathing area) on surface layer (red), bottom layer (blue) and the confidence interval (grey) , with the bathing threshold of 900 MPN.100mL^{-1} (dashed line). (c) Indication of opening (green) and closure (red) days of the bathing area, according to simulation results. Data from 20 to 30/06/2022.

The lateral heterogeneity of *E. coli* distribution at the basin area is illustrated on Figure 4.85. On the 23/06/2022, at 12h30, similar distribution is observed in left and right banks. However, at 14h30, the stratification at La Villette basin was more accentuated. At the bottom layer, higher concentration reached earlier the left bank, where the bathing area is located.

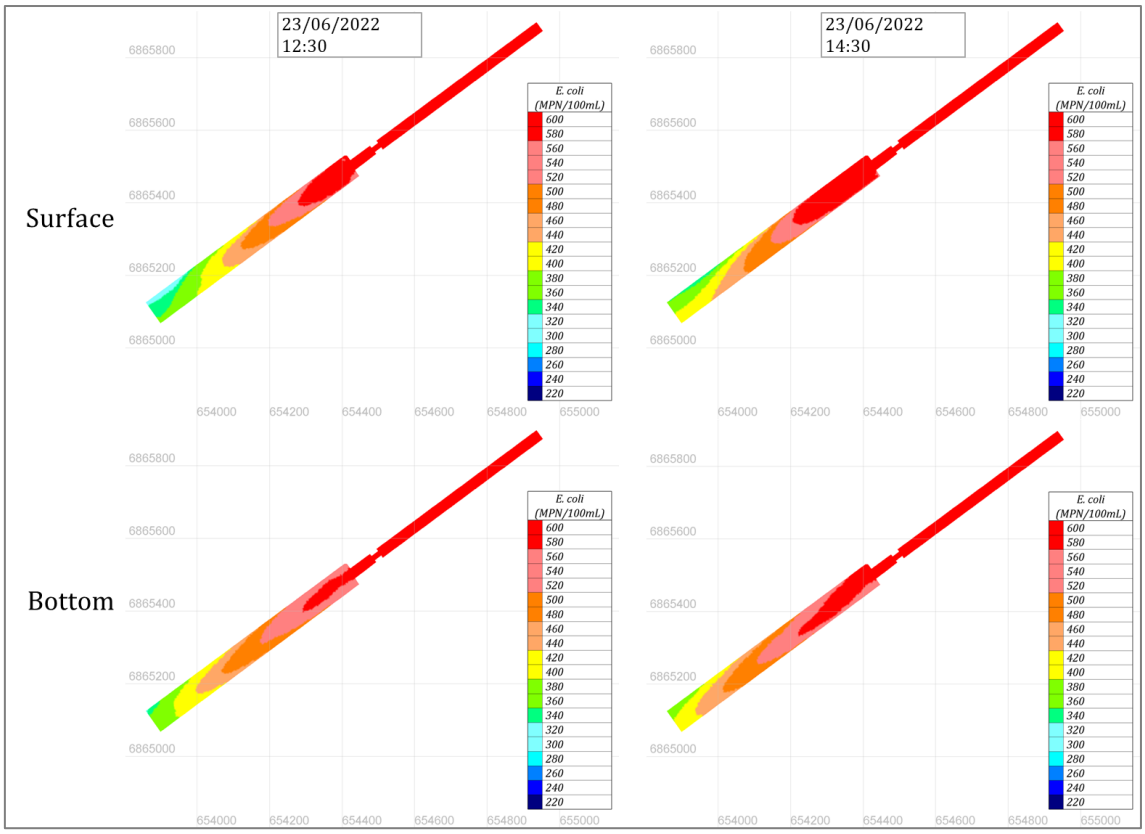


Figure 4.85. June 2022: Spatiotemporal distribution of *E. coli* on the surface (top) and bottom (bottom) layers on 23/06/2022 at 12h30 (left) and 14h30 (right).

The use of the model to compute the transfer of a bacterial contamination consecutive to a rainfall episode allowed to better understand its distribution in time and space. Also, it was possible to estimate the closure duration of the bathing area.

The model results of June 2022 showed that faecal contamination may arrive at the bathing area first from the bottom of the water column. The samples for FIB measurement are collected at the surface, so a contamination from the bottom would not be detected. Also, the lack of UV radiation at the bottom favours the preservation of faecal contamination and active pathogens. Finally, once the water column is mixed, the contamination at the bottom will be present on the surface, exposing the bathers to sanitary risks.

Therefore, to better understand the transport of faecal contamination revealed to be essential on the decision-making to prevent sanitary risks. For that, a modelling tool capable of simulating continuous concentrations of FIB plume concentrations in time and space is of great importance, as already pointed out by Van et al. (2022).

4.4.4.2. Estimation of *E.* coli from precipitation during summer 2021

In this section, the rainfall measurements from Le Bourget meteorological station were used to estimate *E. coli* input on the simulations of faecal contamination in the bathing area after a rainy event.

The *E. coli* input was estimated from the relationship with $P_{1\text{-}3}$, the mean daily rainfall of 1 to 3 days before the day of interest (Equation 4.5, section 4.3.4.2). This equation is intended to be used as a support of the decision-making of the closure of the bathing area from forecasted rainy events.

Two periods were simulated: June 2021 (01-10/06/2021) and July 2021 (10-20/07/2021). These periods were chosen in function of the observed rainfall intensity during summer.

- *June 2021 (01-10/06/2021)*

From 01/06/2021 to 10/06/2021, a rainfall of 37.9 mm in 24 h was observed on the 3rd of June at Le Bourget station (Figure 4.86-a).

According to the simulation results in the bathing area (point B), *E. coli* increased from 100 to 1005 MPN.100mL^{-1} between the 02/06 and the 05/06, overpassing the bathing threshold (900 MPN.100mL^{-1}) between the 05/06 at 10h20 and the 06/06 at 20h30 (Figure 4.86-b).

However, considering the confidence interval of *E. coli* estimation, the bathing threshold was exceeded between the 05/06 at 00h20 and 08/06 at 01h00, with maximum value of 1820 MPN.100mL^{-1} on the 05/06 at 18h30. Therefore, the bathing area should have been closed on the 05/06, reopening on the 08/06 (Figure 4.86-c).

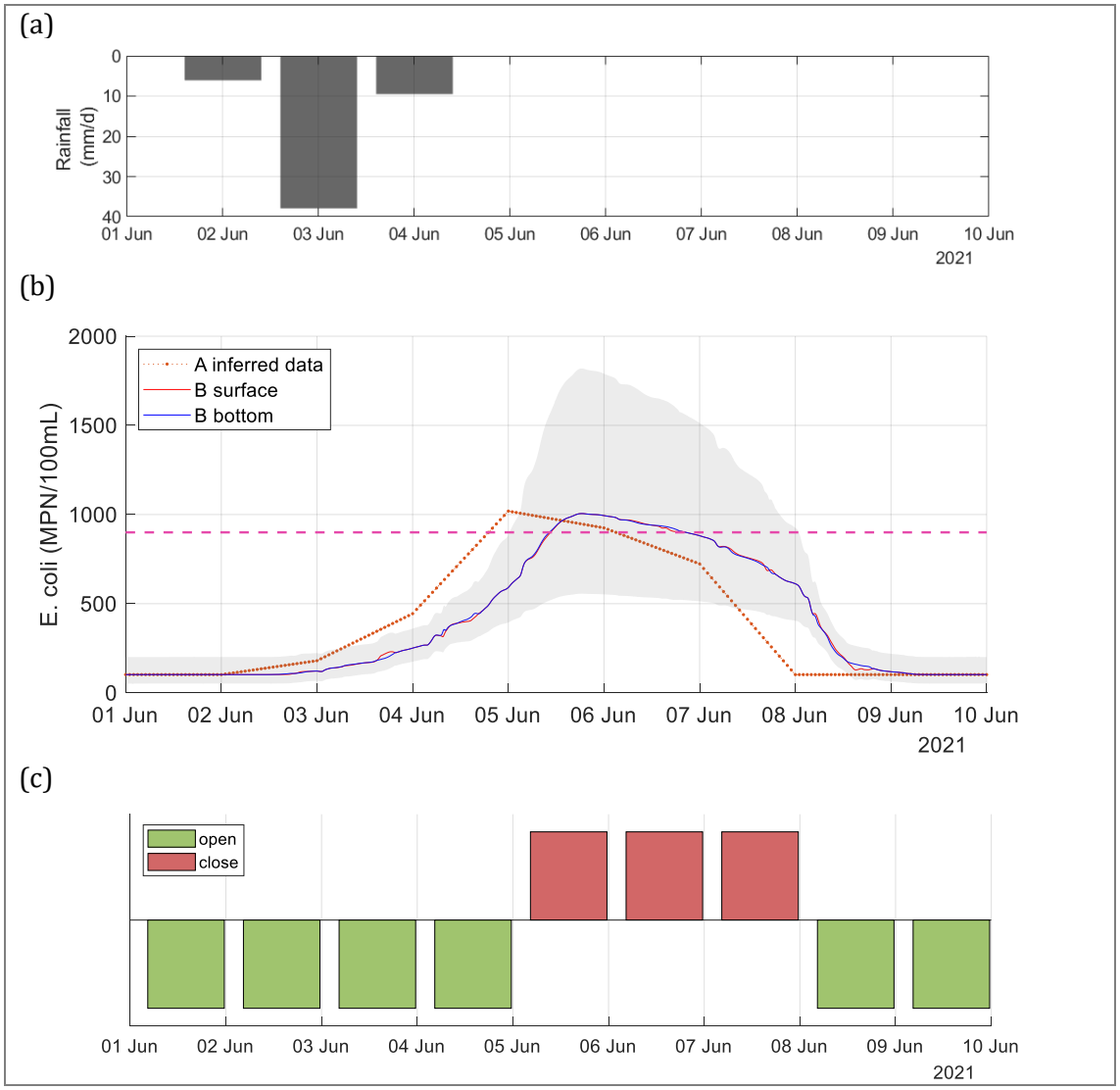


Figure 4.86. June 2021: (a) Daily rainfall at Le Bourget station. (b) *E. coli* simulation results at point B (bathing area) on surface layer (red), bottom layer (blue) and the confidence interval (grey) , with the bathing threshold of 900 MPN.100mL^{-1} (dashed line). (c) Indication of opening (green) and closure (red) days of the bathing area, according to simulation results. Data from 01 to 10/06/2021.

- *July 2021 (10-20/07/2021)*

From 10/07/2021 to 20/07/2021, a rainfall of 49.2 mm in 24 h was observed on the 13th of July at Le Bourget station (Figure 4.87-a).

According to the simulation results in the bathing area, *E. coli* increased from 100 to 1247 MPN.100mL^{-1} between the 12/07 and the 15/07, overpassing the bathing threshold (900 MPN.100mL^{-1}) between the 14/07 at 07h20 and the 16/07 at 20h00 in the surface layer and at 21h20 in the bottom (Figure 4.87-b).

However, considering the confidence interval of *E. coli* estimation, the bathing threshold was exceeded between the 13/07 at 22h20 and 17/07 at 05h20, with maximum value of 2465 MPN.100mL^{-1} on the 15/07 at 17h20. Therefore, the bathing area should have been closed on the 14/07, reopening on the 17/07 (Figure 4.87-c).

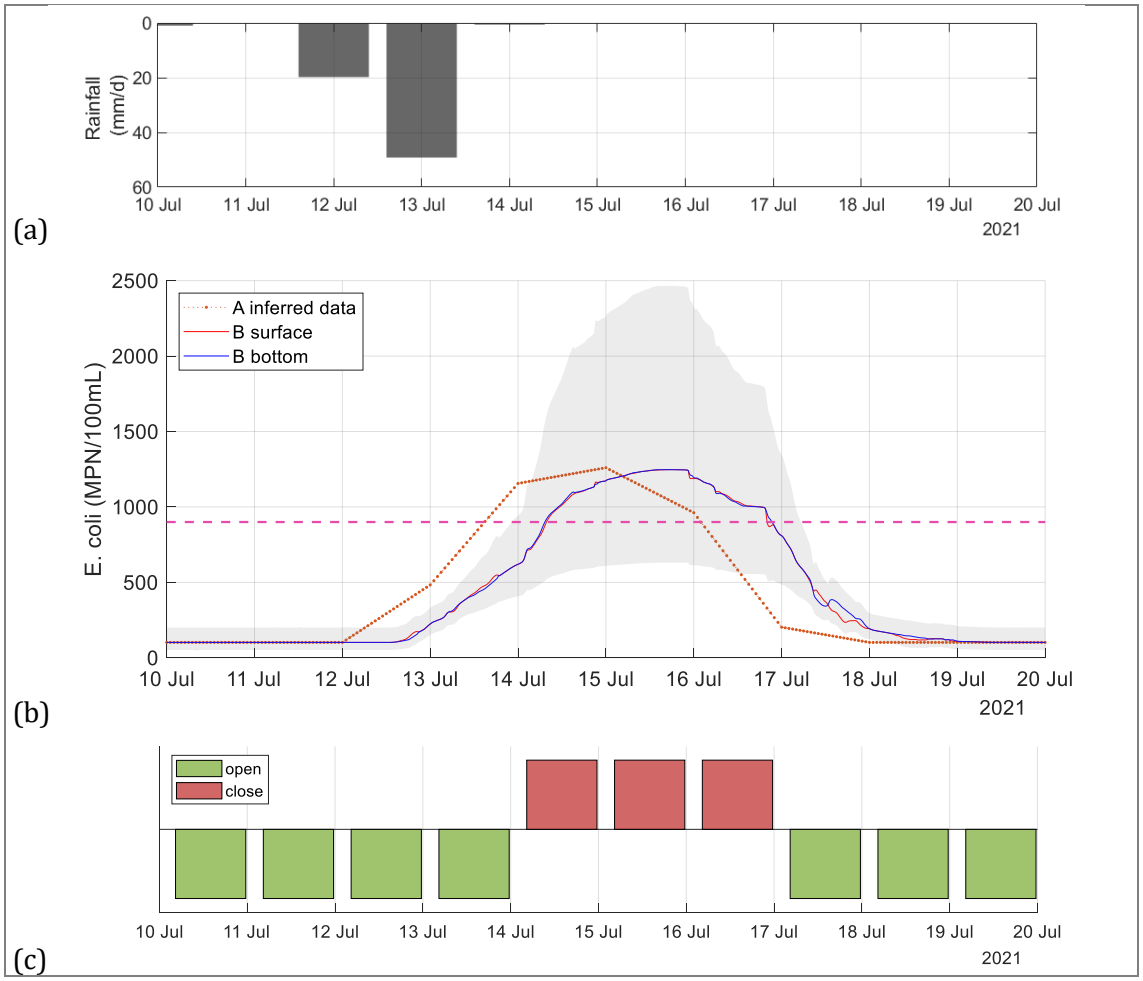


Figure 4.87. July 2021: (a) Daily rainfall at Le Bourget station. (b) *E. coli* simulation results at point B (bathing area) on surface layer (red), bottom layer (blue) and the confidence interval (grey), with the bathing threshold of 900 MPN.100mL^{-1} (dashed line). (c) Indication of opening (green) and closure (red) days of the bathing area, according to simulation results. Data from 10 to 20/07/2021.

The obtained *E. coli* values should be considered qualitatively, as an indicator of whether the bathing threshold of 900 MPN.100mL^{-1} was exceeded.

Nevertheless, the combined use of the model and the estimation of *E. coli* from the rainfall, a predicted variable of easy obtention, proved to be of great interest. It can be applied to anticipate the need of closure of the bathing area after heavy rainfall events, and when it could be reopened. It has the potential to be a valuable decision-making support tool on the closure/opening of the bathing area.

The evolution of water temperature was well simulated (RMSE<0.4°C; R^2>0.7) in both studied periods. The daily thermal stratification and the nightly mixing of the water column were well reproduced. However, the thermal amplitude between surface and bottom layers was underestimated by the model. It can be caused by the different weather forcing used as input and the local conditions. A reliable conclusion would deserve a detailed sensitivity analysis.

To better characterise the thermocline in the water column, a more refined vertical resolution between 1 m and 2 m depth and a calibration of the vertical diffusion coefficient of the tracers would be necessary.

The conductivity evolution from upstream to downstream was very well simulated (RMSE<8 $\mu S.cm^{-1}$; R^2=0.99) in both studied periods. The transfer time estimation with field data and with simulation results was very close (RMSE=33min; R^2=0.97), indicating a good estimation of the mean flow velocity ([1.4-2.6] $cm.s^{-1}$), corresponding to transfer time between 11h and 22h.

An episode when field data of *E. coli* concentration were available was simulated. The transfer of *E. coli* was correctly assessed. The simulation provided the duration of the overpass of the regulatory threshold for bathing activities at the bathing area.

These results show the capability of TELEMAC-3D hydrodynamic model to simulate the impact of hydro-meteorological conditions on the spatiotemporal distribution of contaminants. Combined with a continuous monitoring station, it could be implemented as an early warning system for decision making and sanitary risk attenuation.

Chapter 5. Conclusion and perspectives

Urban bathing helps mitigate urban heat island effects in metropolitan areas and reduce the impact of heatwaves on the local population. However, open water swimming can expose bathers to a risk of waterborne diseases. To control these risks, the concentration of faecal indicator bacteria (FIB), such as *E. coli*, is used to classify the microbiological water quality, and a threshold of satisfactory water quality for bathing activities is established. The reference method for enumerating FIB has a long response time, of over 24 hours.

To determine whether a bathing area in a watercourse should be closed, the analysis of microbiological water quality is conducted upstream of the area, taking into account the transfer time between the sampling site and the bathing area. However, hydro-meteorological conditions, such as variations in flow rate and thermal stratification, can affect the spatiotemporal distribution of FIB and, consequently, the transfer time. Thus, the arrival time and duration of contamination at the bathing area can vary.

An indicator of faecal contamination of fast response time is necessary to increase the frequency of monitoring. A real-time indicator would reduce the risk of undetected contamination between measurements and improve the accuracy of decision making for determining whether or not to close or reopen a bathing area.

In this context, two research lines have been developed in this PhD thesis, aiming to propose a system for monitoring and forecasting the microbiological quality of bathing areas in urban watercourses. The first involves the implementing a 3D hydrodynamic model for velocity field data and water temperature computation. The second focuses on defining a FIB proxy for faster microbiological water quality assessment.

The specific objectives of this study are: (i) implementing the 3D hydrodynamic model TELEMAC-3D, coupled with the WAQTEL thermal module; (ii) designing and installing a monitoring system to collect field data that can be used as input for and validation of the hydrodynamic model; (iii) investigating the impact of thermal stratification on the spatial and temporal distribution of faecal contamination; (iv) identifying a fast method for measuring FIB.

La Villette study site (Paris, France) was equipped with an autonomous monitoring system with two monitoring points: upstream (point A) and at the bathing area (point B). The system was of easy implementation and equipped with a data transfer platform for high-frequency measurements of 20-min time step. The data collected for 20 months was explored. The autonomous system is powered by a photovoltaic panel and battery power system.

The measured physicochemical variables were water temperature at three depths, conductivity (specific conductance at 25 °C) at mid-depth and water level. All variables were validated with reference probes, with $R^2 \geq 0.99$ for all sensors and RMSE ≤ 0.21 °C for temperature sensors and RMSE ≤ 8 $\mu S.cm^{-1}$ for conductivity sensors.

To assure data quality and the reliability of readings, the autonomous system requires regular field campaigns for sensor cleaning and measurement validation. In our system, a regular frequency of once a month, increasing to twice a month during summer due to the higher biofilm formation, proved to be enough. The need of sensor cleaning could be detected with the increase of noise of the conductivity readings.

High-frequency monitoring allowed the observation of hourly, daily, and seasonally variations of the measured physicochemical variables. In La Villette basin, from spring to mid-autumn, a thermal stratification of the water column occurs in a daily basis, breaking

up during the night. The greatest amplitude of thermal stratification was around 2 °C. Large conductivity variations with well-defined inflexion points could be used to estimate transfer time between upstream and downstream points.

The implemented monitoring system was designed to be able to include more sensors. It will improve the surveillance of the long-term evolution of measured variables. It will help measuring the impact of climate change in this urban system on the next years, and possible solutions to mitigate climate change impacts.

A fast method for measuring FIB was explored through the analysis of fluorescence of dissolved organic matter (FDOM). In the collected samples, La Villette basin presented very few elevated *E. coli* values. Other sampling sites were included in the study to cover a wider range of urban waters: the river Marne, two stormwater networks (SN) outlets and one wastewater treatment plant (WWTP) outlet.

The FDOM analysis of the sampling sites resulted in a validated 7-component PARAFAC model, which proved to be effective in characterizing the different types of studied waters. Components C6 and C7, identified as protein-like, as well as a biological index (BIX), were identified as indicative of microbial loadings. These components may serve as valuable markers for the assessment of microbiological contamination in urban waterbodies, particularly after heavy rainfall events.

In SN and WWTP outlets, component C6 exhibited high values. In agreement with previous studies, protein-like fluorescence components, similar to C6, could be used as effective indicators for monitoring microbiological water quality. An increase of C6 level would suggest a microbial contamination and deserves further attention for managing water resources.

The results highlighted the importance of FDOM analysis for assessing microbial loading and water quality. The identification of protein-like fluorescence components, especially C6, provided valuable insights for microbiological water quality monitoring. It may help in the development of effective strategies for microbiological water quality monitoring.

In further studies, measurements of additional parameters such as dissolved organic carbon (DOC) or ammonium could indicate wastewater contamination and therefore could validate the relationship with C6.

In La Villette basin, a fast-measuring proxy for *E. coli* estimation was identified through the protein-like component C6. To confirm that C6 can be an effective indicator for monitoring microbiological water quality, more *E. coli* data in both dry and wet weather are necessary.

For a short-term (3 days) forecast of *E. coli* level, three hydro-meteorological variables were considered: a “natural” flow rate upstream to the entrance of the canal (100 km upstream), rainfall height, and duration of dry weather. The strongest relationship with *E. coli* was with the mean rainfall over one to three days before sampling ($P_{1\text{-}3}$), measured at the meteorological station Le Bourget (9 km northeast).

The *E. coli* estimation from $P_{1\text{-}3}$ gives a daily mean concentration. It does not consider immediate *E. coli* increase related to local runoff. It cannot predict a further increase after storm overflow due to heavy rainfall in a more distant catchment. Nevertheless, it indicated that the impact of a heavy precipitation may last at least for three days. Also, it was able to detect the overpassing of *E. coli* bathing threshold.

Water temperature and solar radiation should be included in the forecasting model to study their relevance on the *E. coli* decay rate. In addition, in order to have more representative rainfall values that impact the upstream watershed from La Villette canal, additional upstream meteorological stations should be considered. This would also enable to track

upstream contamination related to stormwater outlets discharges after a heavy rainfall event at an upstream watershed.

Both expressions of *E. coli* estimation are based on a low range of concentrations ([98-1400] MPN.100mL^{-1}). The validation of these relationships would require a larger range of values. To include the estimation and validation of decay rates, *E. coli* measurements at upstream and downstream should be conducted considering the transfer time between both points. According to field data, the samples should be collected every 5h to 10h.

In addition, higher sampling frequency would enable a better characterisation of *E. coli* evolution after a rainfall episode, capturing its increase and decrease. *E. coli* measurements during heavy rainfall events should be conducted outside the summer period to have a wider range of precipitation intensities and better characterize the effects of rainfall in microbial water quality.

The use of high-frequency field data, combined with *E. coli* estimation, was essential to test and validate the implementation of the hydrodynamic model TELEMAC-3D, coupled with the WAQTEL thermal module. It also enabled to investigate the impact of thermal stratification on the distribution of faecal contamination.

The spatiotemporal distribution of faecal contamination is affected by variations in current velocity caused by flowrate and thermal stratification. Additionally, thermal stratification may influence *E. coli* decay rate due to the heterogeneity of water temperature and UV radiation in the water column.

Our results showed the capability of TELEMAC-3D hydrodynamic model to simulate the impact of hydro-meteorological conditions on the spatiotemporal distribution of contaminants. Combined with a continuous monitoring station, it could be implemented as an early warning system for decision making and sanitary risk attenuation. The evolution of water temperature was well simulated (RMSE<0.4°C; R^2>0.7) in both studied periods. However, the thermal amplitude between surface and bottom layers was underestimated by the model. It can be caused by the different weather forcing used as input and the local conditions. A reliable conclusion would deserve a detailed sensitivity analysis.

In a stratified water column, an upstream contamination may first arrive at the bathing area either by the surface layers, but also by bottom layers. The worst scenario is when contamination arrives first by the bottom because: (i) as the samples for FIB measurement are collected below the surface (generally 30 cm), a contamination at the bottom would not be detected; (ii) the lack of UV radiation at the bottom would decrease the decay rate of faecal bacteria and pathogens; and (iii) once the water column is mixed, the contamination at the bottom will be redistributed to the surface.

The mean velocity field was validated through the conductivity evolution from upstream to downstream. It presented excellent results (RMSE < 8 µS.cm^{-1}, R^2 = 0.99). A variation of the transfer time between upstream to downstream was observed. It highlights the importance of a hydrodynamic model, which can be used to better estimate the sanitary conditions at the bathing area with some hours in advance.

Further studies should include the calibration of *E. coli* decay rate, from WAQTEL water quality module. For model validation with field data, *E. coli* frequent measurements at upstream and downstream during and after rainfall events are required.

This work highlights the relevance of a 3D hydrodynamic model for the computation of contaminant transport and the estimation of sanitary conditions in an area of interest in a river. Combined with a continuous monitoring station, it could be implemented as an early warning system for decision making and sanitary risk attenuation.

Bibliography

Angelescu, D.E., Huynh, V., Hausot, A., Yalkin, G., Plet, V., Mouchel, J.-M., Guérin-Rechdaoui, S., Azimi, S., Rocher, V., 2019. Autonomous system for rapid field quantification of *Escherichia coli* in surface waters. J Appl Microbiol 126, 332–343. https://doi.org/10.1111/jam.14066

Angelotti, N., Guillot-Le Goff, A., Carmigniani, R., Vinçon Leite, B., 2022. Open water swimming in urban areas E. coli distribution with TELEMAC-3D, in: Proceedings of the 2022 TELEMAC-MASCARET User Conference. Saclay Paris France, pp. 231–236.

Angelotti, N., Vinçon-Leite, B., Carmigniani, R., 2021. Hydrodynamic modelling for early warning of sanitary risks in open swimming waters, in: Proceedings of the 2020 TELEMAC-MASCARET User Conference. Antwerpen Belgium, pp. 175–180.

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

Auer, M.T., Niehaus, S.L., 1993. Modeling fecal coliform bacteria—I. Field and laboratory determination of loss kinetics. Water Research 27, 693–701. https://doi.org/10.1016/0043-1354(93)90179-L

Bahram, M., Bro, R., Stedmon, C., Afkhami, A., 2007. Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. J. Chemometrics 20, 99–105. https://doi.org/10.1002/cem.978

Baker, A., Andersen, M.S., Marjo, C.E., Zainuddin, N.S., Rutlidge, H., Graham, P.W., Henderson, R.K., 2019. Investigation of Pollution in Rivers and Groundwater by Fluorescence, in: Meyers, R.A. (Ed.), Encyclopedia of Analytical Chemistry. Wiley, pp. 1–14. https://doi.org/10.1002/9780470027318.a9412.pub2

Baker, A., Cumberland, S.A., Bradley, C., Buckley, C., Bridgeman, J., 2015a. To what extent can portable fluorescence spectroscopy be used in the real-time assessment of microbial water quality? Science of The Total Environment 532, 14–19. https://doi.org/10.1016/j.scitotenv.2015.05.114

Baker, A., Cumberland, S.A., Bradley, C., Buckley, C., Bridgeman, J., 2015b. To what extent can portable fluorescence spectroscopy be used in the real-time assessment of microbial water quality? Science of The Total Environment 532, 14–19. https://doi.org/10.1016/j.scitotenv.2015.05.114

Baker, A., Elliott, S., Lead, J.R., 2007. Effects of filtration and pH perturbation on freshwater organic matter fluorescence. Chemosphere 67, 2035–2043. https://doi.org/10.1016/j.chemosphere.2006.11.024

Ballance, R., Bartram, J. (Eds.), 1996. Water Quality Monitoring: A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes. Chapter 10 - Microbiological Analyses. UNEPWHO. https://doi.org/10.4324/9780203476796

Bedell, E., Harmon, O., Fankhauser, K., Shivers, Z., Thomas, E., 2022. A continuous, in-situ, near-time fluorescence sensor coupled with a machine learning model for detection of fecal contamination risk in drinking water: Design, characterization and field validation. Water Research 220, 118644. https://doi.org/10.1016/j.watres.2022.118644

Bedell, E., Sharpe, T., Purvis, T., Brown, J., Thomas, E., 2020. Demonstration of Tryptophan-Like Fluorescence Sensor Concepts for Fecal Exposure Detection in Drinking Water in Remote and Resource Constrained Settings. Sustainability 12, 3768. https://doi.org/10.3390/su12093768

Bezerra, F.M., 2019. Open-water swimming in urban areas: How a three-dimensional hydrodynamic model can help in the microbiological contamination monitoring?

Bhurtun, P., Lesven, L., Ruckebusch, C., Halkett, C., Cornard, J.-P., Billon, G., 2019. Understanding the impact of the changes in weather conditions on surface water quality. Science of The Total Environment 652, 289–299. https://doi.org/10.1016/j.scitotenv.2018.10.246

Boavida, M.-J., Wetzel, R.G., 1998. Inhibition of phosphatase activity by dissolved humic substances and hydrolytic reactivation by natural ultraviolet light. Freshwater Biology 40, 285–293. https://doi.org/10.1046/j.1365-2427.1998.00349.x

Boehrer, B., Schultze, M., 2008. Stratification of lakes. Rev. Geophys. 46, RG2005. https://doi.org/10.1029/2006RG000210

Brigode, P., Génot, B., Lobligeois, F., Delaigue, O., 2020. Summary sheets of watershed-scale hydroclimatic observed data for France. https://doi.org/10.15454/UV01P1

Burnet, J.-B., Dinh, Q.T., Imbeault, S., Servais, P., Dorner, S., Prévost, M., 2019. Autonomous online measurement of β-D-glucuronidase activity in surface water: is it suitable for rapid E. coli monitoring? Water Research 152, 241–250. https://doi.org/10.1016/j.watres.2018.12.060

Chen, H., Liao, Z., Gu, X., Xie, J., Li, H., Zhang, J., 2017. Anthropogenic Influences of Paved Runoff and Sanitary Sewage on the Dissolved Organic Matter Quality of Wet Weather Overflows: An Excitation–Emission Matrix Parallel Factor Analysis Assessment. Environ. Sci. Technol. 51, 1157–1167. https://doi.org/10.1021/acs.est.6b03727

Chow, V. te, 1959. Open-channel hydraulics, International student edition. ed, McGraw-Hill Civil Engeneering Series. McGraw-Hill.

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., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Marine Chemistry 51, 325–346. https://doi.org/10.1016/0304-4203(95)00062-3

Coble, P.G., Schultz, C.A., Mopper, K., 1993. Fluorescence contouring analysis of DOC intercalibration experiment samples: a comparison of techniques. Marine Chemistry, Measurement of Dissolved Organic Carbon and Nitrogen in Natural Waters 41, 173–178. https://doi.org/10.1016/0304-4203(93)90116-6

Cohen, E., Levy, G.J., Borisover, M., 2014. Fluorescent components of organic matter in wastewater: Efficacy and selectivity of the water treatment. Water Research 55, 323–334. https://doi.org/10.1016/j.watres.2014.02.040

ColiMinder VWMS, 2021. Technology – ColiMinder Measurement Approach [WWW Document]. ColiMinder radis microbiology. URL https://www.coliminder.com/technology/ (accessed 4.30.21).

Cory, R.M., McKnight, D.M., 2005. Fluorescence Spectroscopy Reveals Ubiquitous Presence of Oxidized and Reduced Quinones in Dissolved Organic Matter. Environ. Sci. Technol. 39, 8142–8149. https://doi.org/10.1021/es0506962

Costa Medeiros, M.J., 2021. Internship report: Predictive modelling of sanitary risks for early warning systems in La Villette (Paris, France). LEESU - ENPC, IME.

Cumberland, S., Bridgeman, J., Baker, A., Sterling, M., Ward, D., 2012a. Fluorescence spectroscopy as a tool for determining microbial quality in potable water applications. Environmental Technology 33, 687–693. https://doi.org/10.1080/09593330.2011.588401

Cumberland, S., Bridgeman, J., Baker, A., Sterling, M., Ward, D., 2012b. Fluorescence spectroscopy as a tool for determining microbial quality in potable water applications. Environ. Technol. 33, 687–693. https://doi.org/10.1080/09593330.2011.588401

Deltares, 2019. D-Water Quality User Manual.

Edberg, S.C., Rice, E.W., Karlin, R.J., Allen, M.J., 2000. Escherichia coli: the best biological drinking water indicator for public health protection. Journal of Applied

Microbiology 88, 106S-116S. https://doi.org/10.1111/j.1365-2672.2000.tb05338.x
EDF R&D, Telemac3D consortium, 2021a. TELEMAC-3D User manual, Version v8p3. TELEMAC-MASCARET.
EDF R&D, Telemac3D consortium, 2021b. WAQTEL User manual, Version v8p3. TELEMAC-MASCARET.
EDF R&D, Telemac3D consortium, 2021c. TELEMAC-3D Theory guide, Version v8p3. TELEMAC-MASCARET.
EDF R&D, Telemac3D consortium, 2018. TELEMAC-3D User Manual, Version v8p0. TELEMAC-MASCARET.
EEA, 2018. Quality of bathing waters — European Environment Agency [WWW Document]. URL https://www.eea.europa.eu/airs/2018/environment-and-health/bathing-water-quality (accessed 10.23.23).
Environment Agency, 2002. The Microbiology of Drinking Water. Part 1 – Water Quality and Public Health. Methods for the examination of waters and associated materials.
Erichsen, A.C., Kaas, H., Dannisøe, J., Mark, O., Jørgensen, C., 2006. Implementation and description of different early warning systems for bathing water quality Original title: Etablering af badevandsprofiler og varslingssystemer i henhold til EU's nye badevandsdirektiv (in Danish). Miljøprojekt 92.
EU, 2006. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC.
Evans, C.D., Monteith, D.T., Cooper, D.M., 2005. Long-term increases in surface water dissolved organic carbon: Observations, possible causes and environmental impacts. Environmental Pollution, Recovery from acidificationin the UK: Evidence from 15 years of acid waters monitoring 137, 55–71. https://doi.org/10.1016/j.envpol.2004.12.031
Faris, G.W., Copeland, R.A., 1997. Wavelength dependence of the Raman cross section for liquid water. Appl Opt 36, 2686–2688. https://doi.org/10.1364/ao.36.002686
Fellman, J.B., Hood, E., Spencer, R.G.M., 2010. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. Limnology and Oceanography 55, 2452–2462. https://doi.org/10.4319/lo.2010.55.6.2452
Fellman, J.B., Miller, M.P., Cory, R.M., D'Amore, D.V., White, D., 2009. Characterizing Dissolved Organic Matter Using PARAFAC Modeling of Fluorescence Spectroscopy: A Comparison of Two Models. Environ. Sci. Technol. 43, 6228–6234. https://doi.org/10.1021/es900143g
Fraîcheur de Paris, 2023. Réseau de froid urbain [WWW Document]. URL https://www.fraicheurdeparis.fr/reseau-de-froid-urbain (accessed 7.29.23).
Frank, S., Goeppert, N., Goldscheider, N., 2018a. Fluorescence-based multi-parameter approach to characterize dynamics of organic carbon, faecal bacteria and particles at alpine karst springs. Science of The Total Environment 615, 1446–1459. https://doi.org/10.1016/j.scitotenv.2017.09.095
Frank, S., Goeppert, N., Goldscheider, N., 2018b. Fluorescence-based multi-parameter approach to characterize dynamics of organic carbon, faecal bacteria and particles at alpine karst springs. Science of The Total Environment 615, 1446–1459. https://doi.org/10.1016/j.scitotenv.2017.09.095
Galletti, Y., Gonnelli, M., Retelletti Brogi, S., Vestri, S., Santinelli, C., 2019. DOM dynamics in open waters of the Mediterranean Sea: New insights from optical properties. Deep Sea Research Part I: Oceanographic Research Papers 144, 95–114. https://doi.org/10.1016/j.dsr.2019.01.007
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
Halliday, S.J., Skeffington, R.A., Wade, A.J., Bowes, M.J., Gozzard, E., Newman, J.R., Loewenthal, M., Palmer-Felgate, E.J., Jarvie, H.P., 2015. High-frequency water quality monitoring in an urban catchment: hydrochemical dynamics, primary production and implications for the Water Framework Directive. Hydrological Processes 29, 3388–3407. https://doi.org/10.1002/hyp.10453
Hamilton, W.P., Kim, M., Thackston, E.L., 2005. Comparison of commercially available Escherichia coli enumeration tests: Implications for attaining water quality standards. Water Research 39, 4869–4878. https://doi.org/10.1016/j.watres.2005.02.006
Harmel, R.D., Hathaway, J.M., Wagner, K.L., Wolfe, J.E., Karthikeyan, R., Francesconi, W., McCarthy, D.T., 2016. Uncertainty in monitoring E. coli concentrations in streams and stormwater runoff. Journal of Hydrology 534, 524–533. https://doi.org/10.1016/j.jhydrol.2016.01.040
Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J., 2020. Statistical methods in water resources (USGS Numbered Series No. 4-A3), Statistical methods in water resources, Techniques and Methods. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/tm4A3
Herrig, I., Seis, W., Fischer, H., Regnery, J., Manz, W., Reifferscheid, G., Böer, S., 2019. Prediction of fecal indicator organism concentrations in rivers: the shifting role of environmental factors under varying flow conditions. Environmental Sciences Europe 31, 59. https://doi.org/10.1186/s12302-019-0250-9
Hong, Y., Li, C., Lemaire, B.J., Soulignac, F., Scarati Martins, J.R., Roguet, A., Lucas, F., Vinçon-Leite, B., 2019. An Integrated Approach for Assessing the Impact of Urban Stormwater Discharge on the Fecal Contamination in a Recreational Lake Near Paris, in: Mannina, G. (Ed.), New Trends in Urban Drainage Modelling. Springer International Publishing, Cham, pp. 334–338. https://doi.org/10.1007/978-3-319-99867-1_56
Huang, G., Falconer, R.A., Lin, B., 2017. Integrated hydro-bacterial modelling for predicting bathing water quality. Estuarine, Coastal and Shelf Science 188, 145–155. https://doi.org/10.1016/j.ecss.2017.01.018
Hudson, N., Baker, A., Reynolds, D., 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review. River Res. Applic. 23, 631–649. https://doi.org/10.1002/rra.1005
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
HydroPortail, E., 2022. Station hydrométrique - F664 0001 04 : La Marne à Gournay-sur-Marne - Pont - Fiche de synthèse - Données hydrologiques de synthèse | SCHAPI - HydroPortail [WWW Document]. HydroPortail. URL https://www.hydro.eaufrance.fr/stationhydro/F664000104/synthese (accessed 11.29.22).
IDEXX US, 2020. Quanti-Tray System [WWW Document]. URL https://www.idexx.com/en/water/water-products-services/quanti-tray-system/ (accessed 5.29.20).
Infoclimat, 2021. HistorIC - Episode pluies et inondations exceptionnel - Infoclimat [WWW Document]. URL https://www.infoclimat.fr/historic-details-evenement-2922-episode-pluies-et-inondations-exceptionnel.html (accessed 10.23.23).
Institut Paris Region, 2022. Occupation du sol 2021 en 11 postes de légende d'Île-de-France.
IPCC, Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P.W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W.W.L., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., Jotzo, F., Krug, T., Lasco, R., Lee, Y.-Y., Masson-Delmotte, V., Meinshausen, M., Mintenbeck, K.,

Mokssit, A., Otto, F.E.L., Pathak, M., Pirani, A., Poloczanska, E., Pörtner, H.-O., Revi, A., Roberts, D.C., Roy, J., Ruane, A.C., Skea, J., Shukla, P.R., Slade, R., Slangen, A., Sokona, Y., Sörensson, A.A., Tignor, M., Van Vuuren, D., Wei, Y.-M., Winkler, H., Zhai, P., Zommers, Z., Hourcade, J.-C., Johnson, F.X., Pachauri, S., Simpson, N.P., Singh, C., Thomas, A., Totin, E., Arias, P., Bustamante, M., Elgizouli, I., Flato, G., Howden, M., Méndez-Vallejo, C., Pereira, J.J., Pichs-Madruga, R., Rose, S.K., Saheb, Y., Sánchez Rodríguez, R., Ürge-Vorsatz, D., Xiao, C., Yassaa, N., Alegría, A., Armour, K., Bednar-Friedl, B., Blok, K., Cissé, G., Dentener, F., Eriksen, S., Fischer, E., Garner, G., Guivarch, C., Haasnoot, M., Hansen, G., Hauser, M., Hawkins, E., Hermans, T., Kopp, R., Leprince-Ringuet, N., Lewis, J., Ley, D., Ludden, C., Niamir, L., Nicholls, Z., Some, S., Szopa, S., Trewin, B., Van Der Wijst, K.-I., Winter, G., Witting, M., Birt, A., Ha, M., Romero, J., Kim, J., Haites, E.F., Jung, Y., Stavins, R., Birt, A., Ha, M., Orendain, D.J.A., Ignon, L., Park, S., Park, Y., Reisinger, A., Cammaramo, D., Fischlin, A., Fuglestvedt, J.S., Hansen, G., Ludden, C., Masson-Delmotte, V., Matthews, J.B.R., Mintenbeck, K., Pirani, A., Poloczanska, E., Leprince-Ringuet, N., Péan, C., 2023. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

ISO, 2012. ISO 9308-2:2012 - Water quality — Enumeration of Escherichia coli and coliform bacteria — Part 2: Most probable number method [WWW Document]. ISO. URL https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/22/52246.html (accessed 6.15.20).

ISO, 1998. ISO 9308-3:1998 - Water quality — Detection and enumeration of Escherichia coli and coliform bacteria — Part 3: Miniaturized method (Most Probable Number) for the detection and enumeration of E. coli in surface and waste water [WWW Document]. ISO. URL https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/02/08/20878.html (accessed 6.15.20).

Jacob, P., Henry, A., Meheut, G., Charni-Ben-Tabassi, N., Ingrand, V., Helmi, K., 2015. Health Risk Assessment Related to Waterborne Pathogens from the River to the Tap. IJERPH 12, 2967–2983. https://doi.org/10.3390/ijerph120302967

Khac, V.T., Hong, Y., Plec, D., Lemaire, B., Dubois, P., Saad, M., Vinçon-Leite, B., 2018. An Automatic Monitoring System for High-Frequency Measuring and Real-Time Management of Cyanobacterial Blooms in Urban Water Bodies. Processes 6, 11. https://doi.org/10.3390/pr6020011

King, J., Ahmadian, R., Falconer, R.A., 2021. Hydro-epidemiological modelling of bacterial transport and decay in nearshore coastal waters. Water Research 196, 117049. https://doi.org/10.1016/j.watres.2021.117049

King, J.A., 2019. Investigation and prediction of pollution in coastal and estuarine waters, using experimental and numerical methods (phd). Cardiff University.

Koschelnik, J., Vogl, W., Epp, M., Lackner, M., 2015. Rapid analysis of β-D-glucuronidase activity in water using fully automated technology. Presented at the WATER RESOURCES MANAGEMENT 2015, A Coruña, Spain, pp. 471–481. https://doi.org/10.2495/WRM150401

Kraus, T.E.C., O'Donnell, K., Downing, B.D., Burau, J.R., Bergamaschi, B.A., 2017. Using Paired In Situ High Frequency Nitrate Measurements to Better Understand Controls on Nitrate Concentrations and Estimate Nitrification Rates in a Wastewater-Impacted River. Water Resources Research 53, 8423–8442. https://doi.org/10.1002/2017WR020670

Krommyda, M., Theodoropoulos, T., Sdongos, E., Amditis, A., 2019. Integrated monitoring system for environmental and river data measurements, in: 2019 13th International Conference on Sensing Technology (ICST). Presented at the 2019 13th International

Conference on Sensing Technology (ICST), IEEE, Sydney, Australia, pp. 1–6. https://doi.org/10.1109/ICST46873.2019.9047725
Kubista, M., Sjöback, R., Eriksson, S., Albinsson, B., 1994. Experimental correction for the inner-filter effect in fluorescence spectra. Analyst 119, 417–419. https://doi.org/10.1039/AN9941900417
Lackner, M., Grabow, W., Stadler, P., 2017. Handbook of Online and Near-real-time Methods in Microbiology. CRC Press.
Lambert, T., Bouillon, S., Darchambeau, F., Massicotte, P., Borges, A.V., 2016. Shift in the chemical composition of dissolved organic matter in the Congo River network. Biogeosciences 13, 5405–5420. https://doi.org/10.5194/bg-13-5405-2016
Lambert, T., Bouillon, S., Darchambeau, F., Morana, C., Roland, F.A.E., Descy, J.-P., Borges, A.V., 2017. Effects of human land use on the terrestrial and aquatic sources of fluvial organic matter in a temperate river basin (The Meuse River, Belgium). Biogeochemistry 136, 191–211. https://doi.org/10.1007/s10533-017-0387-9
Lawaetz, A.J., Stedmon, C.A., 2009. Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water. Appl Spectrosc 63, 936–940. https://doi.org/10.1366/000370209788964548
Le Vu, B., Vinçon-Leite, B., Lemaire, B.J., Bensoussan, N., Calzas, M., Drezen, C., Deroubaix, J.F., Escoffier, N., Dégrés, Y., Freissinet, C., Groleau, A., Humbert, J.F., Paolini, G., Prévot, F., Quiblier, C., Rioust, E., Tassin, B., 2011. High-frequency monitoring of phytoplankton dynamics within the European water framework directive: application to metalimnetic cyanobacteria. Biogeochemistry 106, 229–242. https://doi.org/10.1007/s10533-010-9446-1
Lewis, E.L., 1982. The Practical Salinity Scale of 1978 and Its Antecedents. Mar. Geod. 5, 351–357.
Li, P., Chen, L., Zhang, W., Huang, Q., 2015. Spatiotemporal Distribution, Sources, and Photobleaching Imprint of Dissolved Organic Matter in the Yangtze Estuary and Its Adjacent Sea Using Fluorescence and Parallel Factor Analysis. PLOS ONE 10, e0130852. https://doi.org/10.1371/journal.pone.0130852
Lopez-Roldan, R., Tusell, P., Cortina, J.L., Courtois, S., Cortina, J.L., 2013. On-line bacteriological detection in water. TrAC Trends in Analytical Chemistry 44, 46–57. https://doi.org/10.1016/j.trac.2012.10.010
Mancini, J.L., 1978. Numerical estimates of coliform mortality rates under various conditions. Journal of the Water Pollution Control Federation 50, 2477–2484.
Marcé, R., Verdura, L., Leung, N., 2021. Dissolved organic matter spectroscopy reveals a hot spot of organic matter changes at the river–reservoir boundary. Aquat Sci 83, 67. https://doi.org/10.1007/s00027-021-00823-6
Markager, S., Vincent, W.F., 2000. Spectral light attenuation and the absorption of UV and blue light in natural waters. Limnology and Oceanography 45, 642–650. https://doi.org/10.4319/lo.2000.45.3.0642
Martin, N.A., Sala-Comorera, L., Gao, G., Corkery, A., Reynolds, L.J., Nolan, T.M., Whitty, M., O'Sullivan, J.J., Meijer, W.G., 2023. Inclusion of hydrodynamic properties of bathing waters is critical in selecting faecal indicators to assess public health impacts of faecal contamination. Water Research 242, 120137. https://doi.org/10.1016/j.watres.2023.120137
Maul, A., 1982. DEFINITION D'UNE STRATEGIE DE PRELEVEMENT POUR L'ANALYSE MICROBIOLOGIQUE DES EAUX DE SURFACE (PhD thesis). Université de Metz, Metz.
McCabe, K.M., Smith, E.M., Lang, S.Q., Osburn, C.L., Benitez-Nelson, C.R., 2021. Particulate and Dissolved Organic Matter in Stormwater Runoff Influences Oxygen Demand in Urbanized Headwater Catchments. Environ. Sci. Technol. 55, 952–961. https://doi.org/10.1021/acs.est.0c04502
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

Météo-France, 2023. Bilan annuel 2022 - 1ère partie - Synthèse, année mois par mois et année au fil des saisons [WWW Document]. meteofrance.fr. URL https://meteofrance.fr/actualite/publications/2022-les-bilans-climatiques (accessed 11.14.23).

Meyerhoff, J., Dehnhardt, A., Hartje, V., 2010. Take your swimsuit along: the value of improving urban bathing sites in the metropolitan area of Berlin. Journal of Environmental Planning and Management 53, 107–124. https://doi.org/10.1080/09640560903399863

Mladenov, N., Zheng, Y., Miller, M.P., Nemergut, D.R., Legg, T., Simone, B., Hageman, C., Rahman, M.M., Ahmed, K.M., McKnight, D.M., 2010. Dissolved Organic Matter Sources and Consequences for Iron and Arsenic Mobilization in Bangladesh Aquifers. Environ. Sci. Technol. 44, 123–128. https://doi.org/10.1021/es901472g

Moatar, F., Miquel, J., Poirel, A., 2001. A quality-control method for physical and chemical monitoring data. Application to dissolved oxygen levels in the river Loire (France). Journal of Hydrology 252, 25–36. https://doi.org/10.1016/S0022-1694(01)00439-5

Morris, D.P., Zagarese, H., Williamson, C.E., Balseiro, E.G., Hargreaves, B.R., Modenutti, B., Moeller, R., Queimalinos, C., 1995. The attenuation of solar UV radiation in lakes and the role of dissolved organic carbon. Limnology and Oceanography 40, 1381–1391. https://doi.org/10.4319/lo.1995.40.8.1381

Mourad, M., Bertrand-Krajewski, J.-L., 2002. A method for automatic validation of long time series of data in urban hydrology. Water Science and Technology 45, 263–270. https://doi.org/10.2166/wst.2002.0601

Murphy, K.R., Butler, K.D., Spencer, R.G.M., Stedmon, C.A., Boehme, J.R., Aiken, G.R., 2010. Measurement of Dissolved Organic Matter Fluorescence in Aquatic Environments: An Interlaboratory Comparison. Environ. Sci. Technol. 44, 9405–9412. https://doi.org/10.1021/es102362t

Murphy, K.R., Hambly, A., Singh, S., Henderson, R.K., Baker, A., Stuetz, R., Khan, S.J., 2011. Organic Matter Fluorescence in Municipal Water Recycling Schemes: Toward a Unified PARAFAC Model. Environ. Sci. Technol. 45, 2909–2916. https://doi.org/10.1021/es103015e

Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. Anal. Methods 5, 6557. 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

Nguyen, P.T., 2014. Study of the aquatic dissolved organic matter from the Seine River catchment (France) by optical spectroscopy combined to asymmetrical flow field-flow fractionation (phdthesis). Université de Bordeaux.

Nnane, D.E., Ebdon, J.E., Taylor, H.D., 2011. Integrated analysis of water quality parameters for cost-effective faecal pollution management in river catchments. Water Research 45, 2235–2246. https://doi.org/10.1016/j.watres.2011.01.018

Nowicki, S., Lapworth, D.J., Ward, J.S.T., Thomson, P., Charles, K., 2019. Tryptophan-like fluorescence as a measure of microbial contamination risk in groundwater. Science of The Total Environment 646, 782–791. https://doi.org/10.1016/j.scitotenv.2018.07.274

Offenbaume, K.L., Bertone, E., Stewart, R.A., 2020. Monitoring Approaches for Faecal Indicator Bacteria in Water: Visioning a Remote Real-Time Sensor for E. coli and Enterococci. Water 12, 2591. https://doi.org/10.3390/w12092591

Ohno, T., 2002. Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. Environ. Sci. Technol. 36, 742–746. https://doi.org/10.1021/es0155276

Ordulj, M., Jozić, S., Baranović, M., Krželj, M., 2022. The Effect of Precipitation on the Microbiological Quality of Bathing Water in Areas under Anthropogenic Impact. Water 14, 527. https://doi.org/10.3390/w14040527

Parker, C.A., Barnes, W.J., 1957. Some experiments with spectrofluorimeters and filter fluorimeters. Analyst 82, 606–618. https://doi.org/10.1039/AN9578200606

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

Paruch, A.M., Mæhlum, T., 2012. Specific features of Escherichia coli that distinguish it from coliform and thermotolerant coliform bacteria and define it as the most accurate indicator of faecal contamination in the environment. Ecological Indicators 23, 140–142. https://doi.org/10.1016/j.ecolind.2012.03.026

Piccolroaz, S., Toffolon, M., Majone, B., 2013. A simple lumped model to convert air temperature into surface water temperature in lakes. Hydrol. Earth Syst. Sci. 17, 3323–3338. https://doi.org/10.5194/hess-17-3323-2013

Polubesova, T., Sherman-Nakache, M., Chefetz, B., 2007. Binding of Pyrene to Hydrophobic Fractions of Dissolved Organic Matter: Effect of Polyvalent Metal Complexation. Environ. Sci. Technol. 41, 5389–5394. https://doi.org/10.1021/es070722r

Portail assainissement collectif, 2021. Stations de traitement des eaux usées [WWW Document]. Portail assainissement collectif. URL https://www.assainissement.developpement-durable.gouv.fr/PortailAC/data (accessed 3.14.23).

Poulin, M., Servais, P., Mouchel, J.-M., Thérial, C., Lesage, L., Rocher, V., Goncalves, A., Masnada, S., Lucas, F., Flipo, N., 2013. Modélisation de la contamination fécale en Seine : impact des rejets de temps de pluie (report). PIREN-Seine.

Prats, J., Garcia-Armisen, T., Larrea, J., Servais, P., 2008. Comparison of culture-based methods to enumerate Escherichia coli in tropical and temperate freshwaters. Letters in Applied Microbiology 46, 243–248. https://doi.org/10.1111/j.1472-765X.2007.02292.x

Pronk, M., Goldscheider, N., Zopfi, J., 2006. Dynamics and interaction of organic carbon, turbidity and bacteria in a karst aquifer system. Hydrogeol J 14, 473–484. https://doi.org/10.1007/s10040-005-0454-5

Pruss, A., 1998. Review of epidemiological studies on health effects from exposure to recreational water. International Journal of Epidemiology 27, 1–9. https://doi.org/10.1093/ije/27.1.1

Ren, W., Wu, X., Ge, X., Lin, G., Zhou, M., Long, Z., Yu, X., Tian, W., 2021. Characteristics of dissolved organic matter in lakes with different eutrophic levels in southeastern Hubei Province, China. J. Ocean. Limnol. 39, 1256–1276. https://doi.org/10.1007/s00343-020-0102-x

Reynolds, D.M., 2003. Rapid and direct determination of tryptophan in water using synchronous fluorescence spectroscopy. Water Research 37, 3055–3060. https://doi.org/10.1016/S0043-1354(03)00153-2

Romero González-Quijano, C., Herrero Ortega, S., Casper, P., Gessner, M., Singer, G., 2022. Dissolved organic matter signatures in urban surface waters: spatio-temporal patterns and drivers (preprint). Biogeochemistry: Limnology. https://doi.org/10.5194/bg-2021-340

Ryan, K.A., Palacios, L.C., Encina, F., Graeber, D., Osorio, S., Stubbins, A., Woelfl, S., Nimptsch, J., 2022. Assessing inputs of aquaculture-derived nutrients to streams using dissolved organic matter fluorescence. Science of The Total Environment 807, 150785. https://doi.org/10.1016/j.scitotenv.2021.150785

SAFEGE, 2017a. Elaboration d'un profil de baignade pour une baignade publique sur le bassin de la Villette à Paris. Rapport de phase 1: Etats des lieux.

SAFEGE, 2017b. Elaboration d'un profil de baignade pour une baignade publique sur le bassin de la Villette à Paris.

Seis, W., Zamzow, M., Caradot, N., Rouault, P., 2018. On the implementation of reliable early warning systems at European bathing waters using multivariate Bayesian regression modelling. Water Research 143, 301–312. https://doi.org/10.1016/j.watres.2018.06.057

Selméus, L., 2018. Dynamic modelling of bathing water quality with biodegradation of Escherichia coli in TELEMAC-3D.

Seo, M., Lee, H., Kim, Y., 2019. Relationship between Coliform Bacteria and Water Quality Factors at Weir Stations in the Nakdong River, South Korea. Water 11, 1171. https://doi.org/10.3390/w11061171

Servais, P., Billen, G., Goncalves, A., Garcia-Armisen, T., 2007. Modelling microbiological water quality in the Seine river drainage network: past, present and future situations. Hydrology and Earth System Sciences 11, 1581–1592.

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

Skeffington, R.A., Halliday, S.J., Wade, A.J., Bowes, M.J., Loewenthal, M., 2015. Using high-frequency water quality data to assess sampling strategies for the EU Water Framework Directive. Hydrol. Earth Syst. Sci. 19, 2491–2504. https://doi.org/10.5194/hess-19-2491-2015

Sokolova, E., Astrom, J., Pettersson, T.J.R., Bergstedt, O., Hermansson, M., 2012. Estimation of pathogen concentrations in a drinking water source using hydrodynamic modelling and microbial source tracking. J. Water Health 10, 358–370. https://doi.org/10.2166/wh.2012.183

Sorensen, J.P.R., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G., Ward, J.S.T., 2018. Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values. Science of The Total Environment 622–623, 1250–1257. https://doi.org/10.1016/j.scitotenv.2017.11.162

Sorensen, J.P.R., Lapworth, D.J., Marchant, B.P., Nkhuwa, D.C.W., Pedley, S., Stuart, M.E., Bell, R.A., Chirwa, M., Kabika, J., Liemisa, M., Chibesa, M., 2015. In-situ tryptophan-like fluorescence: A real-time indicator of faecal contamination in drinking water supplies. Water Research 81, 38–46. https://doi.org/10.1016/j.watres.2015.05.035

Sorensen, J.P.R., Nayebare, J., Carr, A.F., Lyness, R., Campos, L.C., Ciric, L., Goodall, T., Kulabako, R., Curran, C.M.R., MacDonald, A.M., Owor, M., Read, D.S., Taylor, R.G., 2021. In-situ fluorescence spectroscopy is a more rapid and resilient indicator of faecal contamination risk in drinking water than faecal indicator organisms. Water Research 206, 117734. https://doi.org/10.1016/j.watres.2021.117734

Sorensen, J.P.R., Sadhu, A., Sampath, G., Sugden, S., Dutta Gupta, S., Lapworth, D.J., Marchant, B.P., Pedley, S., 2016. Are sanitation interventions a threat to drinking water supplies in rural India? An application of tryptophan-like fluorescence. Water Research 88, 923–932. https://doi.org/10.1016/j.watres.2015.11.006

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., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Marine Chemistry 82, 239–254. https://doi.org/10.1016/S0304-4203(03)00072-0

Tatari, K., Corfitzen, C.B., Albrechtsen, J., Christensen, S.C.B., 2016. Sensors for microbial drinking water quality. DTU 46.

Tedetti, M., Longhitano, R., Garcia, N., Guigue, C., Ferretto, N., Goutx, M., 2012. Fluorescence properties of dissolved organic matter in coastal Mediterranean waters influenced by a municipal sewage effluent (Bay of Marseilles, France). Environ. Chem. 9, 438. https://doi.org/10.1071/EN12081

Tiberti, R., Caroni, R., Cannata, M., Lami, A., Manca, D., Strigaro, D., Rogora, M., 2021. Automated high frequency monitoring of Lake Maggiore through in situ sensors: system design, field test and data quality control. J Limnol 80. https://doi.org/10.4081/jlimnol.2021.2011

Tiwari, A., Niemelä, S.I., Vepsäläinen, A., Rapala, J., Kalso, S., Pitkänen, T., 2016. Comparison of Colilert-18 with miniaturised most probable number method for monitoring of Escherichia coli in bathing water. Journal of Water and Health 14, 121–131. https://doi.org/10.2166/wh.2015.071

Tucker, L.R., 1951. A method for synthesis of factor analysis studies (Personnel Research Section Report no. 984). Department of the Army, Washington D.C.

US EPA, O., 2013a. Indicators: Enterococci [WWW Document]. US EPA. URL https://www.epa.gov/national-aquatic-resource-surveys/indicators-enterococci (accessed 5.16.21).

US EPA, O., 2013b. Indicators: Conductivity [WWW Document]. US EPA. URL https://www.epa.gov/national-aquatic-resource-surveys/indicators-conductivity (accessed 5.16.21).

USEPA, 1986. Ambient Water Quality Criteria for Bacteria - 1986 (No. EPA 440/5-84-002). Office of Water, Washington DC.

Van, L.-A., Nguyen, K.-D., Le Marrec, F., Jairy, A., 2022. Development of a Tool for Modeling the Fecal Contamination in Rivers with Turbulent Flows—Application to the Seine et Marne Rivers (Parisian Region, France). Water 14, 1191. https://doi.org/10.3390/w14081191

Vanhalst, M., 2018. GESTION DES RISQUES SANITAIRES LIES A LA BAIGNADE URBAINE Problématique : Développements et améliorations des mesures de gestion active pour la prévention des risques sanitaires relatifs à la baignade urbaine.

Velapoldi, R.A., Mielenz, K.D., 1980. A Fluorescence Standard Reference Material: Quinine Sulfate Dihydrate. Department of Commerce, National Bureau of Standards.

Vinçon-Leite, B., Angelotti, N., Arthur Guillot-Le Goff, Carmigniani, R., Lacerda, R., Dubois, P., Saad, M., Lucas, F., Naloufi, M., Therial, C., 2023. 3D hydrodynamic modelling for the assessment of water quality in urban waterbodies – application to Bassin de La Villette (Paris, France). Lyon, France.

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

Ward, J.S.T., Lapworth, D.J., Read, D.S., Pedley, S., Banda, S.T., Monjerezi, M., Gwengweya, G., MacDonald, A.M., 2021. Tryptophan-like fluorescence as a high-level screening tool for detecting microbial contamination in drinking water. Science of The Total Environment 750, 141284. https://doi.org/10.1016/j.scitotenv.2020.141284

Ward, J.S.T., Lapworth, D.J., Read, D.S., Pedley, S., Banda, S.T., Monjerezi, M., Gwengweya, G., MacDonald, A.M., 2020. Large-scale survey of seasonal drinking water quality in Malawi using in situ tryptophan-like fluorescence and conventional water quality indicators. Science of The Total Environment 744, 140674. https://doi.org/10.1016/j.scitotenv.2020.140674

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

Wetzel, R.G., 1992. Gradient-dominated ecosystems: sources and regulatory functions of dissolved organic matter in freshwater ecosystems. Hydrobiologia 229, 181–198. https://doi.org/10.1007/BF00007000

WHO, 2018. WHO recommendations on scientific, analytical and epidemiological developments relevant to the parameters for bathing water quality in the Bathing Water Directive (2006/7/EC).

WHO, 1996. Guidelines for drinking-water quality, volume 2: Health criteria and other supporting information. International Programme on Chemical Safety 61, 274. https://doi.org/10.1016/0048-9697(87)90388-3

Wilkes, G., Edge, T.A., Gannon, V.P.J., Jokinen, C., Lyautey, E., Neumann, N.F., Ruecker, N., Scott, A., Sunohara, M., Topp, E., Lapen, D.R., 2011. Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. Water Research 45, 5807–5825. https://doi.org/10.1016/j.watres.2011.06.021

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

Yan, C., Sheng, Y., Ju, M., Ding, C., Li, Q., Luo, Z., Ding, M., Nie, M., 2020. Relationship between the characterization of natural colloids and metal elements in surface waters. Environ Sci Pollut Res 27, 31872–31883. https://doi.org/10.1007/s11356-020-09500-x

Yang, X., Yu, X., Cheng, J., Zheng, R., Wang, K., Dai, Y., Tong, N., Chow, A.T., 2018. Impacts of land-use on surface waters at the watershed scale in southeastern China: Insight from fluorescence excitation-emission matrix and PARAFAC. Science of The Total Environment 627, 647–657. https://doi.org/10.1016/j.scitotenv.2018.01.279

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

Annexes

Table of Annexes

I. Complementary information of the bioreactor of ALERT system

The ALERT system instrument is a cylindrical body of 25 cm of diameter and 40 cm of length, weighing 9.3 kg (Guérin-Rechdaoui et al., 2017). It has seven individual bioreactors, each one with a glass sampling vial with an effective volume of 25 mL pre-filled with 1 mL of liquid bioreagent (Figure A.1).

The vial is encased by an aluminium sleeve and a foam thermal insulation. It contains a flexible heater temperature sensor and an optical sensor ring. This ring contains three LEDs for measuring the fluorescence (385 nm excitation), the absorbance (at 430 nm) and the turbidity (at 610 nm). It also contains a photodiode coupled to a low-pass optical filter that blocks the UV excitation light, a primary signal amplification, and an incubator temperature control electronics with the main control circuitry, the microprocessor and the memory, the three of them located on separated electronic boards (Angelescu et al., 2019).

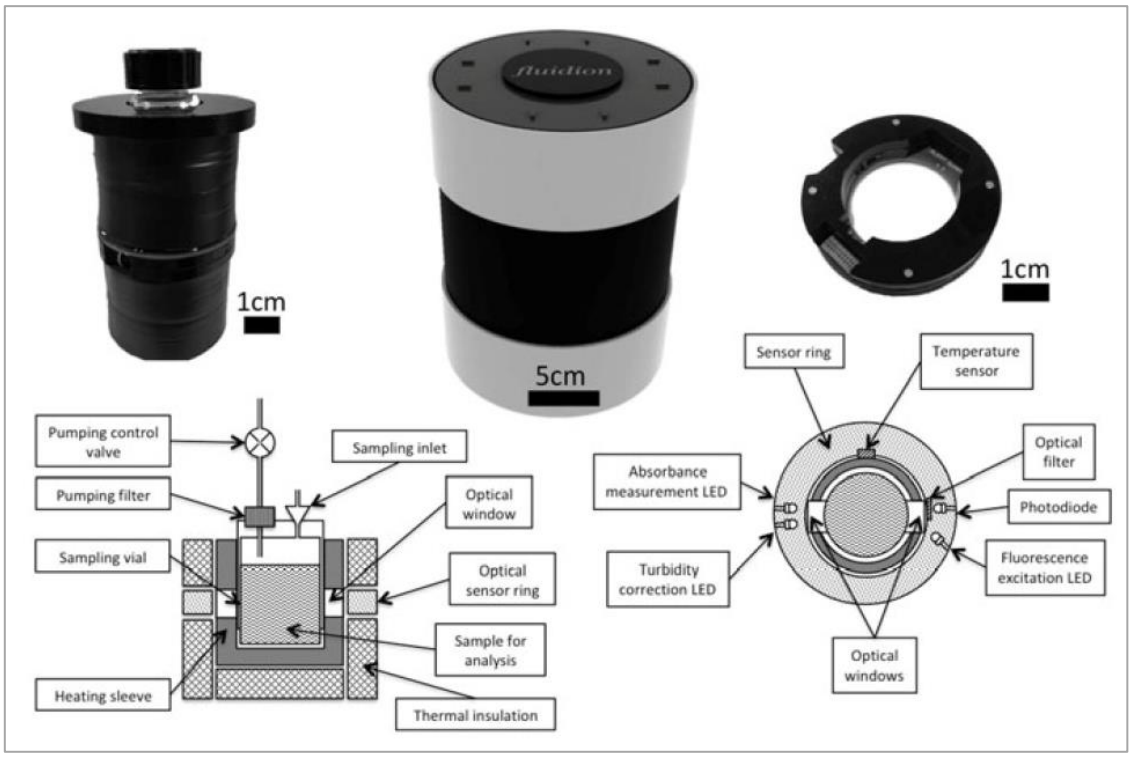


Figure A.1. ALERT system and bioreactor scheme (Angelescu et al., 2019).

- **Measure principle**

The bioreagent is composed by a mixture of growth medium, 4-methylumbelliferyl-β-D-glucuronide (MUG) and ortho-nitrophenyl-β-galactoside (ONPG), being MUG a standard substrate specific to the enzymes of *E. coli* and ONPG to Total Coliforms.

The β-glucuronidase is an *E. coli* enzyme that hydrolyzes the MUG compound into a fluorescent one known as 4-methylumbelliferyl (MUF). The ONPG is metabolised by all types of coliforms (Total Coliforms) and is transformed into ortho-nitrophenol (OP), which has a yellow colour.

When the sampling command is initiated, the new sample is drawn into the next available bioreactor and, in the vial, mixed with the bioreagent and incubated. The incubation is at 37 to 38 °C, the ideal temperature for the growth of *E. coli*. It is a highly selective method to

their cultivation because other organisms do not grow so fast. During this selective culture step, the bacterial metabolism progressively transforms MUG into MUF, generating fluorescence, with the emission peaking around 460 nm.

The appearance of these compounds is monitored over time by spectrophotometric and fluorescence measurements at periodic intervals, typically from 2 to 5 min. The fluorescence signal suffers a slight decrease immediately after the sampling, having a point of inflexion as a valley just before the appearance of fluorescence (Figure A.2). These slightly interferences with fluorescence measurement are mainly given because of the development of turbidity from bacterial growth and of yellow colour from OP. These inferences, however, do not influence the detection of the sharp fluorescence increase.

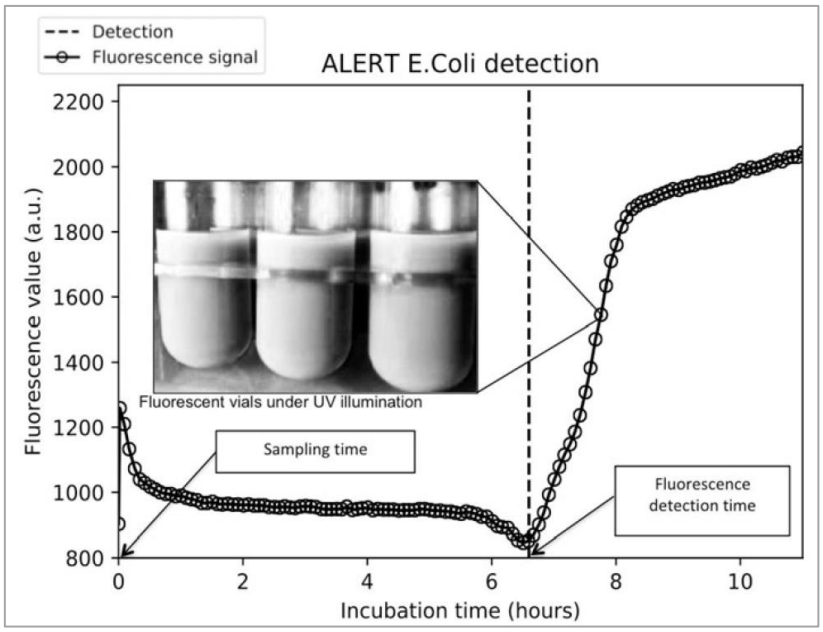


Figure A.2. ALERT system *E. coli* detection by fluorescence signal (Angelescu et al., 2019)

The appearance time of fluorescence is correlated with concentrations of *E. coli* in Colony-forming Units per 100 ml (CFU 100 mL^{-1}) and it ranges from **2 h to 12 h**. After calibration, for a concentration of 1000 MPN 100mL^{-1}, it takes around 9 h to have the results, as shown in Figure A.3 (Guérin-Rechdaoui et al., 2017).

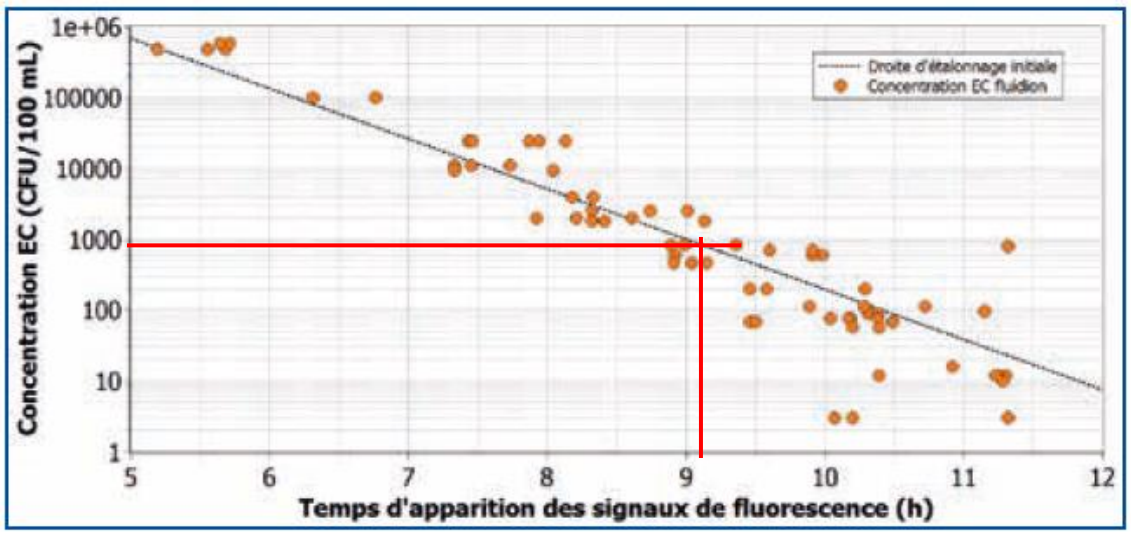


Figure A.3. Initial curve calibration for the Seine river (Guérin-Rechdaoui et al., 2017).

II. Physical processes in water bodies: thermal stratification

Thermal stratification consists in the formation of horizontal layers of water with different densities. The water density is a key variable of the stability of stratification, because of the currents that density differences can drive on the water bodies. The most common mechanism leading to stratification is thermal, due to the density variation with temperature (Boehrer and Schultze, 2008).

Besides temperature, other factors that contribute to density differences in water are dissolved substances, suspended materials and pressure (Boehrer and Schultze, 2008). For a shallow freshwater water body, the temperature affects density the most.

The variation of density with temperature is not linear. The density increases from 0 °C until its peak at 3.98 °C, and then decreases. (Greenwood and Earnshaw, 1997). Equation A1 is used for density (ρ) calculation for water bodies of low salinity (< 0.6 psu), for a temperature range from 0 to 30 °C and pression from 0 to 180 bars (Chen and Millero, 1986):

$$\rho(g.cm^{-3}) = 9.998395 \cdot 10^{-1} + 6.7914 \cdot 10^{-5}\, T - 9.0894 \cdot 10^{-6}\, T^2 + 1{,}0171 \cdot 10^{-7}\, T^3 - 1.2846 \cdot 10^{-9}\, T^4 + 1.1592 \cdot 10^{-11}\, T^5 - 5.0125 \cdot 10^{-14}\, T^6 + (8.181 \cdot 10^{-4} - 3.85 \cdot 10^{-6}\, T + 4.96 \cdot 10^{-8}\, T^2)S \quad \text{A1}$$

The density for pure water (S = 0) at sea level using Chen and Millero (1986) formula is illustrated on Figure A.4, indicating the maximum density at 4 °C.

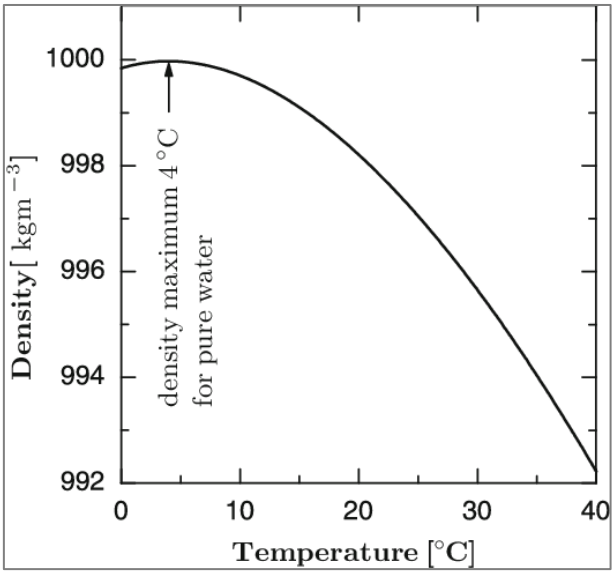


Figure A.4. Density variation of pure water as a function of the temperature, with S = 0 and p = 1 atm (Hutter et al., 2011).

The water temperature is mainly affected by the thermal contact of the surface with the atmosphere. Boehrer and Schultze (2008) list some processes that contribute to heat transfer through the surface. They include "solar radiation, long-wave radiation of atmosphere and surface waters, sensible heat exchange, and heat flux connected with evaporation and precipitation". Other factors that influence the thermal stratification are wind energy, currents, and morphometry (Wetzel, 1983).

Thus, the daily and seasonal variation of incoming solar radiation and wind affects thermal stratification. In addition, the measurements of heat transfer require the measurement of air temperature, air humidity (to estimate the evaporation from the surface) and air pressure (Boehrer and Schultze, 2008).

The surface water layer is exposed to the major sources of heat: solar radiation, heat-loss by longwave radiation and thermal contact with the atmosphere. The layers of a deeper level of the water body are shielded from those sources, which can lead to thermal stratification.

In a stratified water body, the surface layer is called epilimnion. The layer beneath, with higher density, is the hypolimnion. The temperature gradient in the contact zone between both layers is the thermocline. When the temperature of the surface layer falls enough so both layers are mixed, the homogenized one-layer is called mixolimnion (Boehrer and Schultze, 2008).

III. Governing equations of hydrodynamic modelling

This section is based on the theoretical aspects of the book *"Hydrodynamics of free surface flows. Modelling with the finite element method"*, from J-M. Hervouet, editor John Wiley & Sons, Ltd, 2007 (EDF R&D and Telemac3D consortium, 2018).

TELEMAC-3D solves the three-dimensional Navier-Stokes equations, a system of partial differential equations that describes the motion of a viscous fluid. This system expresses the conservation of mass and momentum for a control volume in a Newtonian fluid, with the pressure in terms of depth.

The considered hypothesis includes a free surface changing in time. Water is considered an incompressible fluid, with negligible density variation in mass conservation. Pressure at a given depth is the sum of the air pressure plus the weight of the overlying water (hydrostatic pressure hypothesis). Finally, the density variations are only considered as buoyant forces (Boussinesq approximation).

The equations A2 to A5 are solved by TELEMAC-3D with the hydrostatic pressure hypothesis.

$$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} + \frac{\partial W}{\partial z} = 0 \qquad \text{A2}$$

$$\frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} + V\frac{\partial U}{\partial y} + W\frac{\partial U}{\partial z} = -g\frac{\partial Z_S}{\partial x} + \upsilon\Delta(U) + F_x \qquad \text{A3}$$

$$\frac{\partial V}{\partial t} + U\frac{\partial V}{\partial x} + V\frac{\partial V}{\partial y} + W\frac{\partial V}{\partial z} = -g\frac{\partial Z_S}{\partial y} + \upsilon\Delta(V) + F_y \qquad \text{A4}$$

$$p = p_{atm} + \rho_0 g(Z_S - z) + \rho_0 g\int_z^{Z_S}\frac{\Delta\rho}{\rho_0}dz' \qquad \text{A5}$$

Where:

- U, V and W are the 3D components of velocity ($m.s^{-1}$);
- x, y and z are the 3D space components (m);
- t is time (s);
- g is the acceleration due to gravity ($m.s^{-2}$);
- Z_S is the free surface elevation (m);
- υ is the cinematic viscosity ($m^2.s^{-1}$);
- F_x and F_y are the source terms, denoting the wind, the Coriolis force and the bottom friction ($m.s^{-2}$);
- p is total pressure;
- p_{atm} is atmospheric pressure;
- ρ_0 is the reference density;
- $\Delta\rho$ is the variation of density around the reference density

The tracer transport is computed with Equation A6.

$$\frac{\partial T}{\partial t} + U\frac{\partial T}{\partial x} + V\frac{\partial T}{\partial y} + W\frac{\partial T}{\partial z} = \frac{\partial}{\partial x}\left(\upsilon_T\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(\upsilon_T\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(\upsilon_T\frac{\partial T}{\partial z}\right) + Q \qquad \text{A6}$$

With:

- T is the tracer (passive or active);
- υ_T is the tracer diffusion coefficients ($m^2.s^{-1}$);
- Q is tracer source or sink.

IV. Monitoring system

IV-a. Conversion between salinity and conductivity (Practical Salinity Scale of 1978)

The Practical Salinity Scale (PSS) is based upon electrical conductivity measurements. The equations and constants from PSS of 1978 (Equations A7 to A11) compute salinity from *in-situ* conductivity measurements. Similarly, they can be used to obtain conductivity from salinity data. The equations are valid for water temperatures between -2°C and 35°C.

$$R = \frac{C(S,T,P)}{C(35,15,0)} \quad \text{A7}$$

$$R_T = \frac{R}{r_T(1+\alpha)} \quad \text{A8}$$

$$\alpha = \frac{A_1 P + A_2 P^2 + A_3 P^3}{1 + B_1 T + B_2 T^2 + B_3 R + B_4 T R} \quad \text{A9}$$

$$r_T = c_0 + c_1 T + c_2 T^2 + c_3 T^3 + c_4 T^5 \quad \text{A10}$$

$$S(‰) = a_0 + a_1 {R_T}^{\frac{1}{2}} + a_2 R_T + a_3 {R_T}^{\frac{3}{2}} + a_4 {R_T}^2 + a_5 {R_T}^{5/2} + \frac{(T-15)}{1+k(T-15)}\{b_0 + b_1 {R_T}^{\frac{1}{2}} + b_2 R_T + b_3 {R_T}^{\frac{3}{2}} + b_4 {R_T}^2 + b_5 {R_T}^{5/2}\} \quad \text{A11}$$

Where:

- S is the salinity, in ‰.
- T is the water temperature, in °C.
- P is the pressure, in decibars.
- R is the ratio between *in situ* conductivity C(S,T,P) and the conductivity of reference C(35,15,0).
- C(S,T,P) is the *in-situ* conductivity, in mS.cm^{-1}.
- C(35,15,0) = 42.914 mS.cm^{-1}, conductivity at 35 ppt, 15 °C and 0 pressure.

With the constants:

- $A_1 = 2.070 \cdot 10^{-5}$
- $A_2 = -6.370 \cdot 10^{-10}$
- $A_3 = 3.989 \cdot 10^{-15}$
- $B_1 = 3.426 \cdot 10^{-2}$
- $B_2 = 4.464 \cdot 10^{-4}$
- $B_3 = 4.215 \cdot 10^{-1}$
- $B_4 = -3.107 \cdot 10^{-3}$
- $c_0 = 6.766097 \cdot 10^{-1}$
- $c_1 = 2.00564 \cdot 10^{-2}$
- $c_2 = 1.104259 \cdot 10^{-4}$
- $c_3 = -6.9698 \cdot 10^{-7}$
- $c_4 = 1.0031 \cdot 10^{-9}$
- $a_0 = 0.0080$
- $a_1 = -0.1692$
- $a_2 = 25.3851$
- $a_3 = 14.0941$
- $a_4 = -7.0261$
- $a_5 = 2.7081$
- $b_0 = 0.0005$
- $b_1 = -0.0056$
- $b_2 = -0.0066$
- $b_3 = -0.0375$
- $b_4 = 0.0636$
- $b_5 = -0.0144$
- $k = 0.0162$

IV-b. Validation of field data from basic monitoring system

From 07/08/2020 to 10/062021, the converted conductivity and the temperature measurements from the CTD (nke instruments) sensor installed at sub-surface at point A were compared with the punctual mean values at depths between 0.15 m and 0.35 m (Figure A.5).

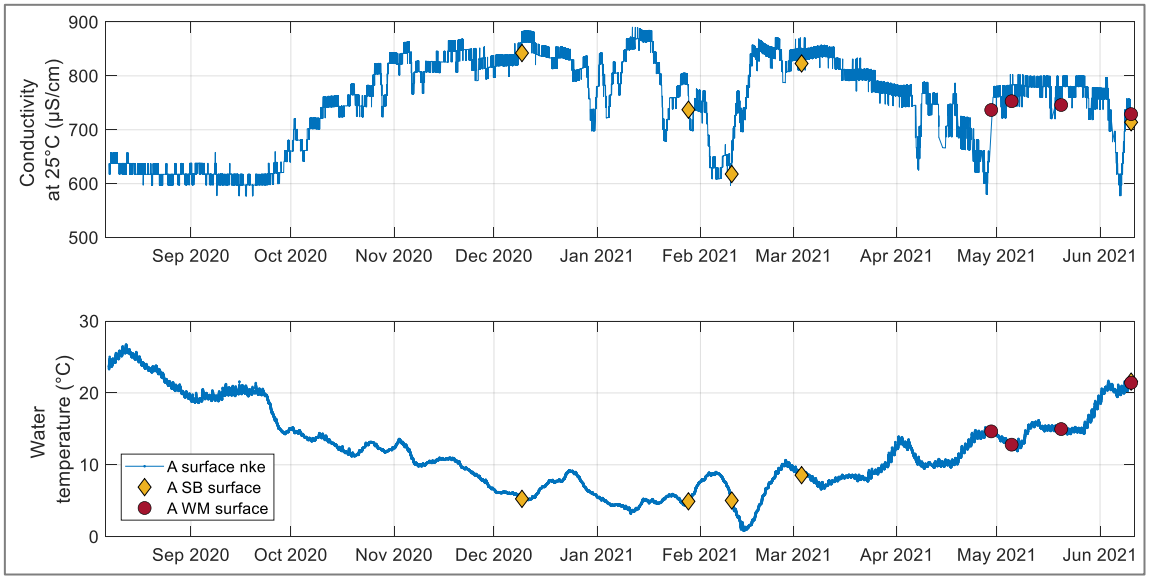


Figure A.5. Conductivity and water temperature measurement from nke sensor (continuous), and from references probes SeaBird (SB) and WiMo (WM), at point B, from 07/08/2021 to 10/06/2021.

A comparison between the reference probes, SeaBird and WiMo, was done. At point A, the conductivity and temperature measurements from SeaBird and the nke continuous measurements presented a good relationship, with RMSE = 0.06 °C for water temperature, RMSE = 4 µS.cm^{-1} and R^2=1.00 for both conductivity and water temperature measurements (Figure A.6 and Table A.1).

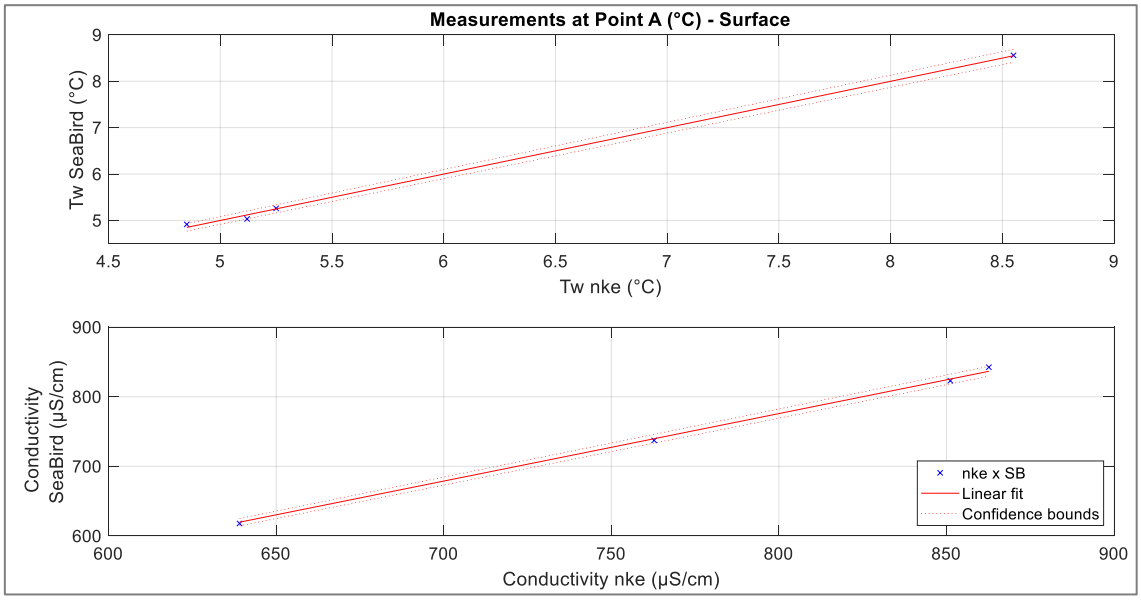


Figure A.6. Water temperature and conductivity: nke and SeaBird measurements at point A.

Table A.1. nke and SeaBird at point A

Point	Equipment	Variable	N. of points	Equation	R^2	RMSE
A surface	Nke and SeaBird	Water temperature (°C)	4	$Tws_{nke} = 1.00 * Tws_{SB}$	1.00	0.06
A surface	Nke and SeaBird	Conductivity ($\mu S.cm^{-1}$)	4	$Cs_{nke} = 0.97 * Cs_{SB}$	1.00	4.06

At point B, the temperature measurements of the sensors were compared with the mean at the depth of the sensor ± 0.10 m for each of three defined depths (Figure A.7).

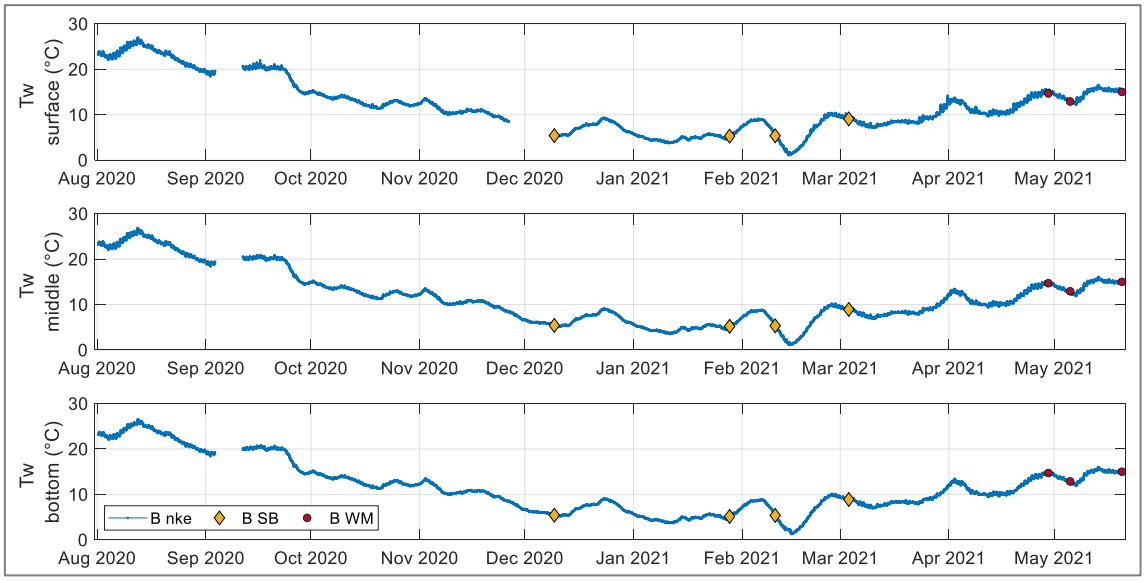


Figure A.7. Water temperature measurement from nke sensor (continuously), and from references probes SeaBird (SB) and WiMo (WM), at point B, from 01/08/2021 to 05/05/2021.

The relationship between the water temperature measurements from nke sensors and SeaBird probe at the middle and bottom depths were good, with RMSE < 0.1 °C and R^2=1.00 (Figure A.8 and Table A.2). At surface, the relationship was not good (RMSE = 1.46 °C, R^2 = 0.39).

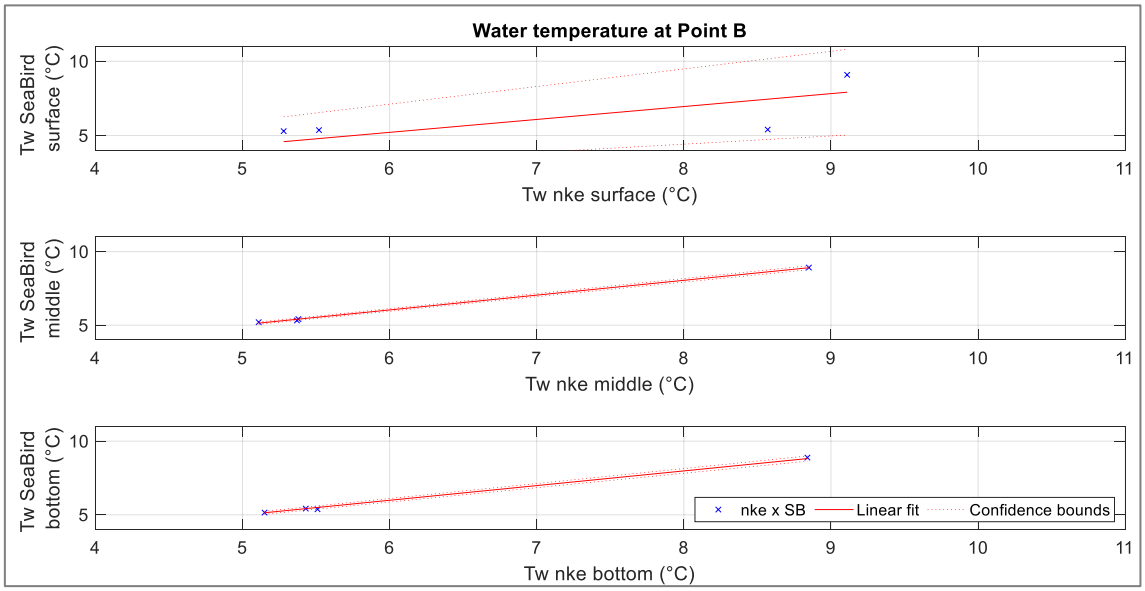


Figure A.8. Water temperature between nke sensors and SeaBird probe measurements at point B.

Table A.2. Relationship between nke and SeaBird measurements at point B.

Point	Equipment	Variable	N. of points	Equation	R^2	RMSE
B surface	Nke and SeaBird	Water temperature (°C)	4	$Tws_{nke} = 0.87 * Tws_{SB}$	0.39	1.46
B middle	Nke and SeaBird	Water temperature (°C)	4	$Twm_{nke} = 1.01 * Twm_{SB}$	1.00	0.06
B bottom	Nke and SeaBird	Water temperature (°C)	4	$Twb_{nke} = 1.00 * Twb_{SB}$	1.00	0.08

IV-c. Follow-up of field campaigns

Written on the 06/08/2020
revised on the 06/11/2023

Site: La Villette basin, Paris 19e

Coordinates:	Point A (cage amont)	Point B0 (baignade)	Point B (marin d'eau douce - ancient point C)
Latitude (N)	48.891732	48.885805	48.885841
Longitude (E)	2.3855	2.373751	2.373831
Altitude (m)	51.9	51.68	51.39

Source: https://www.geoportail.gouv.fr/carte

Pression: Paris (75) Paris Saint-Germain-des-Près (75)
https://www.meteociel.fr/temps-reel/obs_villes.php?code2=2&jour2=9&mois2=11&annee2=2020

	Name	Initials	Name	Initials	Period in LEESU
Participants:	Natalia Angelotti	NA	Francesco Piccioni	FP	2019-2021
	Philippe Dubois	PhD	Guilherme Calabro	GC	2020-2022
	Mohamed	MS	Amanda Tolentino	AT	Jun-Sep 2021
	Brigitte Vinçon-Leite	BVL	Maria Julia	MJ	Sep-Dec 2021
			Arthur Guillot-le goff	AG	Fev-Juil 2022

Suivi des interventions à la Villette

Date	Jour semaine	Heure début	Heure fin	Participants	*Point C (Pont de Crimée)* Mesures effectuées	*Point B (Marin d'eau douce)* Mesures effectuées	Secchi D. (m)	Prof (m)	*Point A (amont/cage)* Mesures effectuées	Secchi D. (m)	Prof (m)	Commentaires
06/07/2020	lun	10:00	11:00	PhD, FP, NA					-	-	-	Installation de la ligne de capteurs au Point B0 avec 3 capteurs de température: surface (nke SP2T 32007), milieu (nke SP2T 32009 - avant modif du 04/09: 32007), fond (nke SP2T 28076). Pas de temps = 10 min.
03/08/2020	lun	19:00	19:30	NA					-	-	-	Récupération des données de température au Point B0
06/08/2020	jeu	11:45	12:45	PhD, NA, M. Service des Canaux					Temperature, salinity	-	2.05	Installation du capteur de conductivité (nke 35013, num. radio 15172) au Point A. Profondeur=0.25m. Pas de temps = 5 min.
24/08/2020	lun	20:00	20:30	NA					Temperature, salinity	-	-	Nettoyage des capteurs. Tout semblait intouché.
28/08/2020	ven	15:00	16:00	BVL, NA, MS					Temp, salinity, Nitrate, Phosphate, Transparence	0.8	1.95	Waterwatch réalisé au niveau de la cage fluidion. Vie aquatique: un pêcheur a peché un poisson de taille moyenne (un palme) à coté de la cage fluidion. Il y a eu une orage entre 15h40 et 15h50.
03/09/2020	jeu	17:30	18:30	FP, NA					-	-	-	Récupération de la ligne de capteurs de la baignade (fin Paris Plage, le ponton sera enlévé). Le 04/09: MAJ du n. de série du capteur du milieu (32009 au lieu de 32007)
11/09/2020	ven	17:30	18:30	NA, MS			0.9	2.75				Installation de la ligne de capteurs au Point B avec 3 capteurs de température: surface (nke SP2T 32007), milieu (nke SP2T 32009), fond (nke SP2T 28076). Pas de temps = 10 min.
21/09/2020	lun	09:00	10:30	NA, FP		Temperature			Temp., Salinity			
12/10/2020	lun	09:00	10:30	NA, Samuel		Temp., FluoroProbe, échantillon (fluospectro, spectrophoto)	1.2	2.75	Temp., salinity, FluoroProbe, échantillon (fluospectro, spectrophoto)	1.25	2.00	Forte présence de macrophytes. Les échantillons pour les spécto ont été prélevé avec un sceau plongé à mi profondeur.
26/10/2020	lun	10:30	11:20	NA		Temp., échantillon (fluospectro, spectrophoto)			Temp., salinity, échantillon (fluospectro, spectrophoto)			Pluie intense à la veille. Macrophytes entremelées dans les capteurs de surface et de milieu au Point B. César (Marin d'Eau Douce) a fourni son numéro pour confirmer les horaires d'ouverture à la veille des interventions (06.51.56.35.14).
09/11/2020	lun	09:40	11:20	NA, Samuel		Temp., FluoroProbe, échantillon (fluospectro, spectrophoto)	1.90	2.80	Temp., salinity, FluoroProbe, échantillon (fluospectro, spectrophoto)	1.2	2.00	FreshWater Watch: Point A: NO3=5mg/L, PO4=0. Point C: NO3 entre 5 et 10 mg/L, PO4=0.
26/11/2020	jeu	09:50	12:00	NA, Samuel		Temp., FluoroProbe, échantillon (fluospectro, spectrophoto)	2.2	2.8	Temp., salinity, FluoroProbe, échantillon (fluospectro, spectrophoto)	2	2	FreshWater Watch: Point A: NO3=5-10 mg/L, PO4=0, Turb<14NTU. Petite couche de graisse au point A causé par des dechets de nourriture en amont. Mesures au point B ont été faites vers le centre du canal car un bateau genait l'accès.
09/12/2020	mer	09:30	11:45	NA, MS, GC, Samuel		ADCP + Temp., FluoroProbe, SeaBird, échantillon (fluospectro, spectrophoto)	1.65	2.65	Temp., salinity, FluoroProbe, SeaBird, échantillon (fluospectro, spectrophoto)	1.2	1.85	FreshWater Watch: Point A: NO3>10mg/L, PO4 entre 0.05 et 0.1mg/L, Turb<14NTU. Point C: NO3 entre 5 et 10 mg/L, PO4 entre 0.05 et 0.1mg/L, Turb<14NTU. Niveau de l'eau bas. Récupération des données de la sonde de surface au point B a été un échec: la sauvegarde a été du point A. Perte des données du 26/11 au 9/12 du capteur de la surface au point B.
07/01/2021	jeu	09:30	11:00	NA, GC		Temp., FluoroProbe, échantillon (fluospectro, spectrophoto)	0.9	2.85	Temp., salinity, FluoroProbe, échantillon (fluospectro, spectrophoto)	0.5	2	FreshWater Watch: Point A: NO3>10 mg/L, PO4 entre 0.05 et 0.1 mg/L, Turb = 17 NTU. Point B: NO3 entre 5 et 10 mg/L, PO4 entre 0.02 et 0.05 mg/L, Turb = 14 NTU. Présence de bcp d'oiseau en arrivant au point B.
28/01/2021	jeu	10:00	12:00	NA, Duong		ADCP + Temp., FluoroProbe, SeaBird, échantillon (fluospectro, spectrophoto)	1	3	Temp., salinity, FluoroProbe, SeaBird, échantillon (fluospectro, spectrophoto)	0.6	2.00	FreshWater Watch: Point A: NO3>10 mg/L, PO4 < 0.2 mg/L, Turb = 14 NTU. Point B: NO3 entre 5 et 10 mg/L, PO4 < 0.2 mg/L, Turb < 14 NTU. Beaucoup de vent direction amont. A la surface, l'eau semble aller vers l'amont.
10/02/2021	mer	10:40	13:40	NA, MS		Temp., FluoroProbe, SeaBird, échantillon (fluospectro, spectrophoto)	0.7	2.9	Temp., salinity, FluoroProbe, SeaBird, échantillon (fluospectro, spectrophoto)	0.5	2	FreshWater Watch: trop froid pour faire NO3 et PO4. Turb = 15 NTU (Point C). Il a neigé pendant la nuit et il y avait encore de la neige. Problème avec la récupération des données des capteurs nke: L'ordinateur était très lent et il s'est mis en veille très rapidement. Besoin de le mettre en charge pour continuer au point C. On n'a pas réussi à supprimer les données du capteur nke du fond au point C. Les données de ce capteur ont été conservées.

03/03/2021	mer	09:50	12:20	NA, MS		Temp., FluoroProbe, SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	1	2.9	Temp., salinity, FluoroProbe, SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	0.65	1.95	FreshWater Watch: Point A: NO3 entre 5 et 10 mg/L, PO4 entre 0.02 et 0.05 mg/L, Turb = 16 NTU. Point B: NO3 entre 5 et 10 mg/L, PO4 entre 0.05 et 0.1 mg/L, Turb < 14 NTU. Test de la sonde WiMo. Au point A, une petite bête était dans le capteur de salinité (un "mini crabbe blanc"). Le capteur était à l'exterieur de la cage.
08/04/2021	jeu	09:30	12:00	NA		Temp., échantillon (fluospectro, spectrophoto)	2.1	2.7	Temp., salinity, échantillon (fluospectro, spectrophoto)	1.1	2	FreshWater Watch: Point A: NO3 entre 5 et 10 mg/L, PO4 entre 0.02 et 0.05 mg/L, Turb < 14 NTU. Point B: NO3 entre 5 et 10 mg/L, PO4 entre 0.05 et 0.1 mg/L, Turb < 14 NTU. Au point A, une famille de petite bêtes était dans le capteur de salinité (des 'mini crevettes blanches'). Le capteur était à l'exterieur de la cage encore une fois. Tous les capteurs étaient très sales (+1 mois depuis le dernier passage)
29/04/2021	jeu	09:50	12:30	NA, Téo		Temp., WiMo, échantillon (fluospectro, spectrophoto)	2.3	2.7	Temp., salinity, WiMo, échantillon (fluospectro, spectrophoto)	2	2.00	FreshWater Watch: Point A et B: NO3 entre 5 et 10 mg/L, PO4 < 0.02 et 0.05 mg/L, Turb < 14 NTU. Au point A, deux mini crevettes blanches étaient encore dans le capteur de salinité. Tous les capteurs étaient très sales.
05/05/2021	mer	09:30	17:30	NA, PhD		Temp., WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	2.4	2.8	Temp., salinity, WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	1.6	2	Installation des modules de transmission automatiques aux points A (matinée) et B (après-midi). M. Mareschal nous a donné une clé pour ouvrir la cage. Changement d'emplacement du point B, maintenant à coté de l'échelle. FLUO3D: échantillons ont été surgélés pendant la nuit. Mesure faite le lendemain.
20/05/2021	jeu	09:30	11:40	NA, PhD		Temp., WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	2.3	2.7	Temp., salinity, WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	1.7	1.95	Récupération des capteurs nke au point B. FreshWater Watch: Point A et B: NO3 entre 5 et 10 mg/L, PO4 entre 0.05 et 0.1 mg/L, Turb < 14 NTU. Au point A, des mini crevettes blanches étaient encore dans le capteur de salinité et la sonde multiparamètres MS5. SPECTRO: cuvette étaient un peu bleue (?). Résultats peuvent être touchés par ça.
07/06/2021	lun	19:30	20:10	NA		échantillon			échantillon			FreshWater Watch: Point A et B: NO3 entre 5 et 10 mg/L, PO4 entre 0.02 et 0.05 mg/L, Turb < 14 NTU. Prise échantillon Point A à 19h35 et Point B à 20h00. Conservation des flacons au frigo, emballés au papier aluminium. Mesures de fluorescence à 10h le landemain, échantillons à la température ambiente (24C)
10/06/2021	jeu	09:45	11:40	NA, MS		SeaBird, WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	1.8	2.70	SeaBird,Temp., salinity, WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	1.3	1.9	Récupération du capteur CTD nke au point A. FreshWater Watch: Point A: NO3 entre 1 et 2 mg/L, PO4 entre 0.02 et 0.05 mg/L, Turb < 14 NTU. Point B: NO3 entre 2 et 5 mg/L, PO4 entre 0.05 et 0.10 mg/L, Turb < 14 NTU. Prise échantillon Point A à 10h15 et Point B à 11h30. Mesures de fluorescence à 14h le même jour, échantillons à la température ambiente (26.6C). Pas de circulation de bateaux (aucun type) la matinée à cause de l'installation des bouées jaunes de la baignade (plongeurs sur place pour attacher les cables - info donnée par Marin d'eau douce). Les algues fixes sont si grandes que MED ont fermé l'espace réservé aux bateaux pour les enfants.
17/06/2021	jeu	20:40	21:10	NA		échantillon			échantillon			Annonce d'orage vers 23h. Prise échantillon Point A à 20h45 et Point B à 21h05. Conservation des flacons au frigo, emballés au papier aluminium. Mesures de fluorescence à 16h le landemain (appareil pris le matin), échantillons à la température ambiente (28.4C)
18/06/2021	ven	09:00	09:30	NA		échantillon			échantillon			Prise échantillon Point A à 9h00 et Point B à 9h20. Conservation des flacons au frigo. Mesures de fluorescence à 14h (appareil pris le matin), échantillons à la température ambiente (28.4C)
22/06/2021	mar	09:50	11:30 (14h20)	NA, MS		SeaBird., WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	1.6	2.83	SeaBird., WiMo, Algaetorch, échantillon (fluospectro, spectrophoto)	0.95	1.95	FreshWater Watch: Point A et B: NO3 entre 5 et 10 mg/L, PO4 entre 0.1 et 0.2 mg/L, Turb < 14 NTU. Prise échantillon Point A à 10h25 et Point B à 11h10. Mesures de fluorescence à 16h le même jour, échantillons à la température ambiente (24.0C). Installation preleveur au pont Crimée, avec prise d'échantillon à 13h35 et 13h40, avec forte pluie. Algaetorche: 2 mesures de profil au point A et 2 au point B (doute du bon fonctionnement de l'équipement)
28/06/2021	lun	09:20	12:00	NA, AT		WiMo, échantillon (fluospectro, spectrophoto)	1.3	2.85	WiMo, échantillon (fluospectro, spectrophoto)	0.95	2	FreshWater Watch: Point A: NO3 entre 2 et 5 mg/L, PO4 entre 0.2 et 0.5 mg/L, Turb < 14 NTU. Point B: NO3 entre 5 et 10 mg/L, PO4 entre 0.05 et 0.1 mg/L, Turb < 14 NTU. Prise échantillon Point A à 10h35 et Point B à 11h40. Mesures de fluorescence à 14h30 le même jour, échantillons à la température ambiente (25.0C). Accompagnement de l'installation fluidion. Pas de souci avec notre équipement.
30/06/2021	mer	14:00	18:00	NA, AT	mesures fluo3D							Mesure de fluorescence: Pont de Crimée (nouveau point C) à la Villette, Sortie de la Station St-Thibault (Point TS) et entrée de la station St-Thibault (Point TEf et TEb: eau filtré et eau brut). Échantillons récupérés ce matin, vers 8h30-9h00. Mesure à 14h.
05/07/2021	lun	09:45	11:45	NA,AT		WiMo, échantillon (fluospectro, spectrophoto)	-	-	WiMo, échantillon (fluospectro, spectrophoto)	-	-	FreshWater Watch: Points A et B: NO3 entre 5 et 10 mg/L, PO4 entre 0.1 et 0.2 mg/L, Turb < 14 NTU. Prise échantillon Point A à 9h45 et Point B à 10h45. Mesures de fluorescence à 13h30 le même jour, échantillons à la température ambiante.
13/07/2021	mar	09:45	11:00	NA, AT		échantillon			échantillon			FreshWater Watch: Point A: NO3 entre 2 et 5 mg/L, PO4 entre 0.05 et 0.1 mg/L, Turb < 14 NTU. Point B: NO3 entre 2 et 5 mg/L, PO4 entre 0.1 et 0.2 mg/L, Turb < 14 NTU. Prise échantillon Point A à 10h00 et Point B à 11h40. Mesures de fluorescence à 13h30 le même jour, échantillons à la température ambiante (24.0C).

16/07/2021	ven	09:15	12:00	NA, AT		SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	1.05	2.8	SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	0.85	1.85	Prise échantillon Point A à 11h40 et Point B à 10h40 (Paris). Mesures de fluorescence à 14h00 le même jour, échantillons à la température ambiante (24.5C). Installation de l'ADCP proche de la plateforme Marin d'Eau Douce, à coté de la bouée en phare, appelé point TB-D (transect Baignade rive Droite). Secchi = 0.85m, total = 2.65 m + profil avec SeaBird.
19/07/2021	lun	08:15	09:00	NA, AT	mesures fluo3D							Mesure de fluorescence: Pont de Crimée (nouveau point C) à la Villette, Sortie de la Station St-Thibault (Point TS) et entrée de la station St-Thibault (Point TEf et TEb: eau filtré et eau brut). Échantillons récupérés ce matin, vers 8h20-8h40. Mesure à 11h. Température ambiante = 27 °C
23/07/2021	ven	09:00	12:00	NA, AT		SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	0.8	2.8	SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	0.9	1.9	Prise échantillon Point A à 11h45 et Point B à 10h10 (Paris). Mesures de fluorescence à 14h00 le même jour, échantillons à la température ambiante (28.5C). Essai de récuperer l'ADCP proche de la plateforme Marin d'Eau Douce, à coté de la bouée en phare, appelé point TB-D (transect Baignade rive Droite), mais sans succès. À coté du point A, il y avait un oiseau mort sur l'eau.
26/07/2021	lun	09:00	12:50	NA, AT		Échantillon (fluospectro, spectrophoto)	-	-	Échantillon (fluospectro, spectrophoto)	-	-	Prise échantillon Point A à 12h45 et Point B à 11h15 (Paris). Mesures de fluorescence à 14h00 le même jour, échantillons à la température ambiante (26.3C). Essai de récuperer l'ADCP au point TB-D (transect Baignade rive Droite) avec un aimant, mais sans succès.
06/08/2021	ven	09:40	10:10	Amanda (AT)		Échantillon (fluospectro, spectrophoto)	-	-	Échantillon (fluospectro, spectrophoto)	-	-	Prise échantillon Point A à 9h40 et Point B à 10h10 (Paris). Mesures de fluorescence à 11h20 le même jour, échantillons à la température ambiante (24.4C).
25/08/2021	mer	10:10	11:40	NA, AT, MJ		Échantillon (fluospectro, spectrophoto)	1.15	2.8	Échantillon (fluospectro, spectrophoto)	1	1.95	WiMo ne demarrait pas. Prise échantillon Point A à 10h25 et Point B à 11h25 (Paris). Mesures de fluorescence à 14h00 le même jour, échantillons à la température ambiante (25.8C).
01/09/2021	mer	10:20	14:00	NA,PhD,João-Miguel								Changement de panneau solaire au point A, avec orientation vers le sud. Équipements fluidions déjà enlevés. Renforcement d'attachement des sondes CTD dans leur corde respective aux points A et B. Préleveur point C: Nettoyage le 02/09/2021 avec eau de javel + eau robinet.
07/09/2021	mar	10:20	10:40	NA		Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h20 et Point B à 10h35. Conservation des flacons au frigo. Mesures de fluorescence à 13h, échantillons à la température ambiente (27.4C)
08/09/2021	mer	08:30	09:00	Mohamed	Échantillon (fluospectro, spectrophoto)							Prise des échantillons journaliers moyens du pont de Crimée et de St-Thibault. Mesures fluo3d et absorbance realisées à partir de 13h. Dilution des échantillons de St-Thibault: d5 et d10, entrée et sortie.
09/09/2021	jeu	08:30	09:00	Mohamed								Prise des échantillons journaliers moyens de St-Thibault. Mesures fluo3d et absorbance realisées à partir de 13h. Dilution des échantillons de St-Thibault: d5 et d10, entrée et sortie, d2 sortie.
13/09/2021	lun	10:20	11:00	NA	Échantillon (fluospectro, spectrophoto)	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Pigeons se baignaient au point A, avec liquide blanc en surface. Prise échantillon Point A à 10h30 et Point B à 10h55. Mesures de fluorescence à 13h, échantillons à la température ambiante (24.9C). Blanc EUP a eu de l'interference. Forte chance que la cuve aille des traces des échantillons de St-Thibault. Nettoyage avec acetone pour les échantillons de demain. Prélèvement au point A à 20h25 et point B à 21h00.
14/09/2021	mar	08:30	08:40	Mohamed	Échantillon (fluospectro, spectrophoto)							Prise des échantillons journaliers moyens du pont de Crimée et de St-Thibault. Mesures fluo3d et absorbance realisées à partir de 12h. Dilution des échantillons de St-Thibault: d10 entrée et d5 sortie. Mesures à la température abiente de 25.7C
15/09/2021	mer	08:20	09:00	NA	Échantillon (fluospectro, spectrophoto)	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 8h20 et Point B à 9h00. Mesures de fluorescence à 13h, échantillons à la température ambiente (26.4C). Prise échantillon point C à 8h30. Pluie constante toute la journée. Problème avec la cuve s'est reproduit.
16/09/2021	jeu	08:30	09:20	NA	Échantillon (fluospectro, spectrophoto)	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 8h45 et Point B à 9h10. Mesures de fluorescence à 11h, échantillons à la température ambiente (25.9C). Prise échantillon point C à 8h30. Dilution des échantillons de St-Thibault: d10 entrée et d5 sortie. Problème avec prélèvement à St-Thibault: les échantillons reçus sont passés à la centrifuge à Créteil avant.
17/09/2021	ven	08:30	09:20	NA	Échantillon (fluospectro, spectrophoto)	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 8h45 et Point B à 9h00. Mesures de fluorescence à 11h, échantillons à la température ambiente (25.8C). Prise échantillon point C à 8h10.
20/09/2021	lun	08:40	09:20	NA	Échantillon (fluospectro, spectrophoto)	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 9h50 et Point B à 11h10. Mesures de fluorescence à 12h, échantillons à la température ambiente (24.8C). Prise échantillon point C à 10h30. Prélèvement point C de 8h à 8h, du 19 au 20/09.
21/09/2021	mar	09:00	11:30	NA	Échantillon (fluospectro, spectrophoto)	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h et Point B à 11h10. Mesures de fluorescence le lendemain, échantillons stockés au frigo. Prélèvement point C de 9h20 à 9h20, du 20 au 21/09.
22/09/2021	mer	08:50	10:00	NA	Échantillon (fluospectro, spectrophoto)	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 8h50 et Point B à 10h00. Mesures de fluorescence à 11h30 (y compris les échantillons d'hier), échantillons à la température ambiante (21.2C). Prélèvement point C de 9h30 à 9h30, du 21 au 22/09.
04/10/2021	lun	09:30	10:00	NA								Prise des échantillons journaliers moyens de St-Thibault. Mesures fluo3d et absorbance realisées à partir de 12h, échantillons à la température ambiante (21.6C). Dilution des échantillons de St-Thibault: d5 sortie et d10 entrée.

19/10/2021	mar	10:30	13:00	NA, Maria Julia	WiMo, échantillon (fluospectro, spectrophoto)	1.45	2.75	WiMo, échantillon (fluospectro, spectrophoto)	1.2	1.9	Prise échantillon Point A à 11h40 et Point B à 12h40. Mesures de fluorescence à 15h, échantillons à la température ambiante (24.5C). Colliers en plastique des capteurs de température au point A étaient cassés. Les capteurs du milieu et en surface étaient au fond de l'eau. Remise d'autres colliers. Besoin de voir altérnative, car les colliers en plastique se cassent très fréquemment à la Villette. Prevoir une cordelette pour attacher les capteurs de température.
24/11/2021	mer	10:30	12:00	NA, PhD, Maria Julia	WiMo, échantillon (fluospectro, spectrophoto)	1.5	2.8	WiMo, échantillon (fluospectro, spectrophoto)	0.8	1.9	Remplacement de batterie au point A. Changement pas de temps de mesure aux points A et B: passé de 10 min à 20 min pour cet hiver. Mesure pH, conductivité spécifique et turbidité au labo. Point A: pHA = 8.00, condA = 801 uS/cm turbA = 8.00 NTU (moyenne entre 8.05 et 7.96 NTU); Point B: pHB = 8.02, condB = 801 uS/cm, turbB = 6.37 NTU (2 mesures égales). Prise échantillon Point A à 10h50 et Point B à 11h45. Mesures de fluorescence à 14h, échantillons à la température ambiente (19.6C). Fermeture Marin d'Eau Douce: du 01/12/21 au 01/03/22.
12/01/2022	mer	10:00	15:00	NA, PhD	SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	0.45	2.8	SeaBird, WiMo, échantillon (fluospectro, spectrophoto)	0.4	1.9	Remplacement des batteries (neuves) et installation de régulateurs entre batterie et panneaux solaires aux points A et B. Remplacement du panneau solaire au point A (puissance 20 W à la place de 10 W). MAJ du programme pour envoyer la valeur de la mesure seulement quand une mesure est faite, avec batterie > 12.2V. Prise échantillon Point A à 11h05 et Point B à 13h35. Mesures de fluorescence à 12h, échantillons à la température ambiante (20.7C).
31/01/2022	lun	17:40	18:50	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 18h10 et Point B à 18h30. Mesures de fluorescence le 02/02, avec les échantillons du même jour. Vérification de la tension des batteries. Point A = 11.1 V, LED indiquant batterie complètement déchargée. Point B = 12.5V, LED vert (ok). Panneaux solaires bien placés.
02/02/2022	mer	09:30	11:40	NA, Joao-Miguel	WiMo, échantillon (fluospectro, spectrophoto)	1	2.8	WiMo, échantillon (fluospectro, spectrophoto)	0.75	1.9	Prise échantillon Point A à 10h40 et Point B à 11h15. Mesures de fluorescence à 13h30 (y compris les échantillons du 31/01), échantillons à la température ambiante (24.2C). Récupération de la sonde OTT MS5 du point A pour mesurer sa consommation de batterie au labo. Mise d'un capteur nke (température et salinité) à sa place, à 0.50m. Le LED panneau solaire était vert clignotant (panneau en cours de production d'énergie).
16/02/2022	mer	10:00	12:00	NA, PhD, Arthur	SeaBird, échantillon (fluospectro, spectrophoto)	1.05	2.8	SeaBird, échantillon (fluospectro, spectrophoto)	0.8	1.9	Prise échantillon Point A à 11h00 et Point B à 11h50. Mesures de fluorescence à 14h00, échantillons à la température ambiante (24.4C). Remplacement de la batterie au point A. Recuperation des donnees du capteur nke et des mesures de niveau d'eau et temperature pas transmises (mesures jusqu'au 30 janvier). WiMo pas utilisée car le dispositif d'activation n'était pas dans le casier (on a trouvé un aimant dans le bureau de NA qui peut remplacer le dispositif pour l'instant). Presence d'une dizaine de mouettes au point A à notre arrivée (une personne était en train de les alimenter).
15/03/2022	mar	09:00	10:10	NA, Arthur	WiMo, échantillon (fluospectro, spectrophoto)	-	-	WiMo, échantillon (fluospectro, spectrophoto)	-	-	Prise échantillon Point A à 09h20 et Point B à 10h00. Mesures de fluorescence à 11h30, échantillons à la température ambiante (24.6C). Liquide blanchâtre huileux à la surface du point A. Mesure de turbidité au labo. Secchi n'a pas été mesuré et les données de salinité du point A n'ont pas été récupérées (Secchi et datapencil utilisé à la campagne du lac de Champs). Toutes les sondes ont été nettoyées.
29/03/2022	mar	09:00	11:00	NA, Arthur	Turbidité,WiMo, échantillon (fluospectro, spectrophoto), Freshwaterwatch	1.7	2.8	Turbidité,WiMo, échantillon (fluospectro, spectrophoto), Freshwaterwatch	1.8	1.9	Prise échantillon Point A à 09h10 et Point B à 10h20. Mesures de fluorescence à 12h00, échantillons à la température ambiante (24.3C). FreshWater Watch: Points A et B: NO3 entre 5 et 10 mg/L, PO4 entre 0.02 et 0.05 mg/L (point A) et entre 0.05-0.1 mg/L (point B), Turb < 14 NTU. Pas de vent en amont ni en aval. Mesure de turbidité en terrain et au labo. Toutes les sondes (sauf température point A) ont été nettoyées.
30/03/2022	mer	09:20	10:20	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 9h30 et Point B à 10h00. Mesures de fluorescence à 11h, échantillons à la température ambiante (23.3C). Liquide huileux en surface au point A.
31/03/2022	jeu	09:20	10:40	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 9h30 et Point B à 10h00. Mesures de fluorescence à 11h, échantillons à la température ambiante (23.3C).
01/04/2022	ven	09:30	10:20	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 9h50 et Point B à 10h30. Mesures de fluorescence à 11h, échantillons à la température ambiante (21.6C).
02/04/2022	sam	09:30	10:20	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h00 et Point B à 10h20. Mesures de fluorescence à 11h, échantillons à la température ambiante (21.8C).
03/04/2022	dim	14:00	15:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 14h30 et Point B à 14h50. Mis au frigo. Mesures de fluorescence le lendamain, avec les nouveaux échantillons.
04/04/2022	lun	09:30	11:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h00 et Point B à 10h20. Mesures de fluorescence à 11h, avec les échantillons d'hier. Echantillons à la température ambiante (24.3C)
08/04/2022	ven	18:00	20:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 19h00 et Point B à 19h20. Mis au frigo. Mesures de fluorescence le lendamain, avec les nouveaux échantillons.
09/04/2022	sam	10:00	12:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h40 et Point B à 11h10. Mesures de fluorescence à 12h, avec les échantillons d'hier. Echantillons à la température ambiante (23.1C)
10/04/2022	dim	13:30	14:30	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 14h10 et Point B à 14h00. Mis au frigo. Mesures de fluorescence le lendamain, avec les nouveaux échantillons.
11/04/2022	lun	09:30	11:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 09h50 et Point B à 10h10. Mesures de fluorescence à 11h, avec les échantillons d'hier. Echantillons à la température ambiante (23.8C)
12/04/2022	mar	08:30	10:40	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 9h20 et Point B à 09h40. Mesures de fluorescence à 11h, échantillons à la température ambiante (24.4C).

20/04/2022	mer	09:00	12:00	NA,Arthur							Installation de la sonde CTD au point A (MS5 OTT). Pas de temps reste à 20 min. La sonde nke reste sur place. Nettoyage de tous les capteurs et de la cage.
22/04/2022	ven	09:00	11:40	NA, Arthur	Turbidité,WiMo, échantillon (fluospectro, spectrophoto),Freshwaterwatch	2	2.7	Turbidité,WiMo, échantillon (fluospectro, spectrophoto),Freshwaterwatch	1.5	1.85	Prise échantillon Point A à 9h20 et Point B à 10h00. Mesures de fluorescence à 12h, échantillons à la température ambiante (23.3C). FreshWater Watch: Point A: NO3 entre 5 et 10 mg/L, PO4 entre 0.02 et 0.05 mg/L,Turb < 14 NTU; Point B: NO3 entre2-5 mg/L et PO4 entre 0.05-0.1 mg/L, Turb < 14 NTU. Mesure de turbidité au labo.
05/05/2022	jeu	09:00	12:00	NA,MS,Arthur	SeaBird,WiMo, échantillon (fluospectro, spectrophoto, turbidité)	2.4	2.75	SeaBird,WiMo, échantillon (fluospectro, spectrophoto, turbidité)	1.5	1.8	2 profils avec WiMo et SeaBird ont été faits aux 2 points, sans interruption de mesure. Prise échantillon Point A à 9h30 et Point B à 10h20. Mesures de fluorescence à 12h, échantillons à la température ambiante (23.6C). Turbidité faite au labo après X jours (appareil était pris)
16/05/2022	lun	18:00	19:30	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 18h50 et Point B à 19h10. Mesures de fluorescence le lendemain, échantillons stockés au frigo. Mesures à 9h00, température ambiante (27.1C)
15/06/2022	mer	09:00	12:00	NA, Arthur	Turbidité,WiMo, échantillon (fluospectro, spectrophoto),Freshwaterwatch	1.1	2.78	Turbidité,WiMo, échantillon (fluospectro, spectrophoto),Freshwaterwatch	1.2	1.88	Récupération de la sonde nke au point A. Côte par rapport à la berge: ptA = 2.72 m. Profondeur amont pont Crimée = 2.30 m (3.80m jusqu'à la berge). Profondeurs en aval (sortie du bassin): rive droite = 2.78 m; milieu = 2.62m; rive gauche écluses = 2.10m. Prise échantillon Point A à 9h50 et Point B à 10h30. Mesures de fluorescence à 12h, échantillons à la température ambiante (28.6C).
20/06/2022	lun	09:30	11:30	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h00 et Point B à 10h30. Mesures de fluorescence à 12h, échantillons à la température ambiante (27.9C).
21/06/2022	mar	09:00	11:00	Arthur	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 09h10 et Point B à 09h30. Mesures de fluorescence à 11h, échantillons à la température ambiante (28.4C).
22/06/2022	mer	08:30	10:00	Arthur	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 08h50 et Point B à 09h10. Mesures de fluorescence à 10h, échantillons à la température ambiante (26.8C). Beaucoup de déchets en surface de l'eau le long du canal et du bassin (fête de la musique hier soir).
23/06/2022	jeu	09:00	11:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 09h50 et Point B à 10h20. Mesures de fluorescence à 12h, échantillons à la température ambiante (28.8C).
24/06/2022	ven	09:30	11:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h00 et Point B à 10h20. Mesures de fluorescence à 12h, échantillons à la température ambiante (28.4C).
21/07/2022	jeu	09:00	11:00	Ana Carolina P.P., Clélia R.	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 09h30 et Point B à 09h50. Mesures de fluorescence à 12h, échantillons à la température ambiante (30.2C).
12/08/2022	ven	09:00	14:00	NA, PhD, Vincent	WiMo, échantillon (fluospectro, spectrophoto)	1	2.8	WiMo, échantillon (fluospectro, spectrophoto)	1.05	1.9	Prise échantillon Point A à 10h00 et Point B à 12h20. Mesures de fluorescence à 15h, échantillons à la température ambiante (30.5C). Intervention au point A: réparation de la hauteur du capteur de niveau d'eau et réparation de la fixation du boîtier imperméable. Les tendeurs se sont cassés vers le 20/07. Une fixation provisoire a été faite par les agents de Fluidion entretemps.
16/08/2022	mar	09:00	11:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 09h20 et Point B à 09h40. Mesures de fluorescence à 11h, échantillons à la température ambiante (27.5C).
17/08/2022	mer	09:00	11:00	NA	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h20 et Point B à 10h40. Mesures de fluorescence à 12h, échantillons à la température ambiante (28.1C).
18/08/2022	jeu	10:00	13:00	NA, Vincent	WiMo, échantillon (fluospectro, spectrophoto)			WiMo, échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h30 et Point B à 11h00. Mesures de fluorescence à 14h, échantillons à la température ambiante (29.5C).
19/08/2022	ven	09:00	11:00	Vincent	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 9h10 et Point B à 9h45. Mesures de fluorescence à 14h, échantillons à la température ambiante (27,1C).
23/08/2022	mar	-	-	Fluidion							Analyseurs Fluidion retirés le 23/08/2022.
09/09/2022	ven	09:00	12:00	NA, Lorena,Romero	WiMo, BBE, échantillon (fluospectro, spectrophoto)	0.6	2.7	WiMo, BBE, échantillon (fluospectro, spectrophoto)	0.5	1.9	**Début campagne au point B. Installation des capteurs avec Manel aux points B et A. Réglage hauteur sonde de niveau d'eau avec le disque Secchi. Un serre-cable a été mis pour empêcher le capteur de glisser vers le fond. Besoin des éponges et de la brosse rouge pour nettoyer les capteurs de conductivité pour la prochaine campagne. Vent très présent, direction amont. Prise échantillon point A à 10h40 et Point B à 10h00. Mesures de fluorescence à 14h, échantillons à la température ambiante (23.2C).**
15/09/2022	jeu	09:00	12:30	NA, Lorena, Pedro	WiMo, BBE, échantillon (fluospectro, spectrophoto)	0.8	2.7	WiMo, BBE, échantillon (fluospectro, spectrophoto)	0.8	1.85	Récupération des données des capteurs de Manel. Prise échantillon Point A à 9h45 et Point B à 10h50. Mesures de fluorescence à 14h, échantillons à la température ambiante (22,5C). Absence de vent (quelques brises vers l'aval).
16/09/2022	ven	10:00	11:30	Lorena	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 10h20 et Point B à 10h50. Mesures de fluorescence à 13h, échantillons à la température ambiante (19.5C).
22/09/2022	jeu	09:30	11:30	NA, Lorena	BBE, échantillon (fluospectro, spectrophoto)	0.9	2.7	BBE, échantillon (fluospectro, spectrophoto)	0.7	1.9	Prise échantillon au Point A à 9h50, pas de vent et environ une dizaine de pigeons près du point A. Prise échantillon au Point B à 10h50, vent très léger direction aval. Mesures de fluorescence à 14h, échantillons à la température ambiante (20.5C).
27/09/2022	mar	08:30	11:00	Lorena	Échantillon (fluospectro, spectrophoto)			Échantillon (fluospectro, spectrophoto)			Prise échantillon Point A à 8h50 et Point B à 10h40, vent très léger direction aval. Copie de clé pour Manel. Mesures de fluorescence à 14h30, échantillons à la température ambiante (17.0C).
14/10/2022	ven	09:30	11:00	Lorena, Arthur	WiMo, échantillon (fluospectro, spectrophoto)	0.8	2.8	WiMo, échantillon (fluospectro, spectrophoto)	1.2	1.8	Prise échantillon Point A à 9h38 et Point B à 10h15. Mesures de fluorescence à 14h, échantillons à la température ambiante (18.7C).

17/10/2022	lun	11:00	12:00	Lorena, Romero	WiMo, échantillon (fluospectro, spectrophoto)	1.1	2.7	WiMo, échantillon (fluospectro, spectrophoto)	1.1	1.8	Prise échantillon Point A à 11h00 et Point B à 11h45. Mesures de fluorescence le jour 18/10/2022 10h00, échantillons à la température ambiante (23.0C).
17/10/2022	lun	13:00	14:00	Lorena, Romero	WiMo, échantillon (fluospectro, spectrophoto)	1.1	2.7	WiMo, échantillon (fluospectro, spectrophoto)	1	1.8	Prise échantillon Point A à 13h00 et Point B à 13h35. Mesures de fluorescence le jour 18/10/2022 10h00, échantillons à la température ambiante (23.0C).
17/10/2022	lun	15:00	16:00	Lorena, Romero	WiMo, échantillon (fluospectro, spectrophoto)	1.1	2.7	WiMo, échantillon (fluospectro, spectrophoto)	0.9	1.8	Prise échantillon Point A à 15h10 et Point B à 15h45. Mesures de fluorescence le jour 18/10/2022 10h00, échantillons à la température ambiante (23.0C).
17/10/2022	lun	17:00	18:00	Lorena, Romero	WiMo, échantillon (fluospectro, spectrophoto)	1.1	2.7	WiMo, échantillon (fluospectro, spectrophoto)	0.9	1.8	Prise échantillon Point A à 17h07 et Point B à 17h40. Mesures de fluorescence le jour 18/10/2022 10h00, échantillons à la température ambiante (23.0C). Prise de 2 flacons pour comparer deux mesures d'un même échantillon..
09/11/2022	mer	16:00	18:00	NA, Romero	WiMo			WiMo			Remplacement des batteries (neuves) aux points A et B. Arrivée à La Villette en fin d'après-midi. La batterie au point A était déchargée (12.1 V). La batterie au point B était chargée (quand nous sommes arrivés), avec 12.6 V. Pas de Secchi: très sombre (manque de lumière solaire).
11/01/2023	mer	09:00	12:00	NA, Arthur	WiMo, BBE, échantillon (fluospectro, spectrophoto)	1	2.7	WiMo, BBE, échantillon (fluospectro, spectrophoto), bidon 2*5L (forbath), échantillon 500ml (contrôle cond., new sondes ott)	0.8	1.9	Prise échantillon Point A à 10h00 et Point B à 11h30. Mesures de fluorescence le jour même, échantillons à la température ambiante (18.5C). Changement des batteries, et nouvelles configuration du cablage PMV (régulateur) et mode C, au deux points. Vérification avec pdt de 1min des mesures effectuées. Remplissage des bidons avec un sceau et rinçage avec l'eau du canal. Faire attention à l'état des vis pour ouvrir le boitier.
17/02/2023	ven	09:30	13:00	Guilherme, Arthur	WiMo, échantillon fluo	0.9	2.7	WiMo, échantillon fluo	0.9	1.8	Pas de nettoyage des capteurs, démontage de la station au point B pour révision sonde OTT et ajout capteur de profondeur. Prise échantillon Point A à 09h30 et Point B à 10h30. Mesures de fluorescence le jour 17/02/2023 13h30, échantillons à la température ambiante (21.0C).
08/03/2023	mer	09:30	12:00	NA, AGLG	WiMo, échantillon (fluospectro, spectrophoto), échantillons 250ml pour bacterio EDP	1.4	2.75	WiMo, échantillon (fluospectro, spectrophoto), échantillons 250ml pour bacterio EDP	1.05	1.85	Prise échantillon Point A à 09h50 et Point B à 10h20. Mesures de fluorescence le jour 08/03/2023 13h30, échantillons à la température ambiante (21.0C).
08/03/2023	mer	19:00	20:30	NA, AGLG	WiMo, échantillon (fluospectro, spectrophoto), échantillons 250ml pour bacterio EDP	NA	NA	WiMo, échantillon (fluospectro, spectrophoto), échantillons 250ml pour bacterio EDP	NA	NA	Prise échantillon Point A à 19h50 et Point B à 20h20. Mesures de fluorescence le jour 09/03/2023 13h00, échantillons à la température ambiante (21.0C).
09/03/2023	jeu	10:00	12:00	NA, AGLG	WiMo, échantillon (fluospectro, spectrophoto), échantillons 250ml pour bacterio EDP	1.2	2.8	WiMo, échantillon (fluospectro, spectrophoto), échantillons 250ml pour bacterio EDP	1.1	1.9	Prise échantillon Point A à 10h15 et Point B à 10h50. Mesures de fluorescence le jour 09/03/2023 13h00, échantillons à la température ambiante (21.0C).

IV-d. Vertical profiles in La Villette system

La Villette profile measurements follow-up

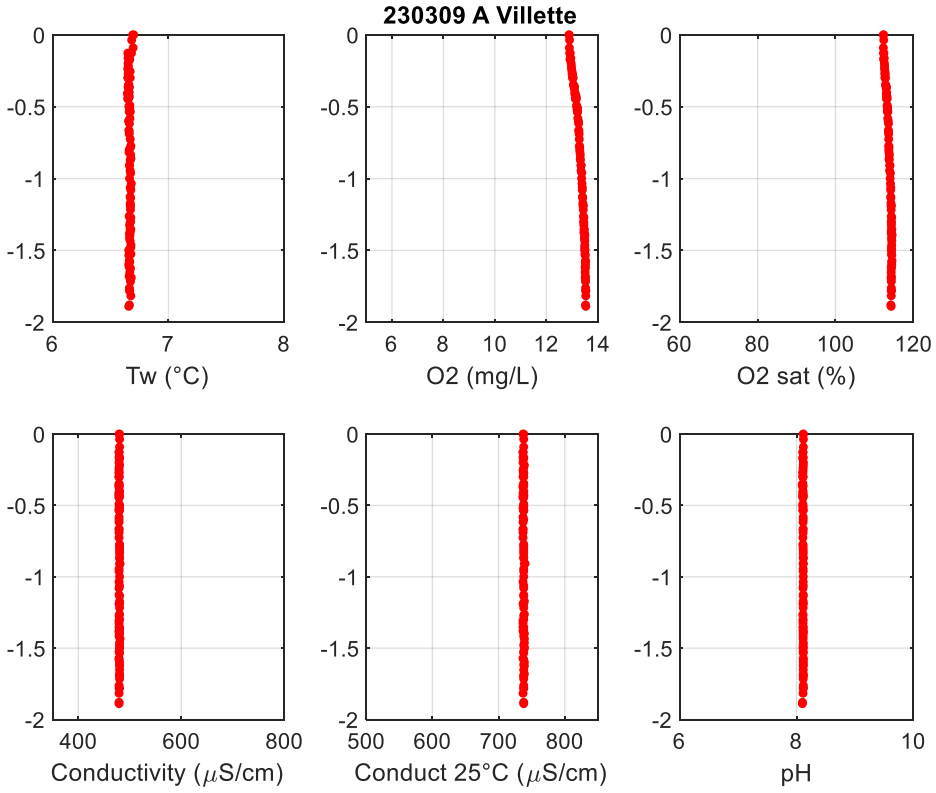
Last update: 12/01/2023

Contents

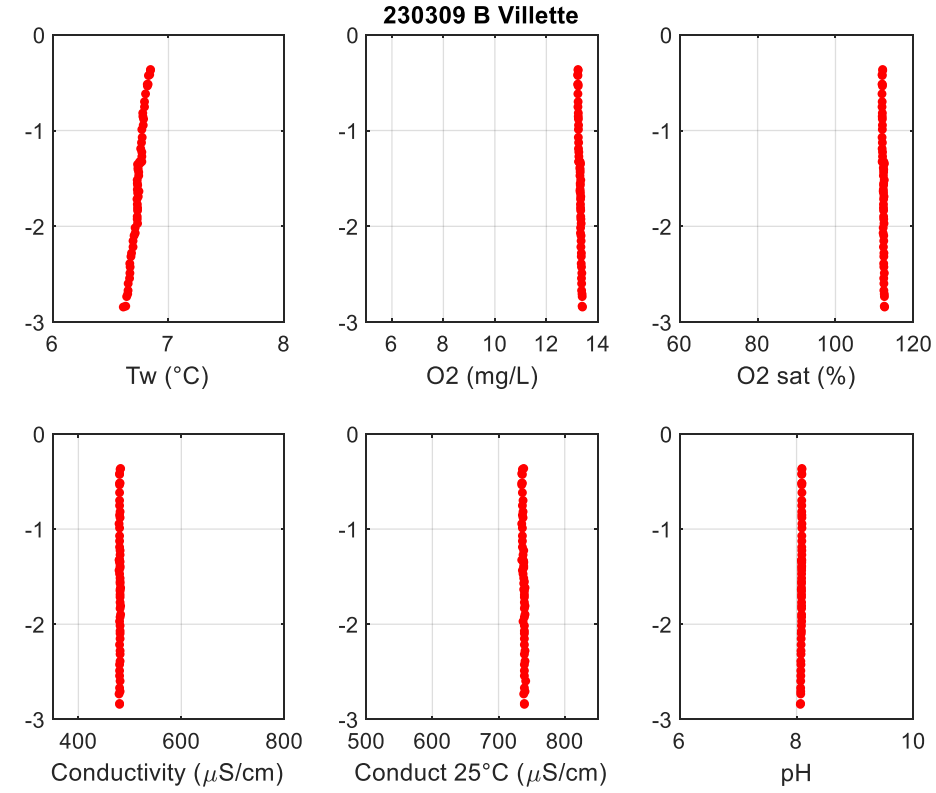
1. WiMo profiles

09/03/2023

Point A – 10:20 (UTC)

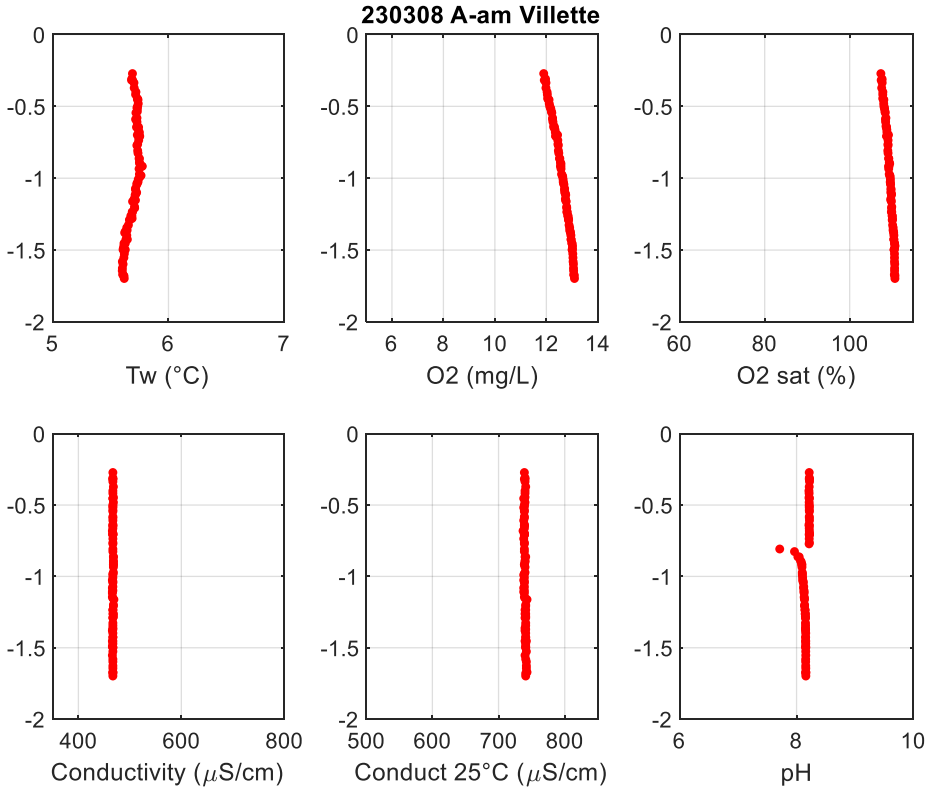


Point B – 11:00 (UTC)

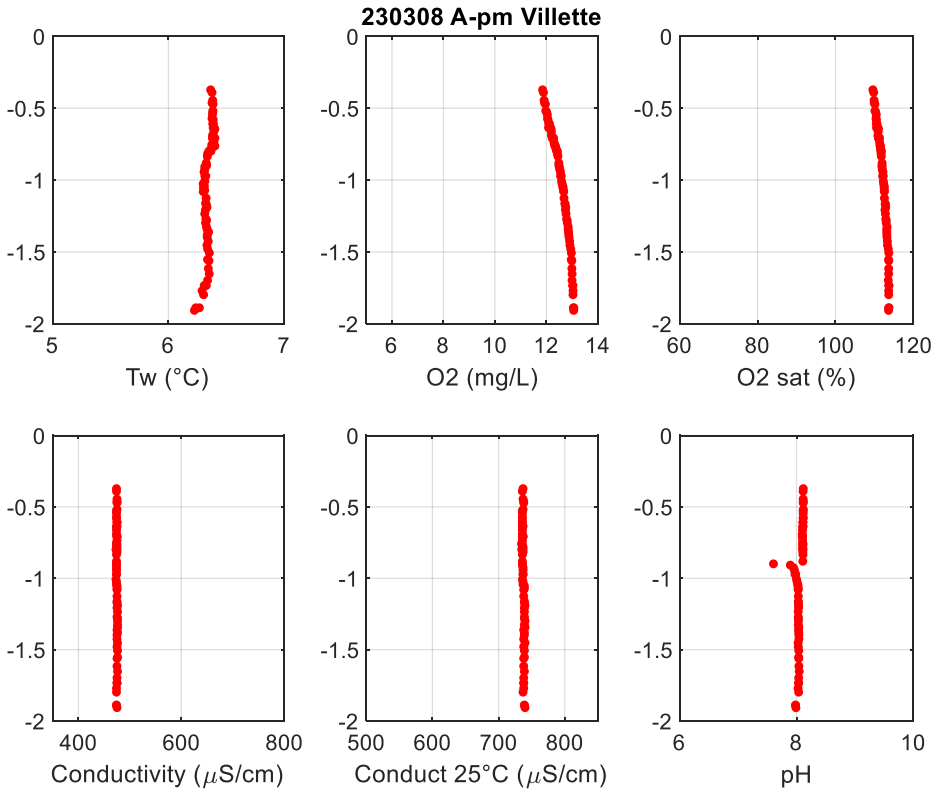


08/03/2023 – a.m. and p.m.

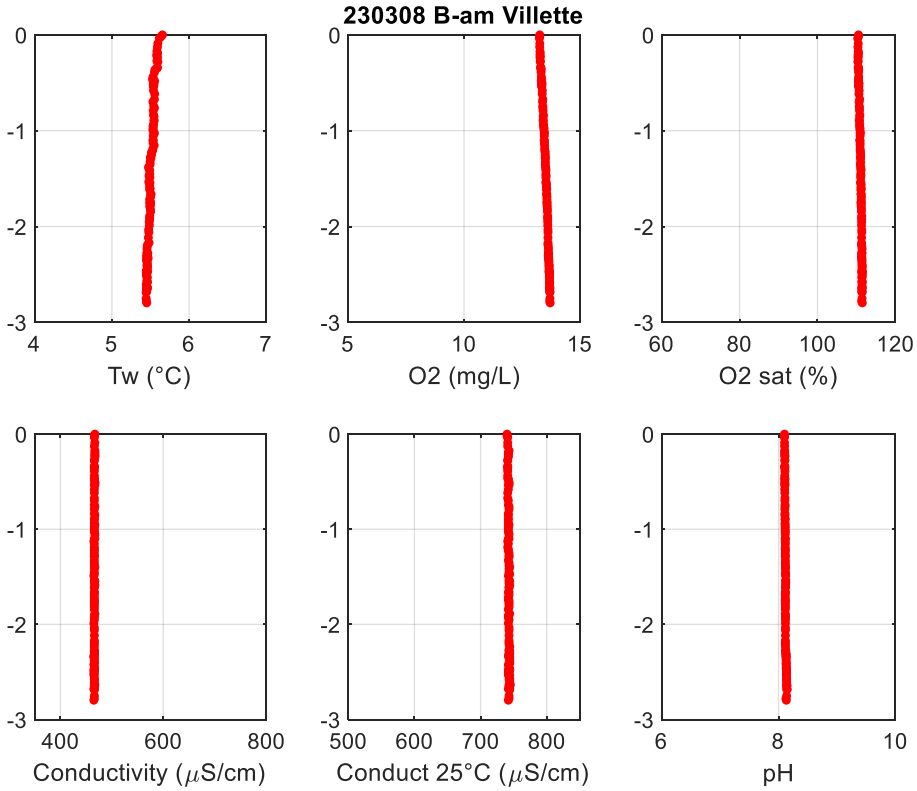
Point A – 09:50 (UTC)



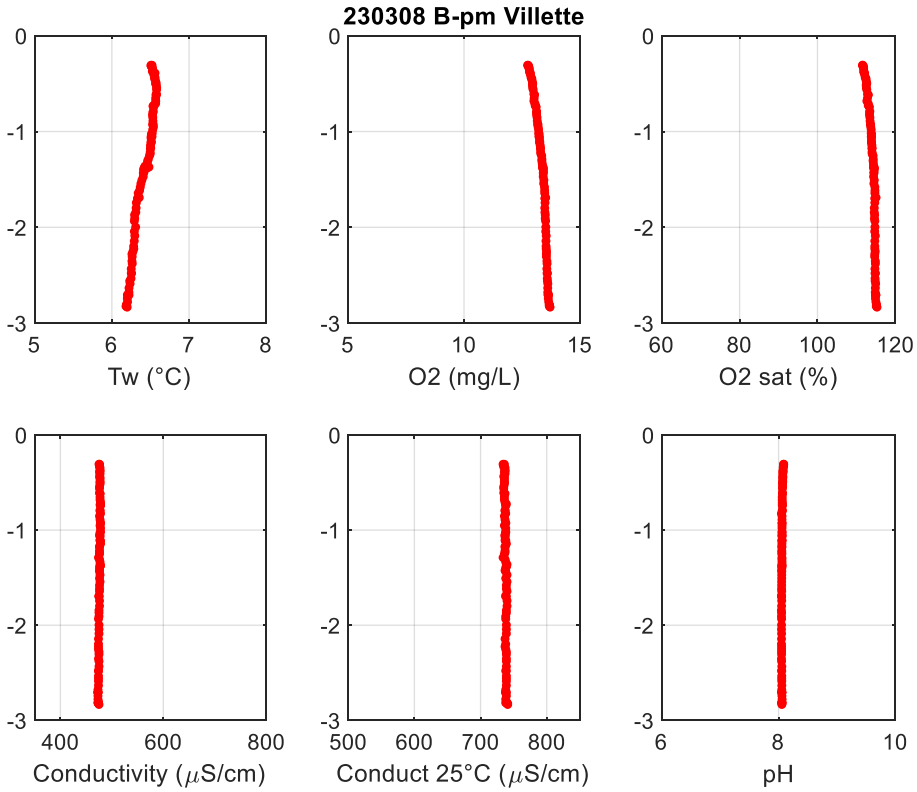
Point A – 19:50 (UTC)



Point B – 10:30 (UTC)

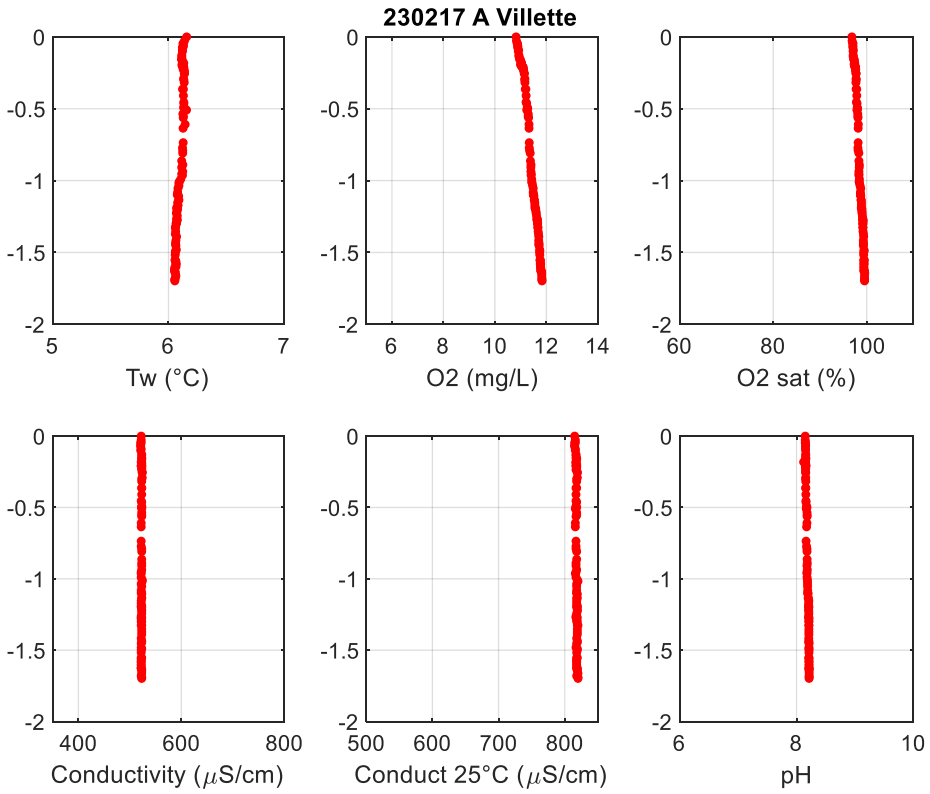


Point B – 20:20 (UTC)

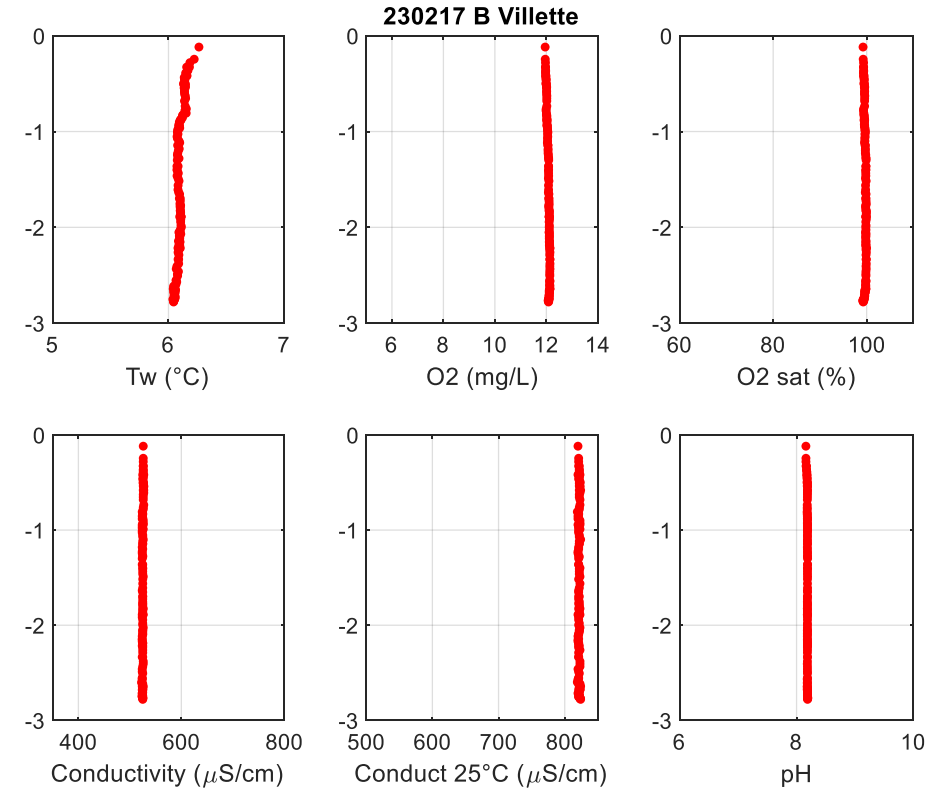


17/02/2023

Point A – 08:50 (UTC)

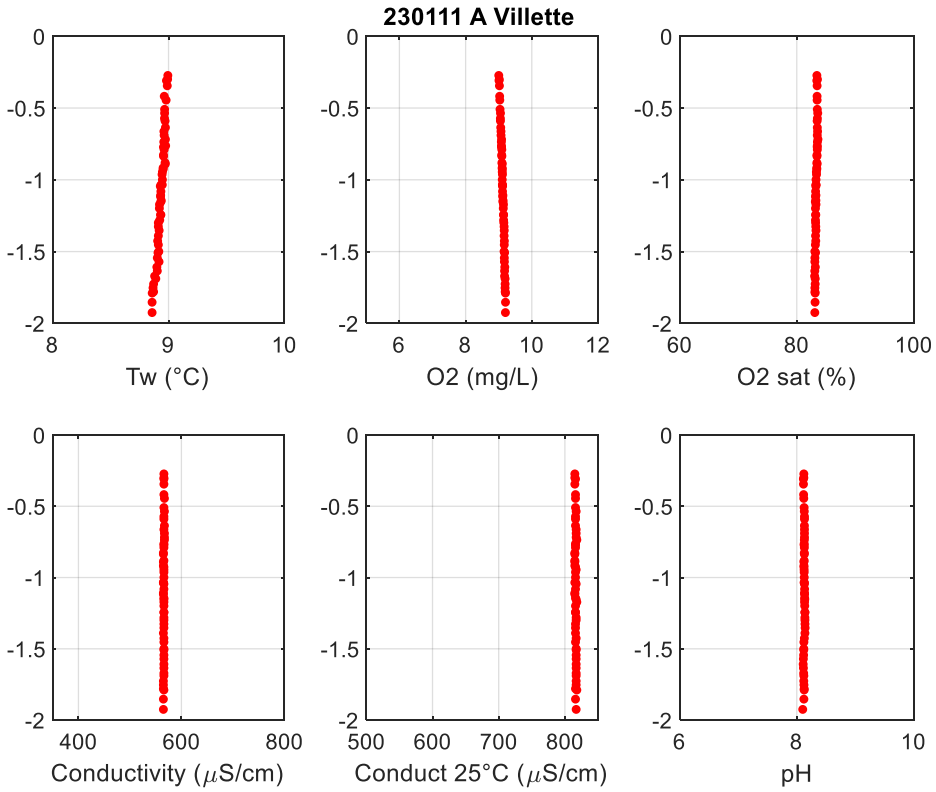


Point B – 09:50 (UTC)

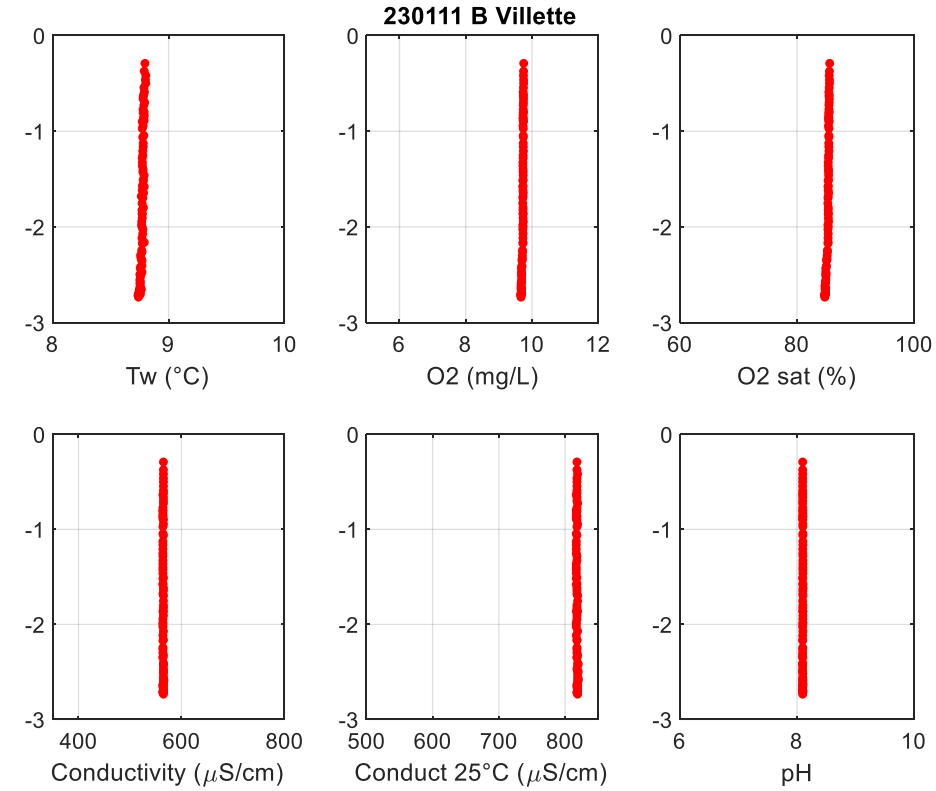


11/01/2023

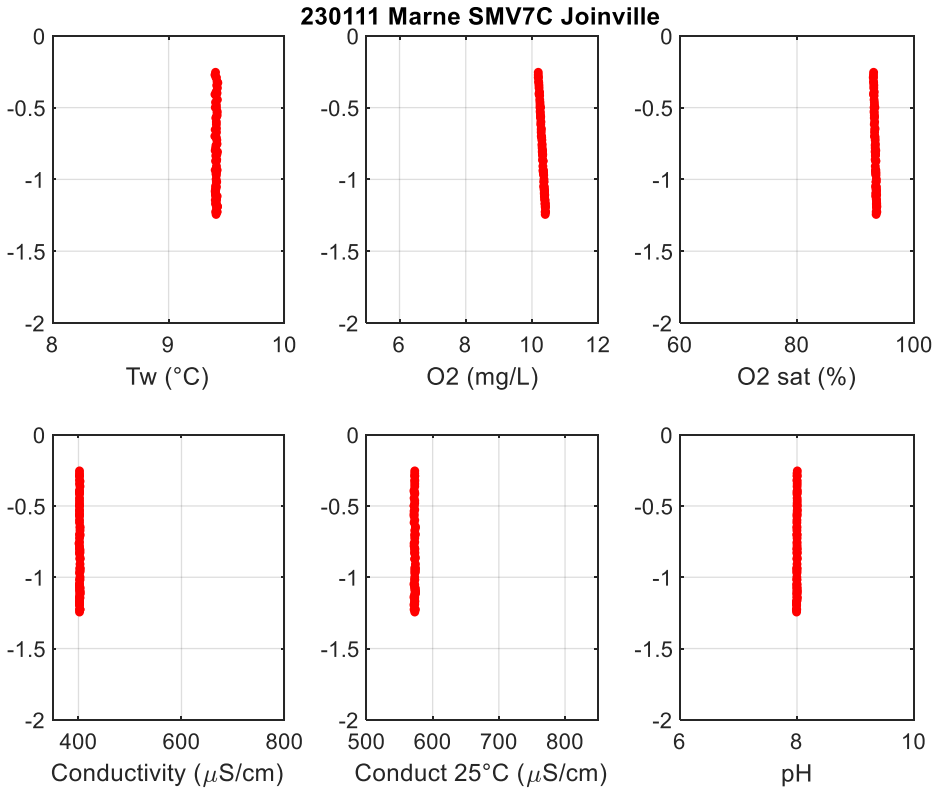
Point A – 09:30 (UTC)



Point B – 11:14 (UTC)

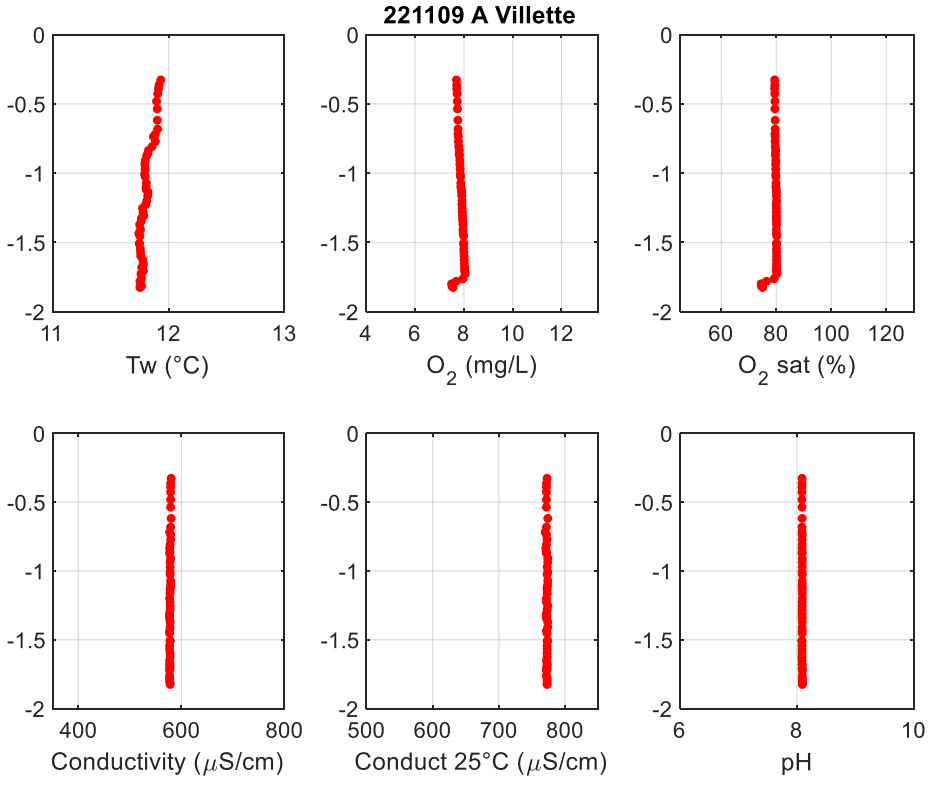


At the same day, measurement at Joinville (Marne), point SMV7C, at 13:30 (UTC):

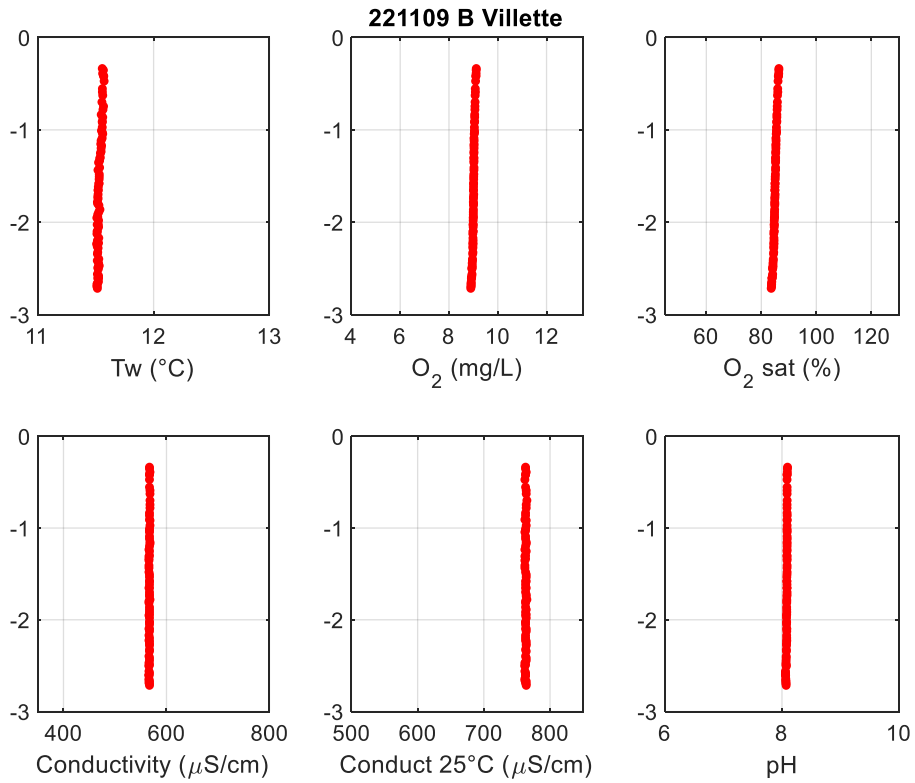


09/11/2022

Point A – 16:20 (UTC)

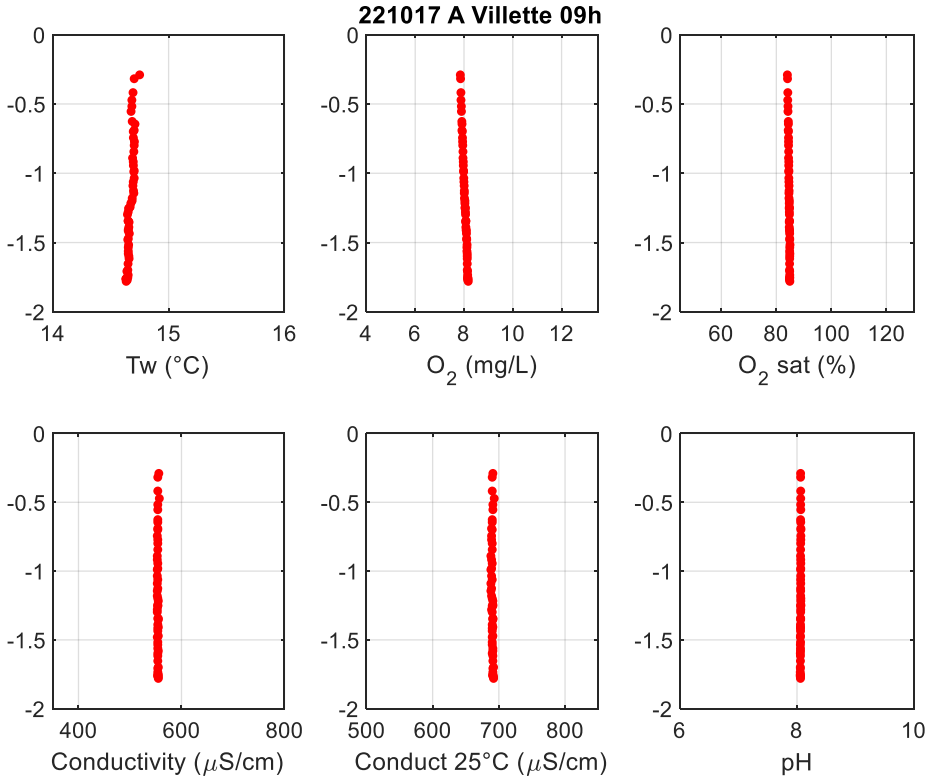


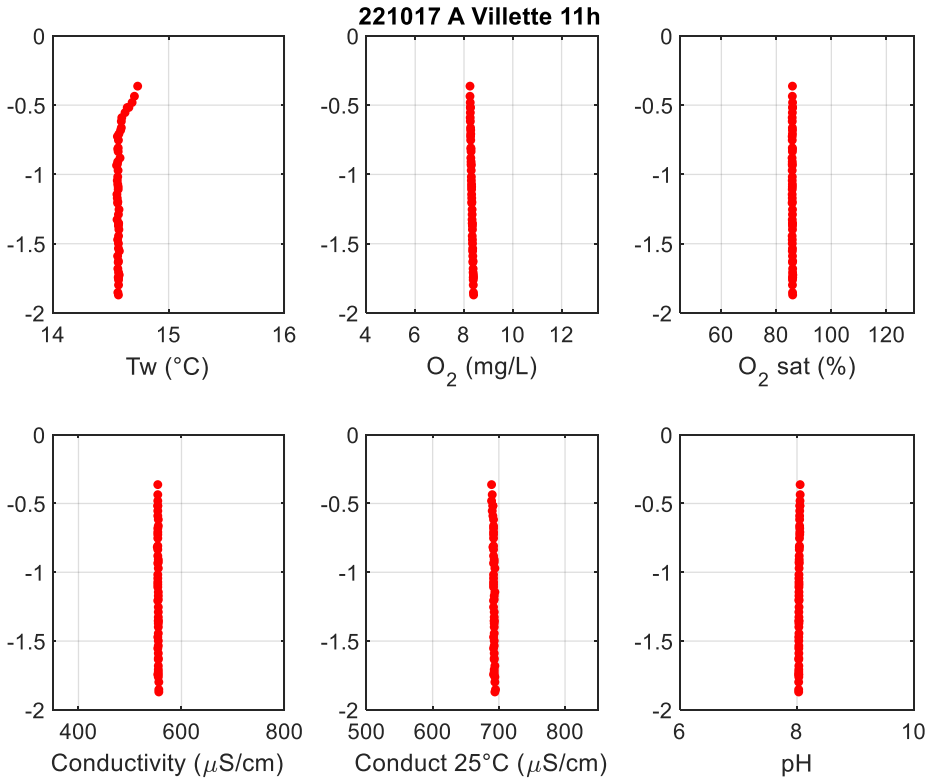
Point B – 16:56 (UTC)

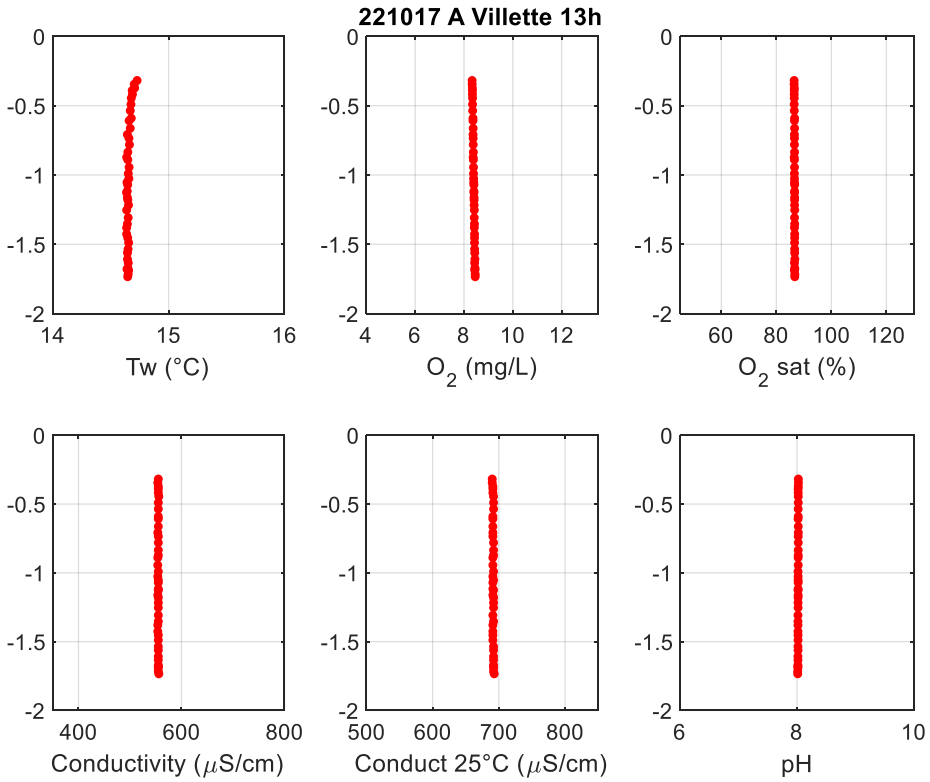


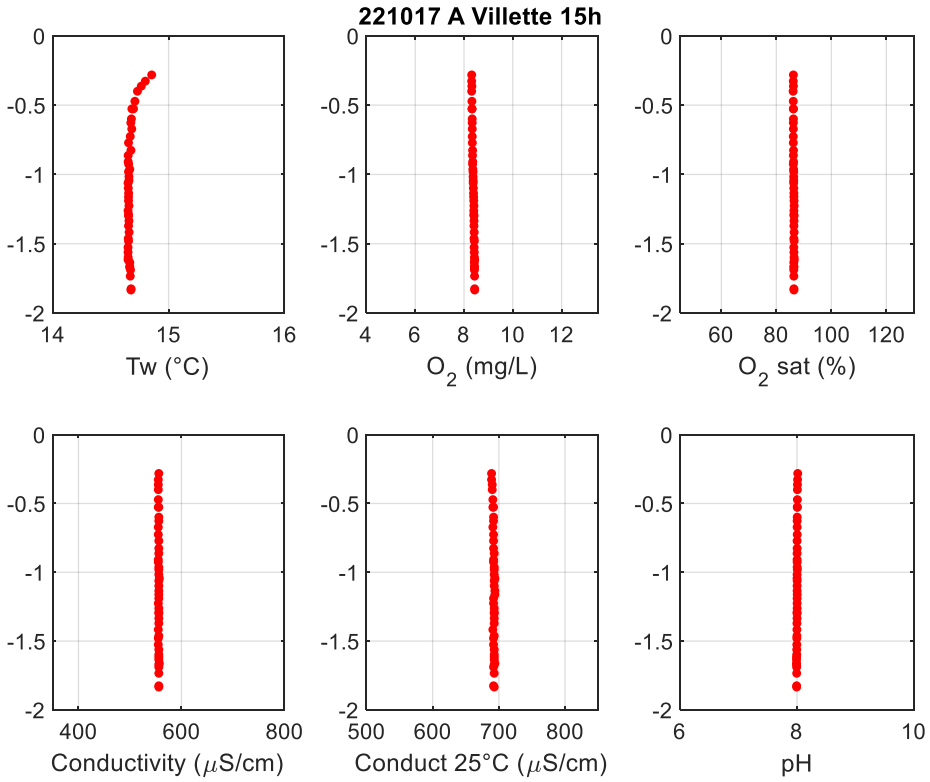
17/10/2022

Point A

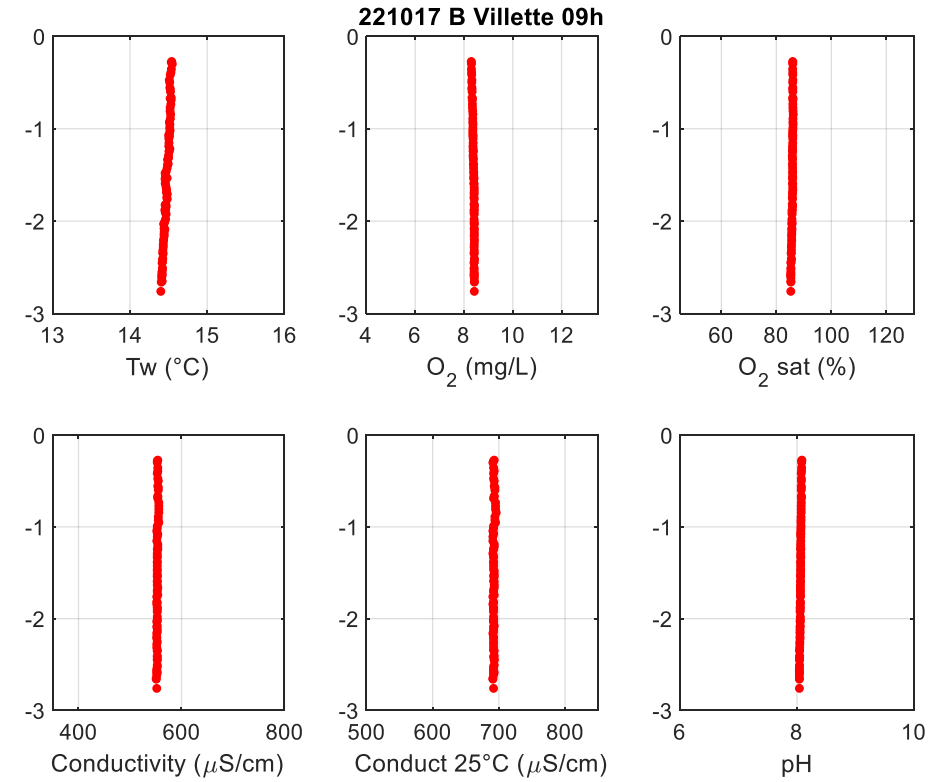


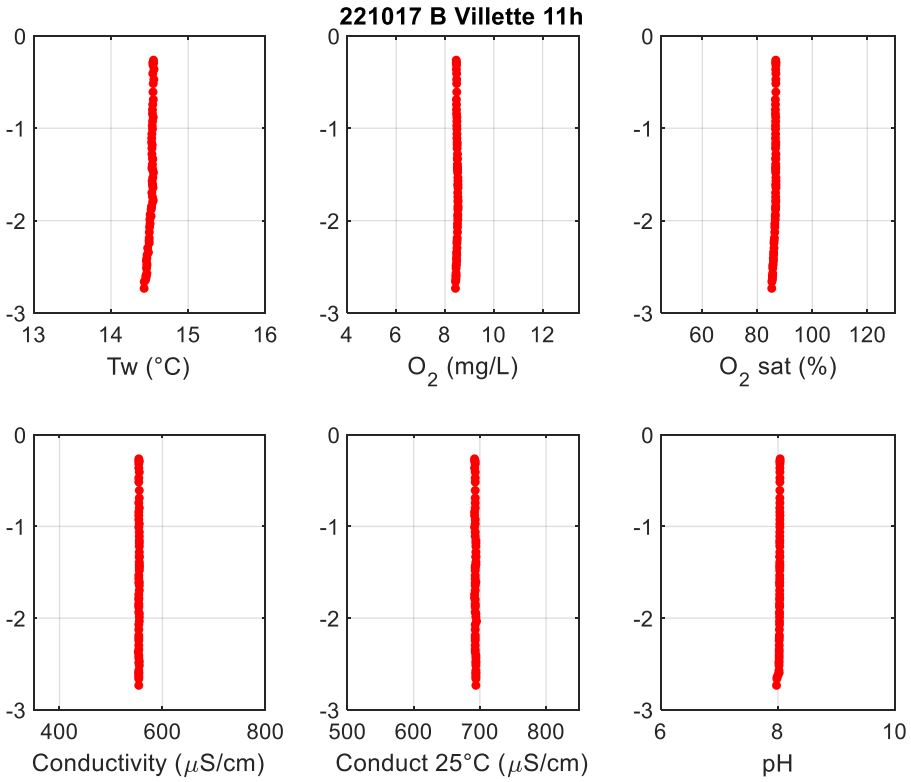

221017 A Villette 11h
0
-0.5
-1
-1.5
-2
14
15
16
Tw (°C)
4
6
8
10
12
O2 (mg/L)
60
80
100
120
O2 sat (%)
400
600
800
Conductivity (μS/cm)
500
600
700
800
Conduct 25°C (μS/cm)
6
8
10
pH

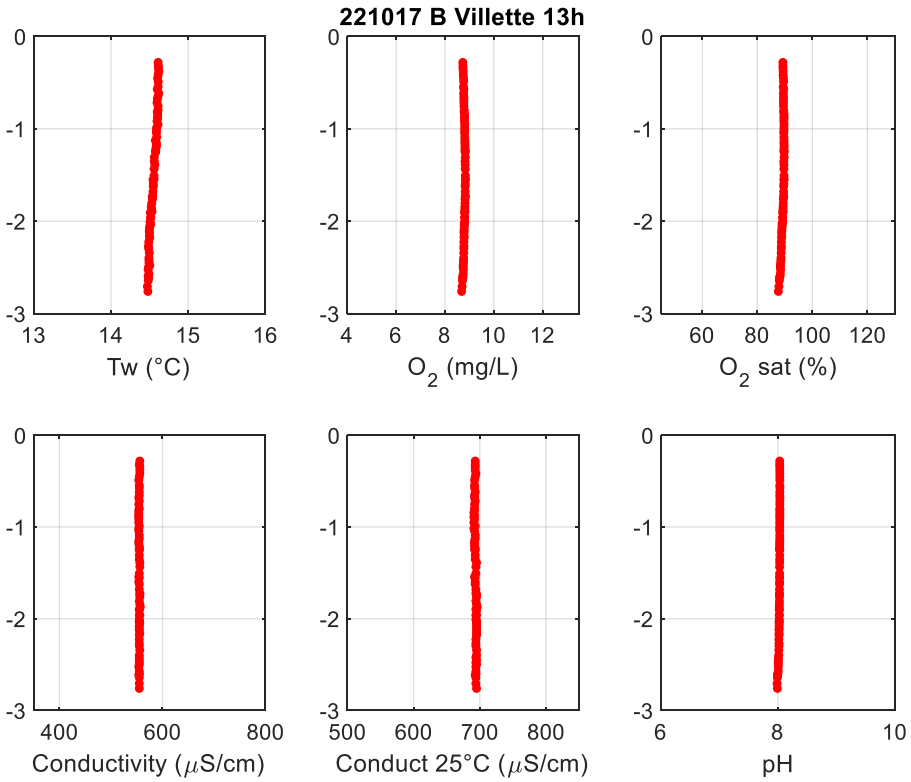

221017 A Villette 13h
0
-0.5
-1
-1.5
-2
14
15
16
Tw (°C)
4
6
8
10
12
O2 (mg/L)
60
80
100
120
O2 sat (%)
400
600
800
Conductivity (μS/cm)
500
600
700
800
Conduct 25°C (μS/cm)
6
8
10
pH

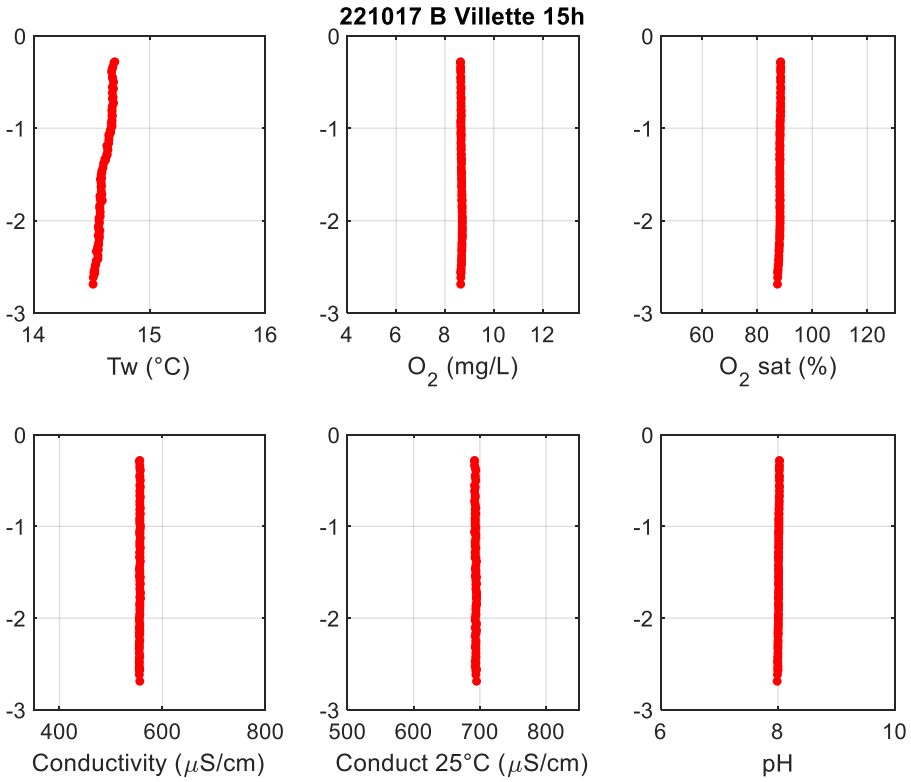

221017 A Villette 15h
Tw (°C)
O2 (mg/L)
O2 sat (%)
Conductivity (µS/cm)
Conduct 25°C (µS/cm)
pH

Point B


221017 B Villette 09h
Tw (°C)
O2 (mg/L)
O2 sat (%)
Conductivity (µS/cm)
Conduct 25°C (µS/cm)
pH

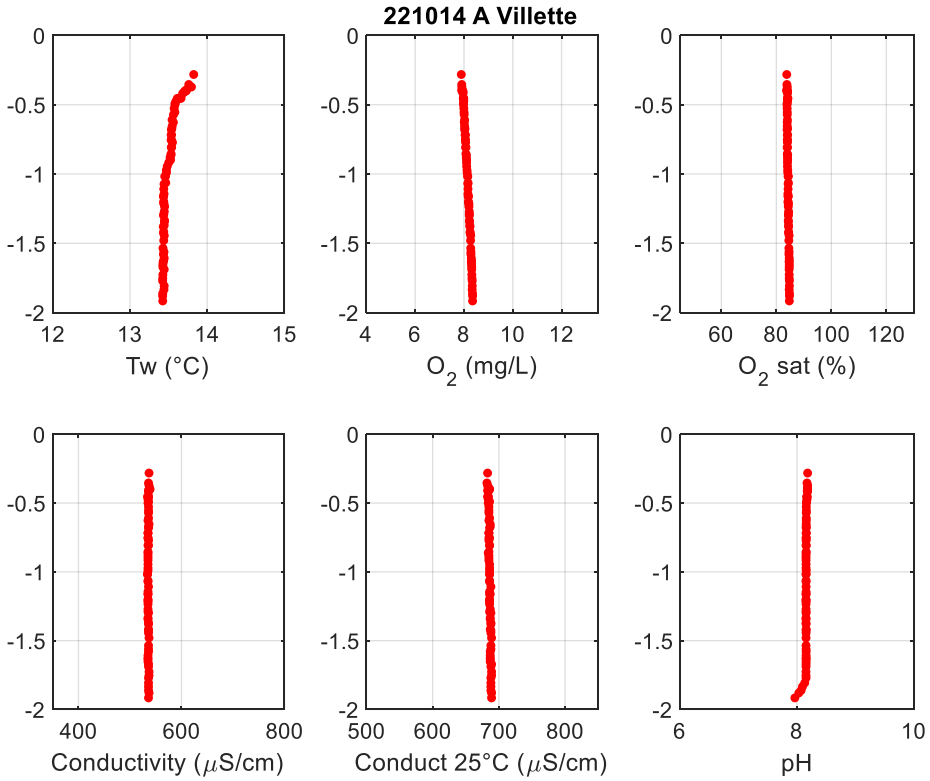

221017 B Villette 11h
0
-1
-2
-3
13
14
15
16
Tw (°C)
4
6
8
10
12
O2 (mg/L)
60
80
100
120
O2 sat (%)
400
600
800
Conductivity (μS/cm)
500
600
700
800
Conduct 25°C (μS/cm)
6
8
10
pH


221017 B Villette 13h
0
-1
-2
-3
13
14
15
16
Tw (°C)
4
6
8
10
12
O2 (mg/L)
60
80
100
120
O2 sat (%)
400
600
800
Conductivity (μS/cm)
500
600
700
800
Conduct 25°C (μS/cm)
6
8
10
pH

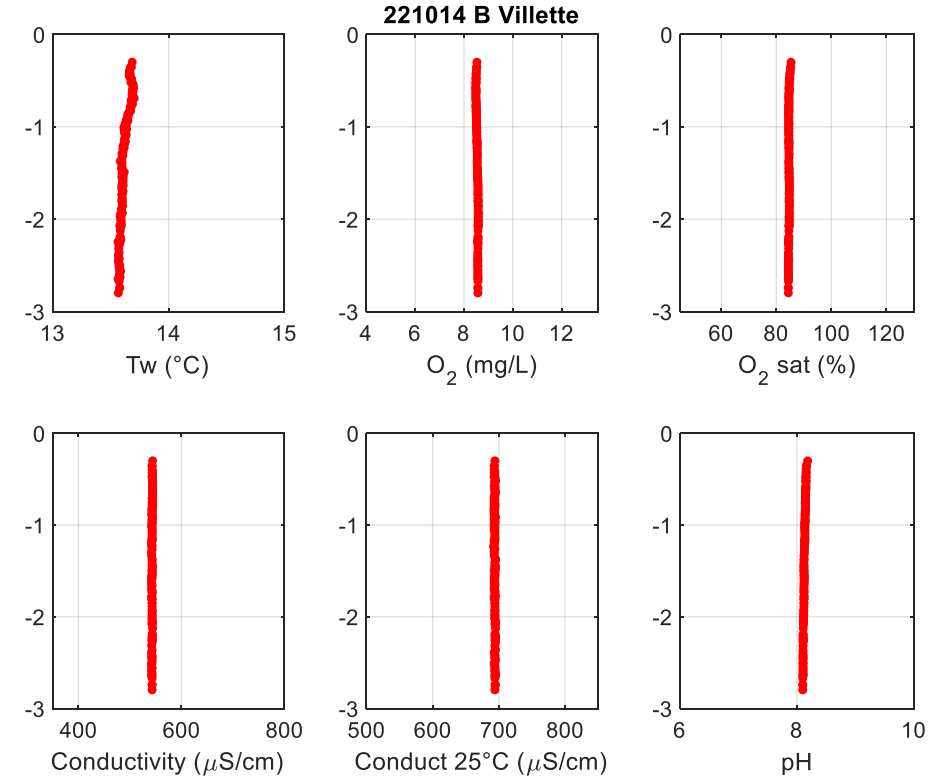

221017 B Villette 15h
0
-1
-2
-3
14
15
16
Tw (°C)
4
6
8
10
12
O2 (mg/L)
60
80
100
120
O2 sat (%)
400
600
800
Conductivity (μS/cm)
500
600
700
800
Conduct 25°C (μS/cm)
6
8
10
pH

14/10/2022

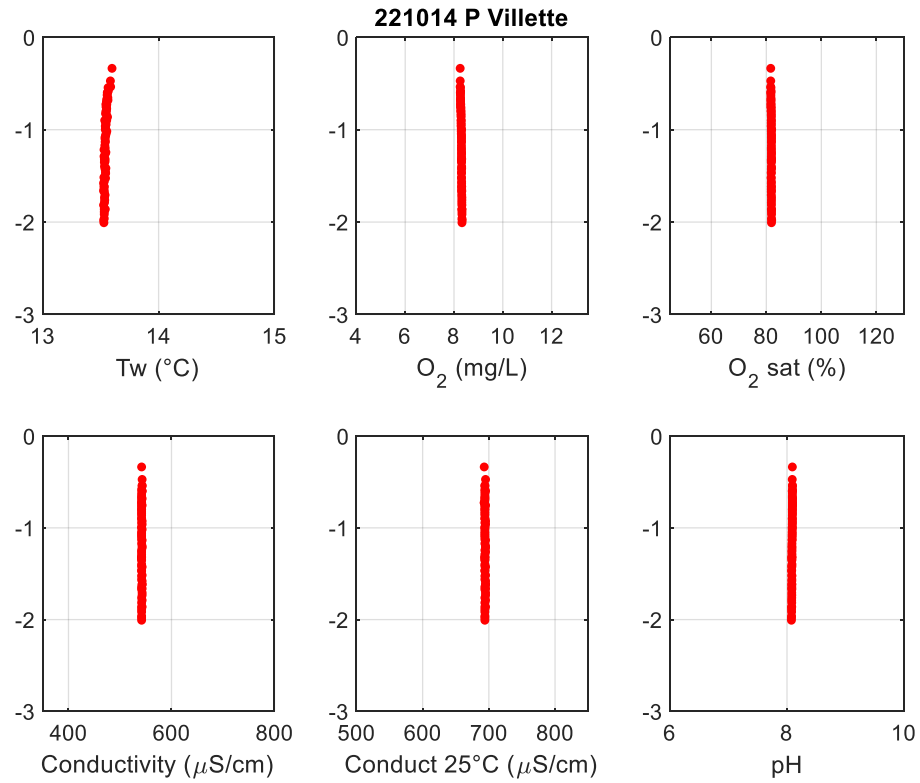
Point A – 07:45 (UTC)



Point B – 08:20 (UTC)

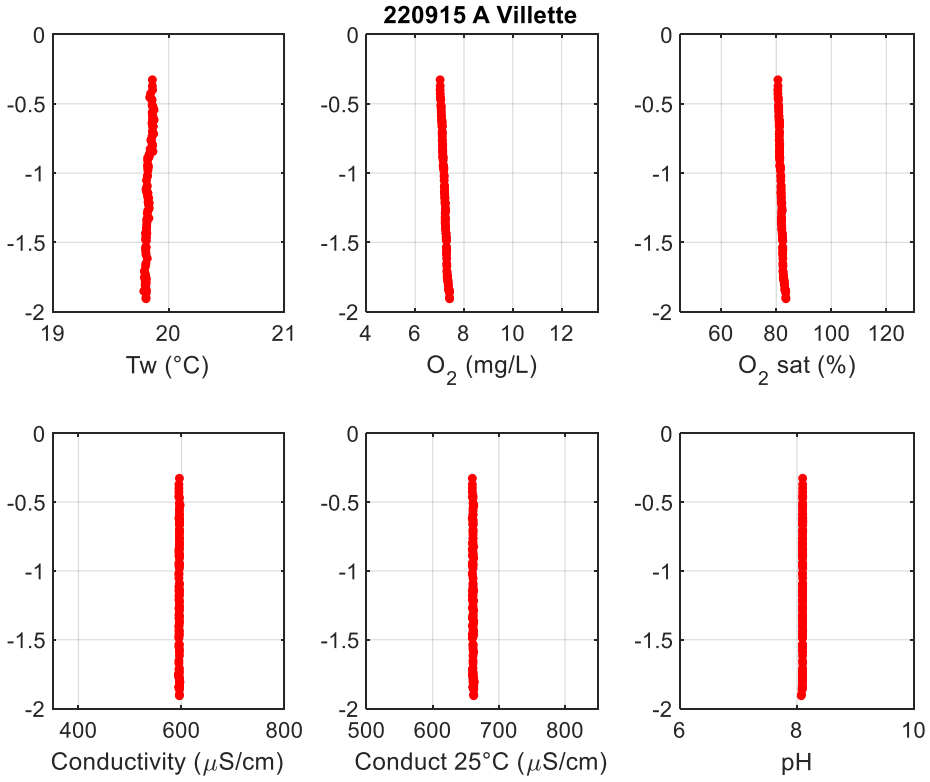


Point P – 08:20 (UTC) – Station de pompage

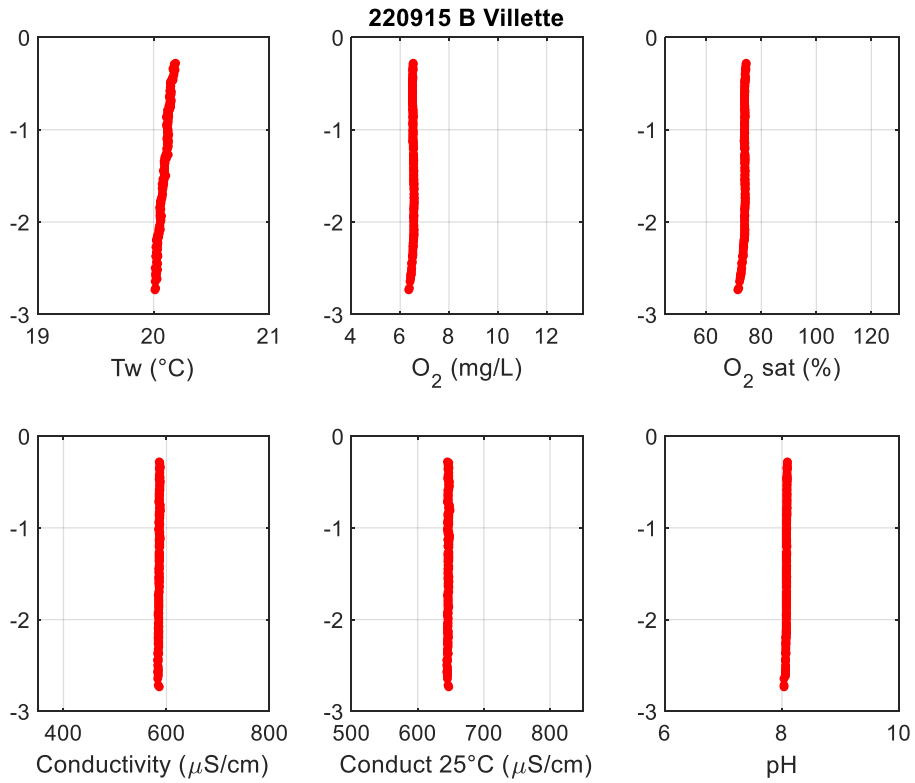


15/09/2022 – Thermal stratification (B)?

Point A – 07:36 (UTC)

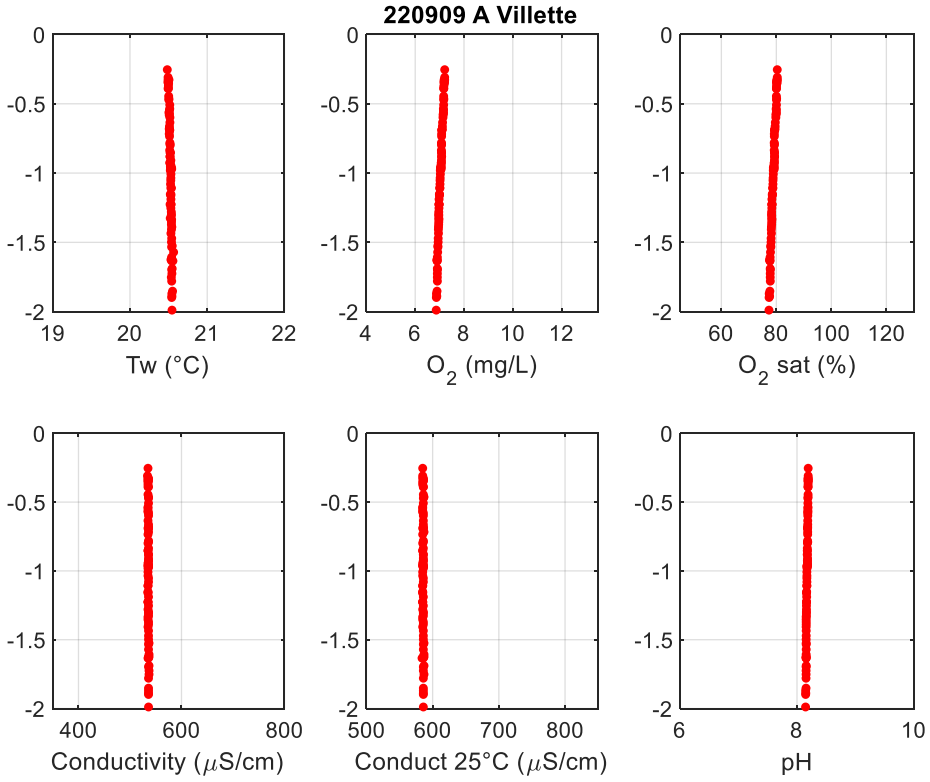


Point B – 08:45 (UTC)

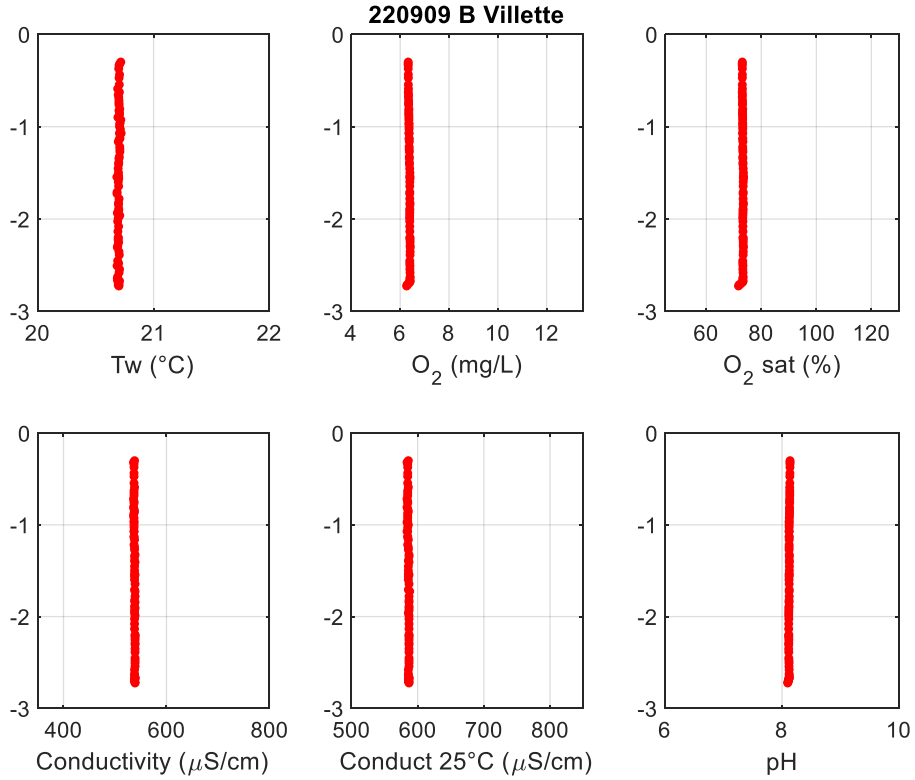


09/09/2022

Point A – 08:35 (UTC)

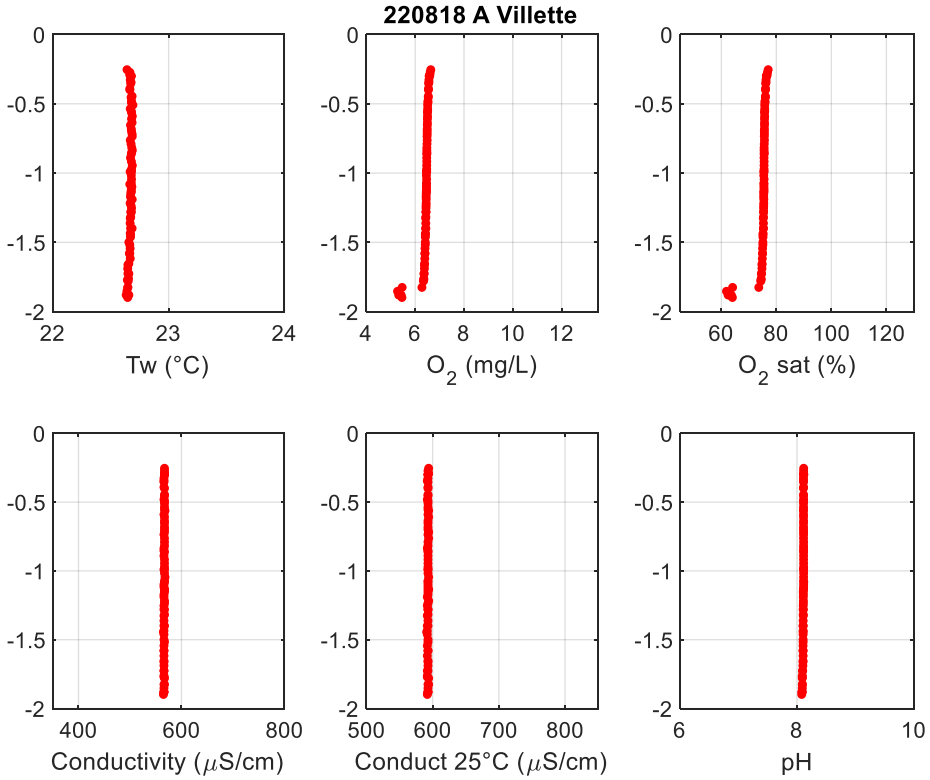


Point B – 07:55 (UTC)

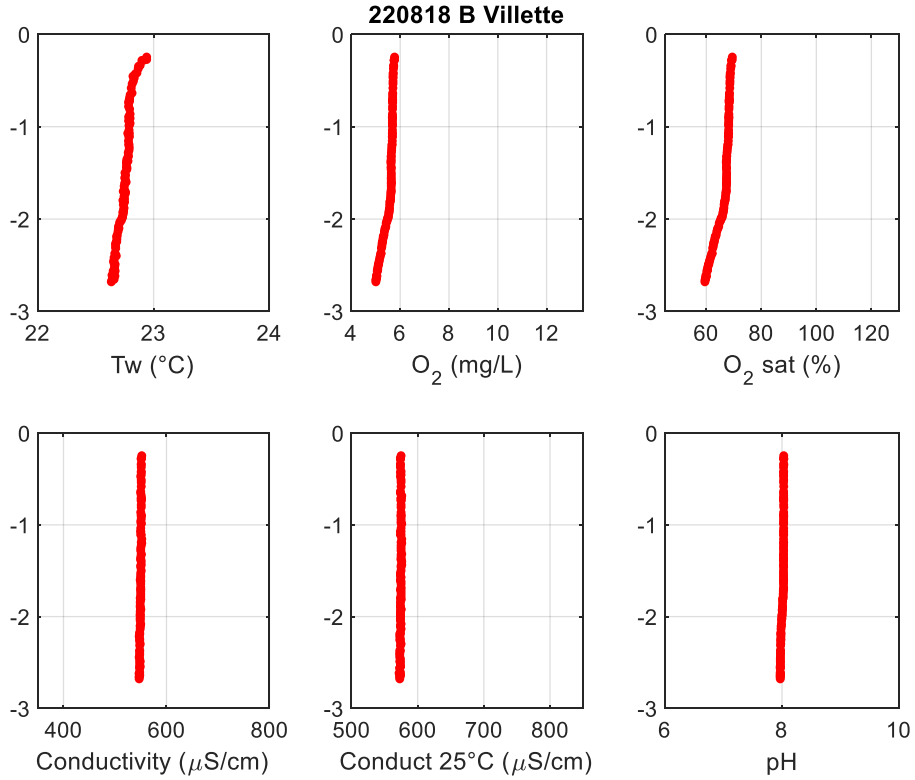


18/08/2022

Point A – 08:28 (UTC)

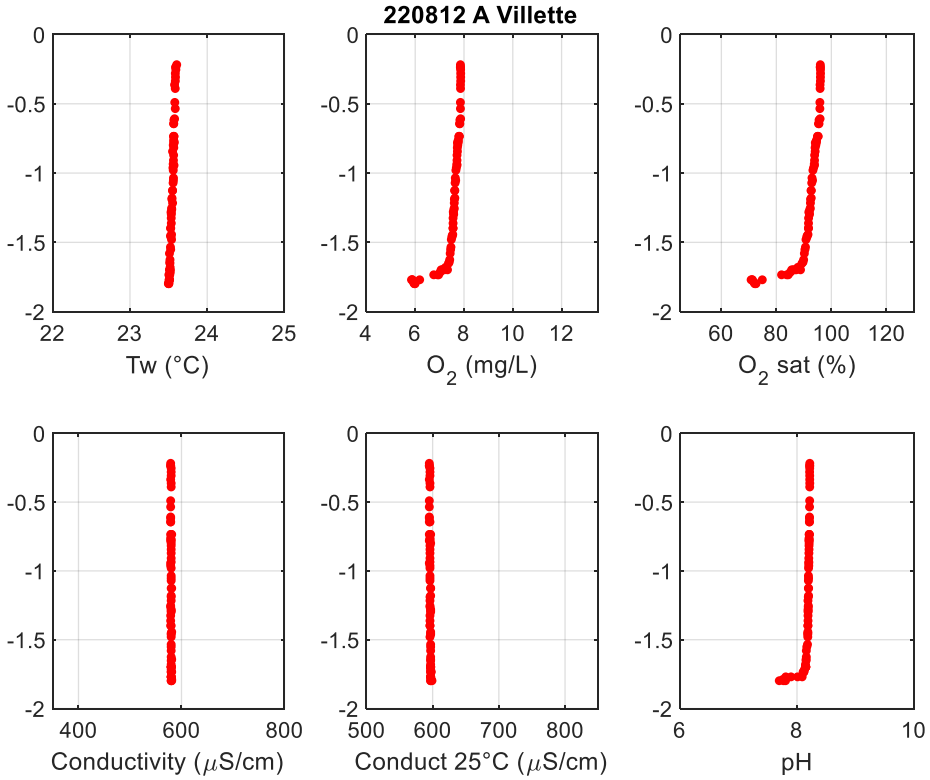


Point B – 09:08 (UTC)

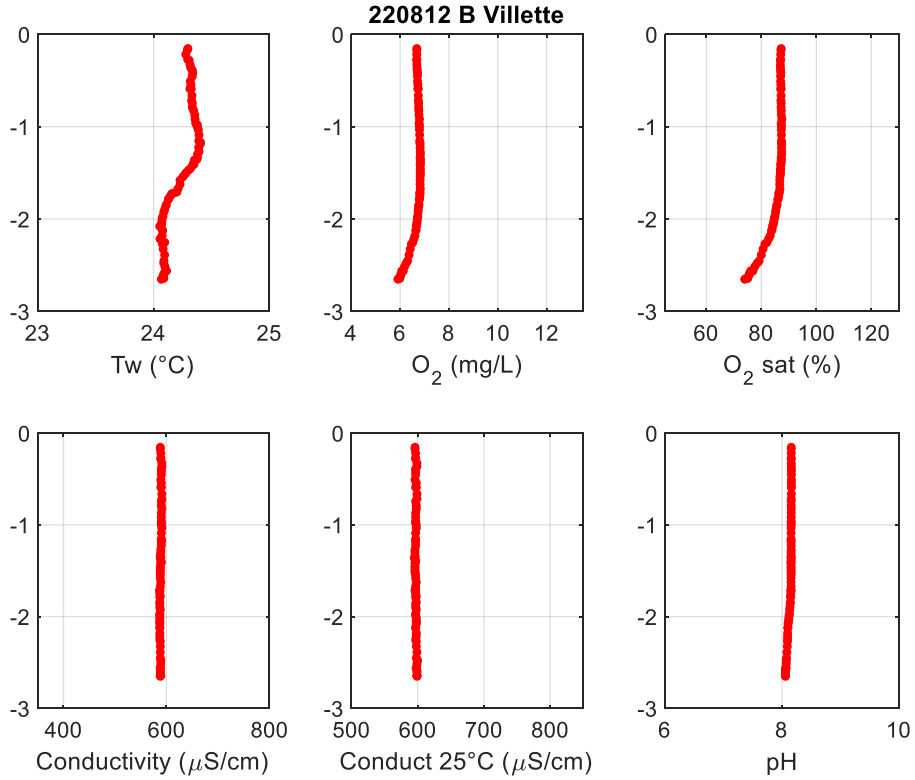


12/08/2022 – Thermal stratification (B)

Point A – 07:50 (UTC)

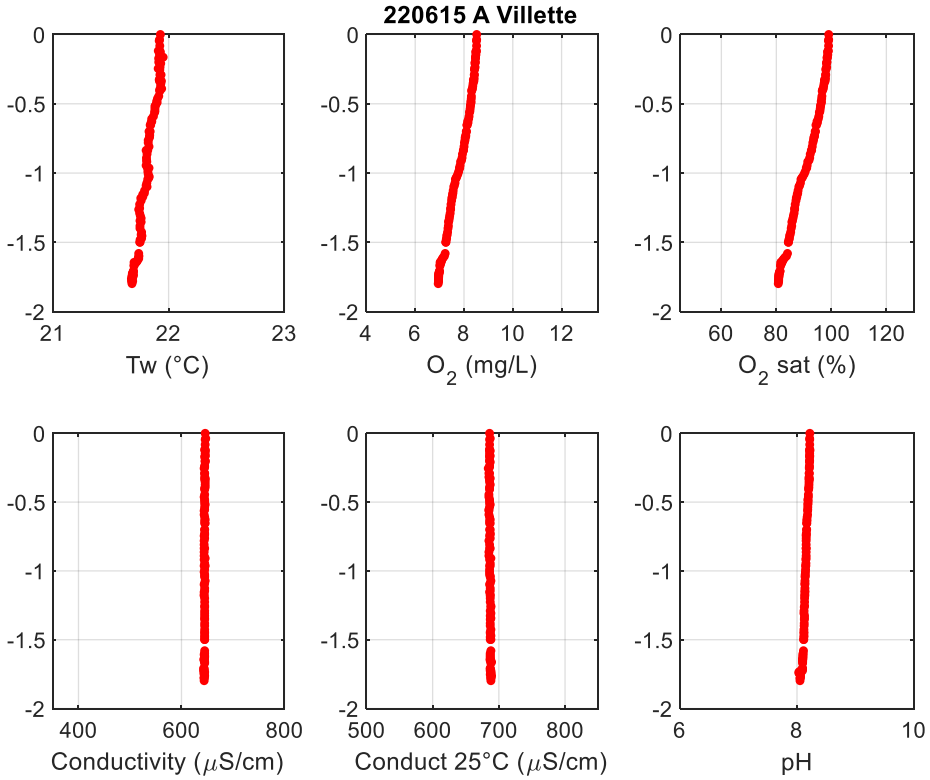


Point B – 10:30 (UTC)

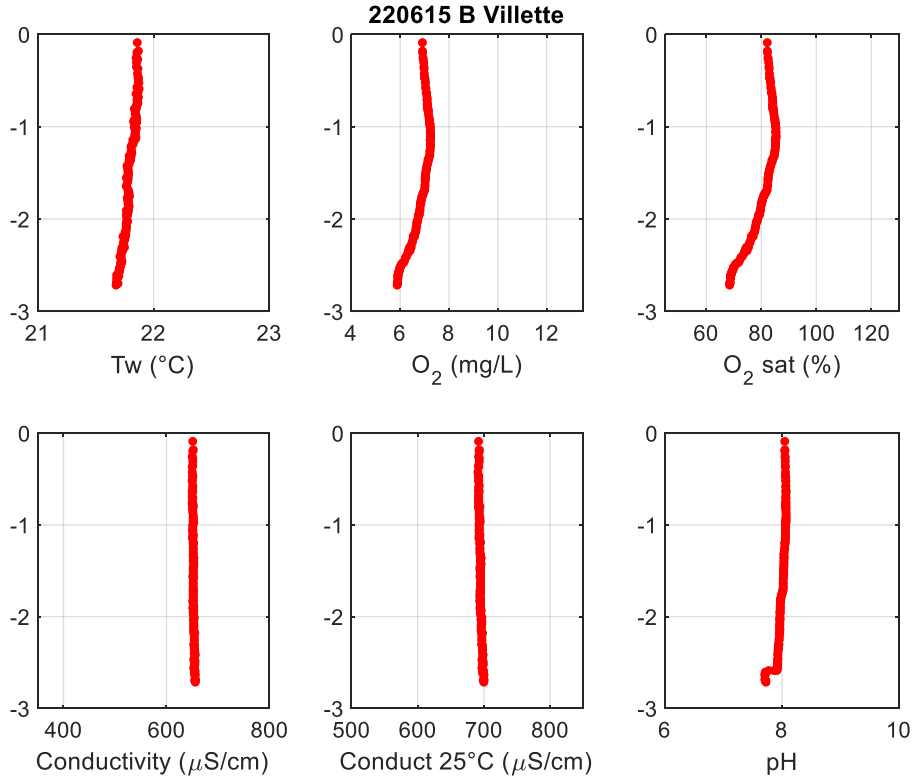


15/06/2022

Point A – 07:25 (UTC)

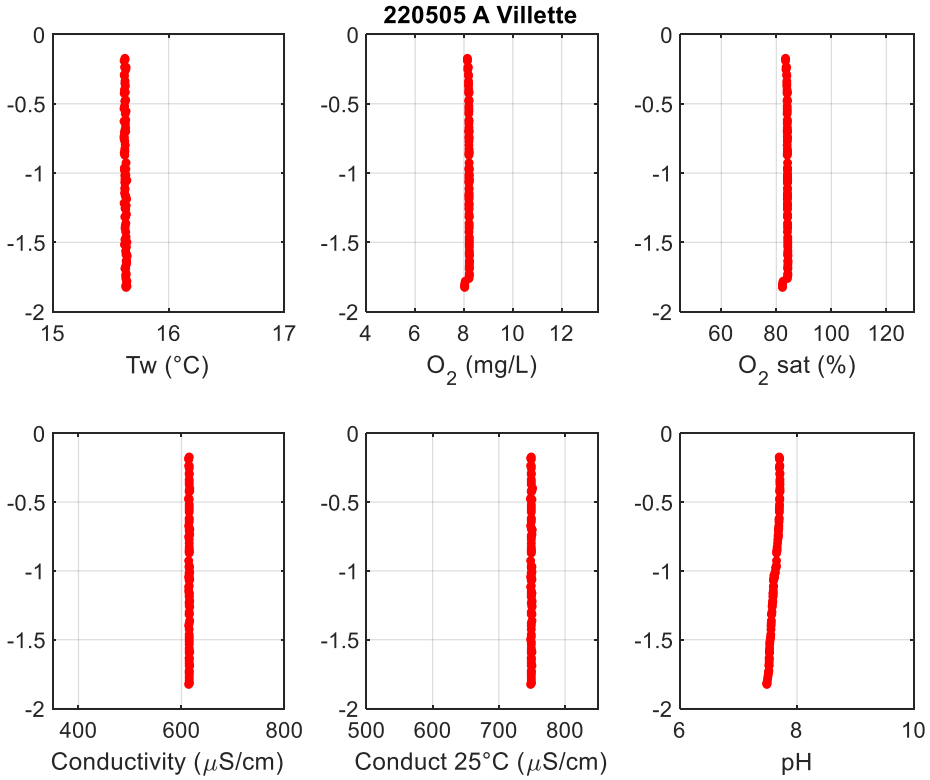


Point B – 08:30 (UTC)

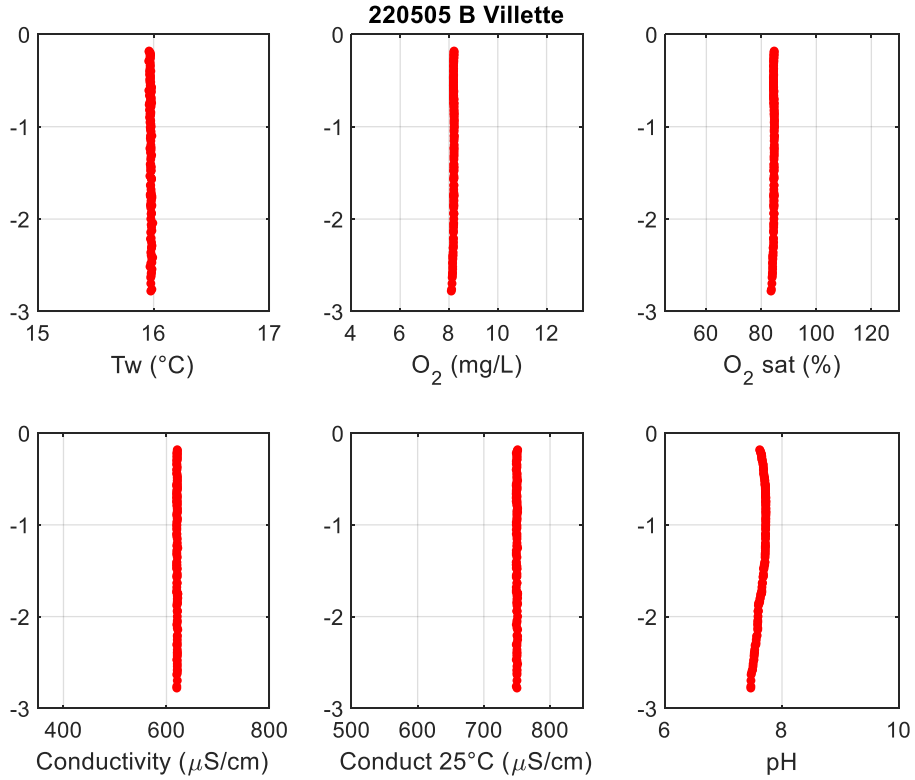


05/05/2022

Point A – 07:40 (UTC)

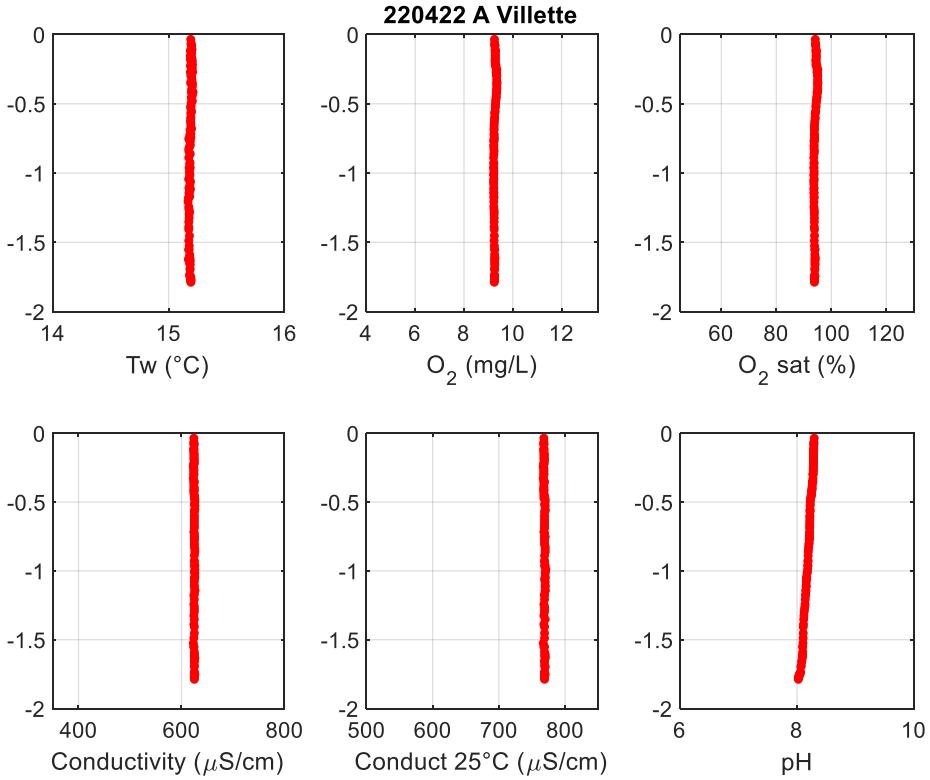


Point B – 08:27 (UTC)

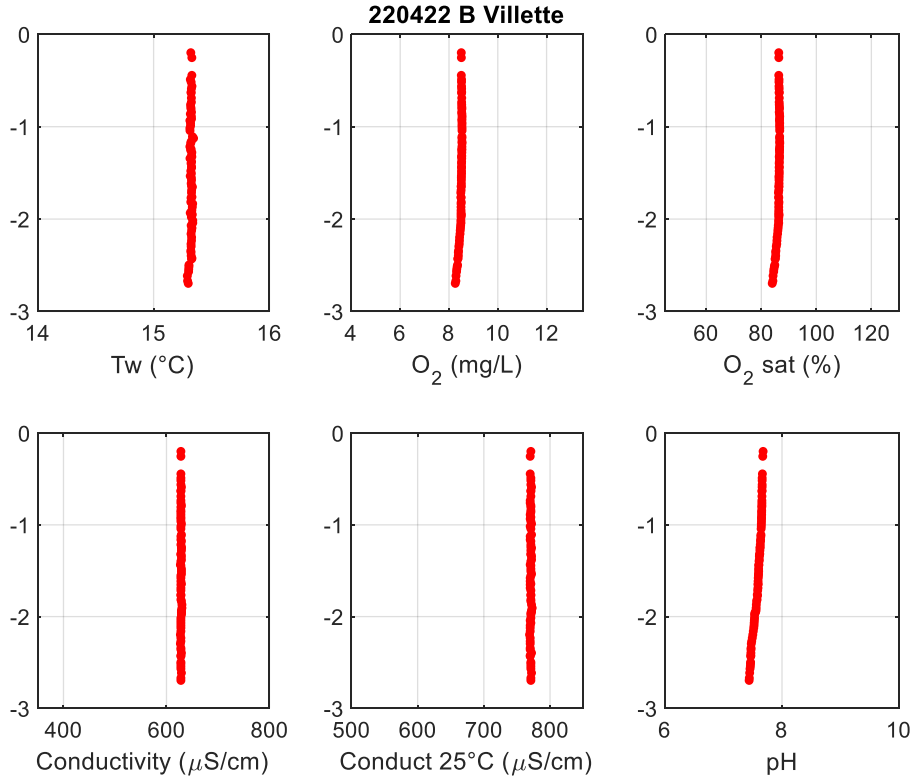


22/04/2022 – return of pH sensor (maintenance nke)

Point A – 07:25 (UTC)

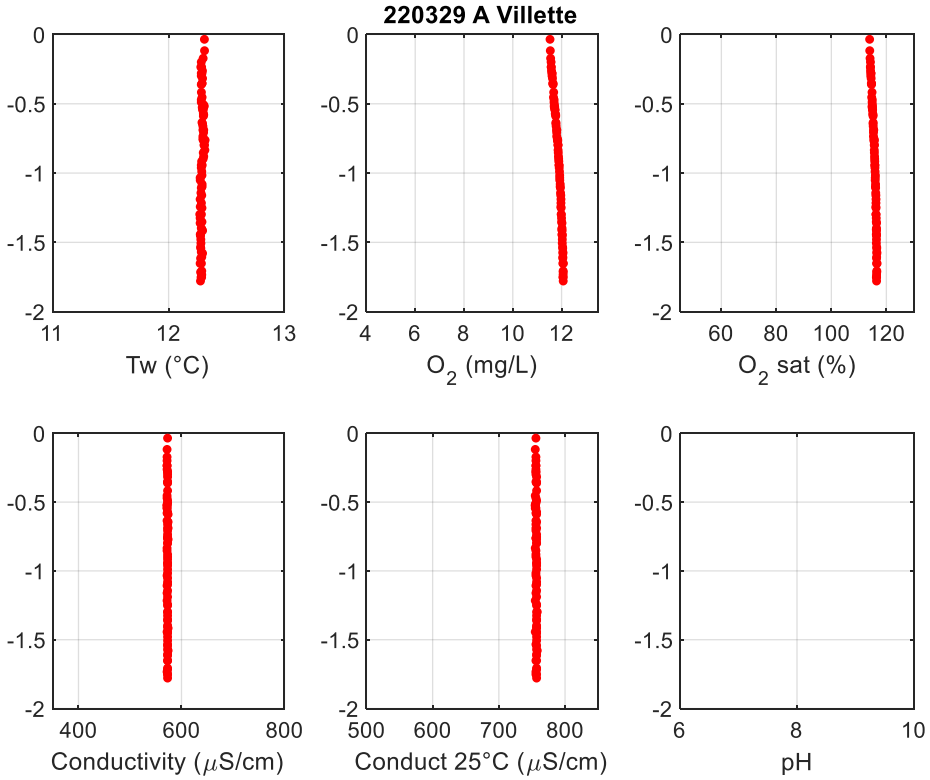


Point B – 08:07 (UTC)

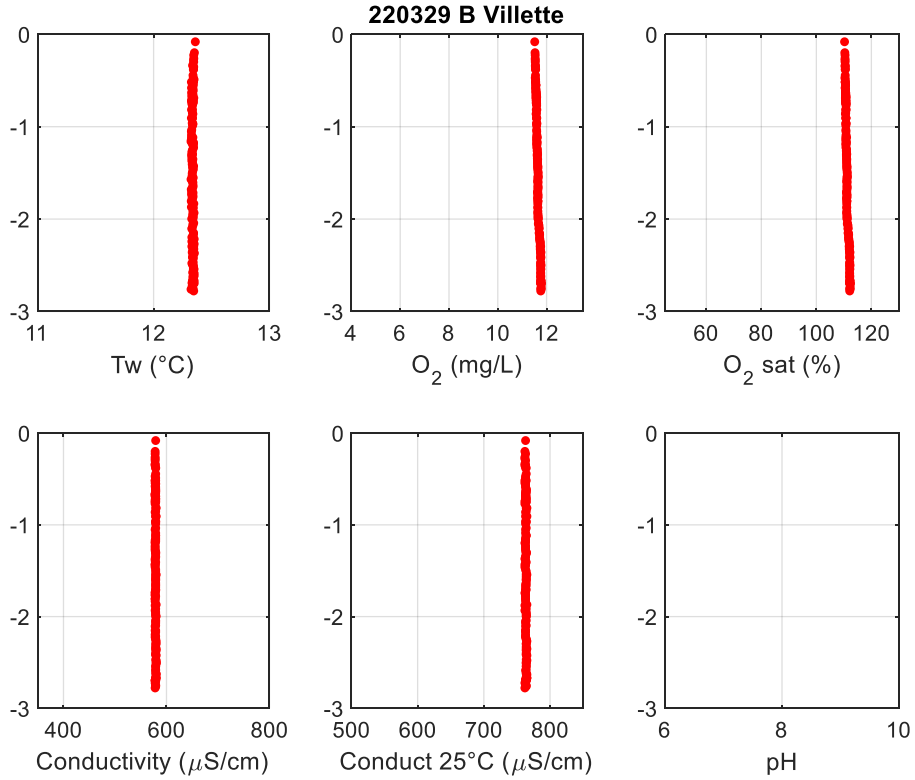


29/03/2022

Point A – 07:25 (UTC)

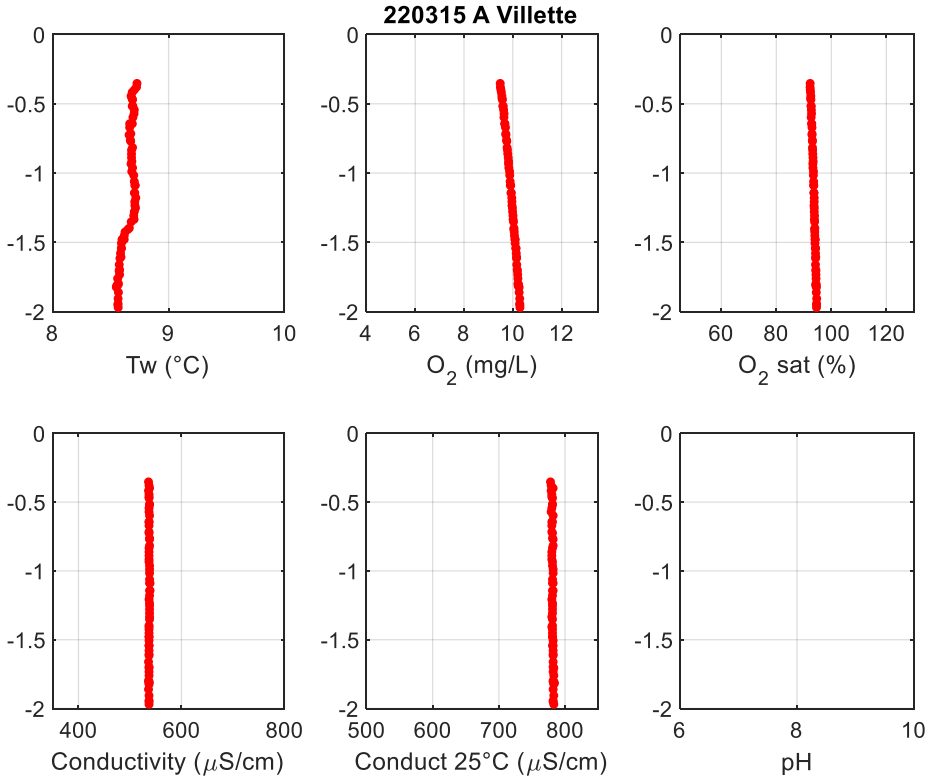


Point B – 08:33 (UTC)

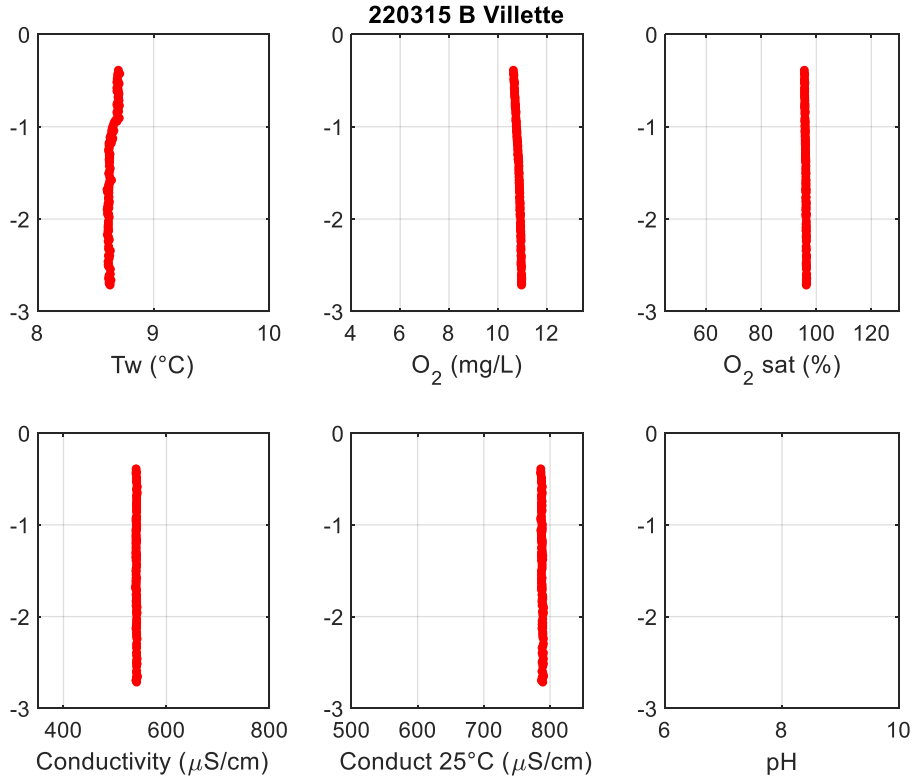


15/03/2022 – pH sensor in maintenance (nke)

Point A – 08:20 (UTC)

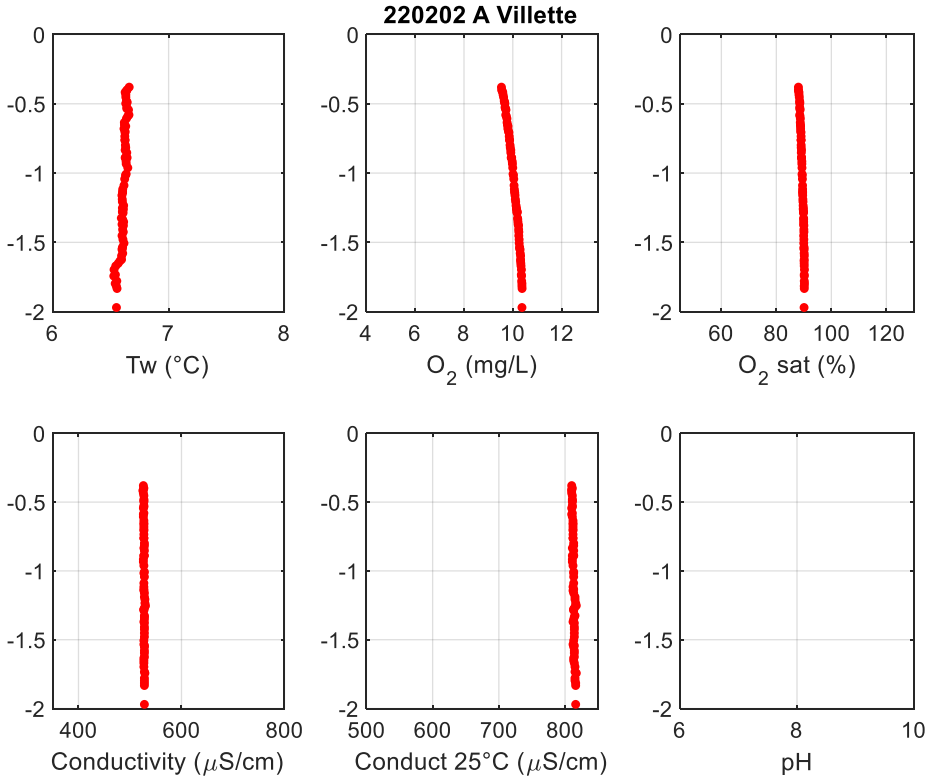


Point B – 09:00 (UTC)

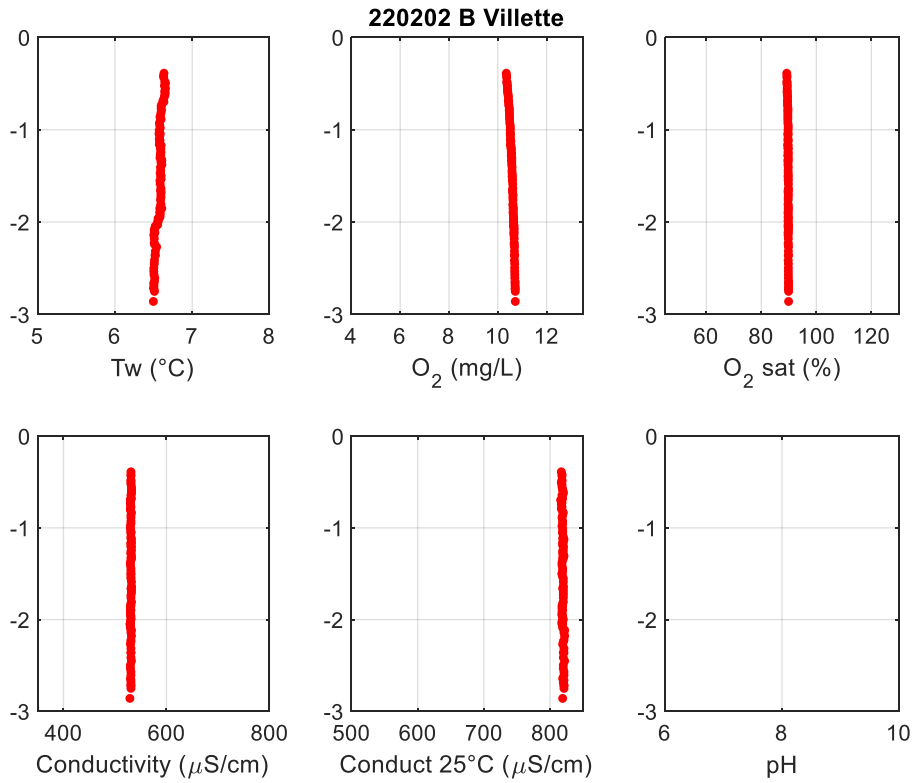


02/02/2022

Point A – 09:40 (UTC)

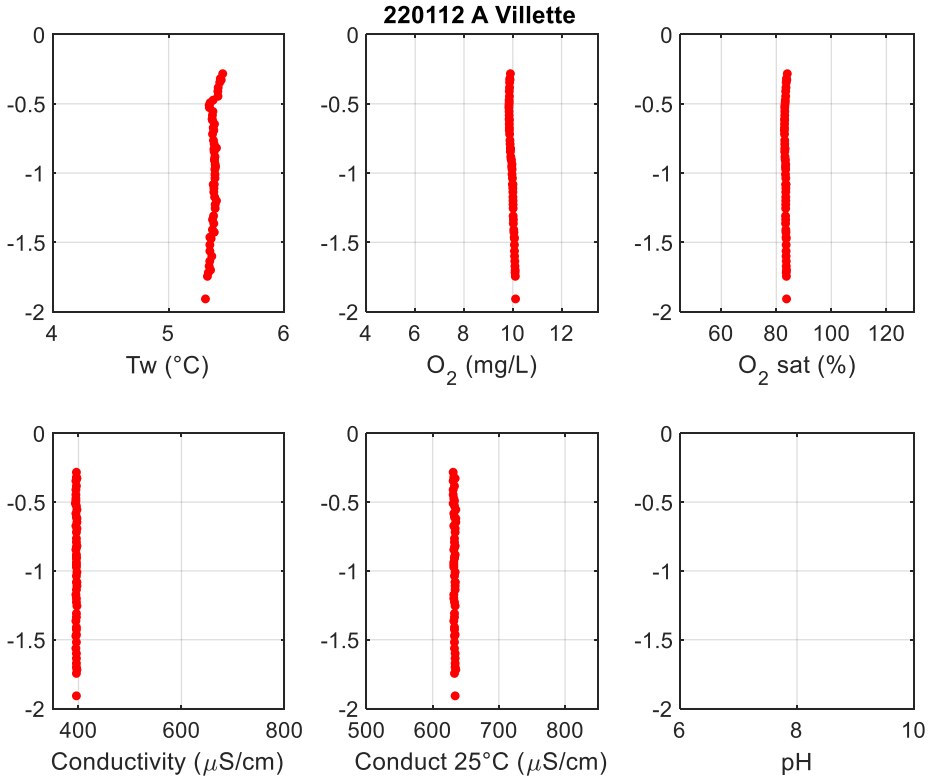


Point B – 10:10 (UTC)

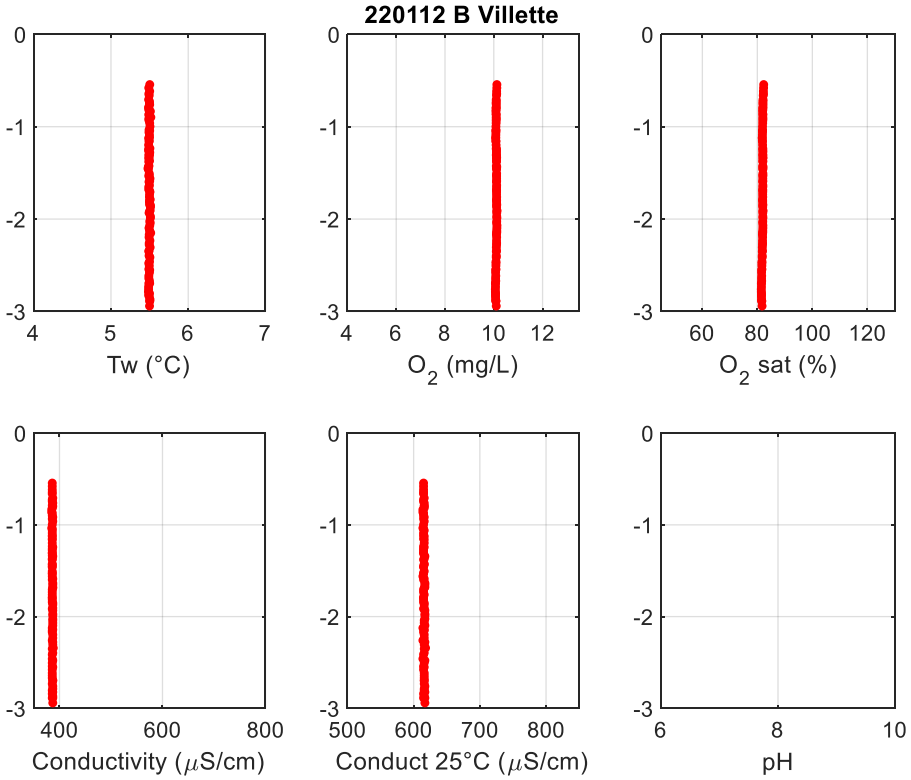


12/01/2022

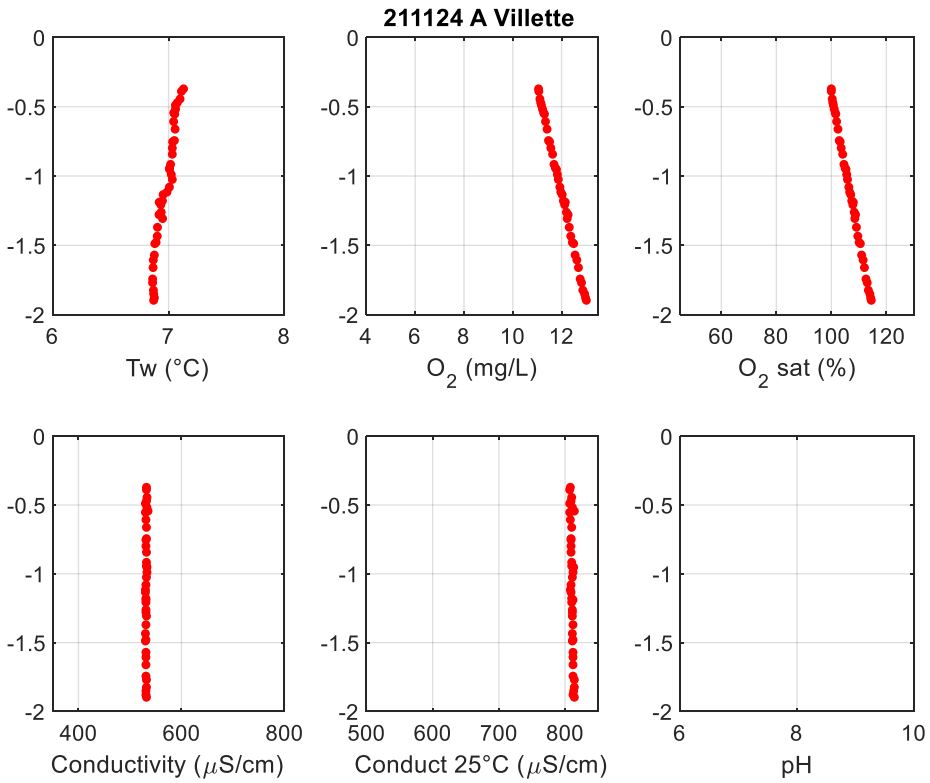
Point A – 09:45 (UTC)

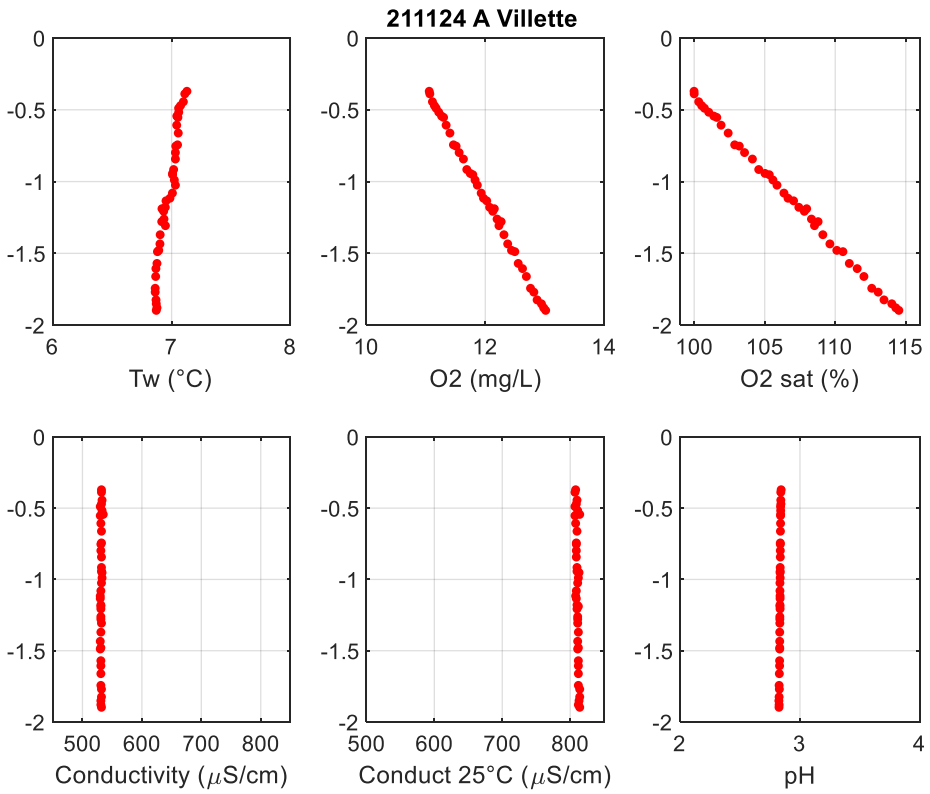


Point B – 12:20 (UTC)

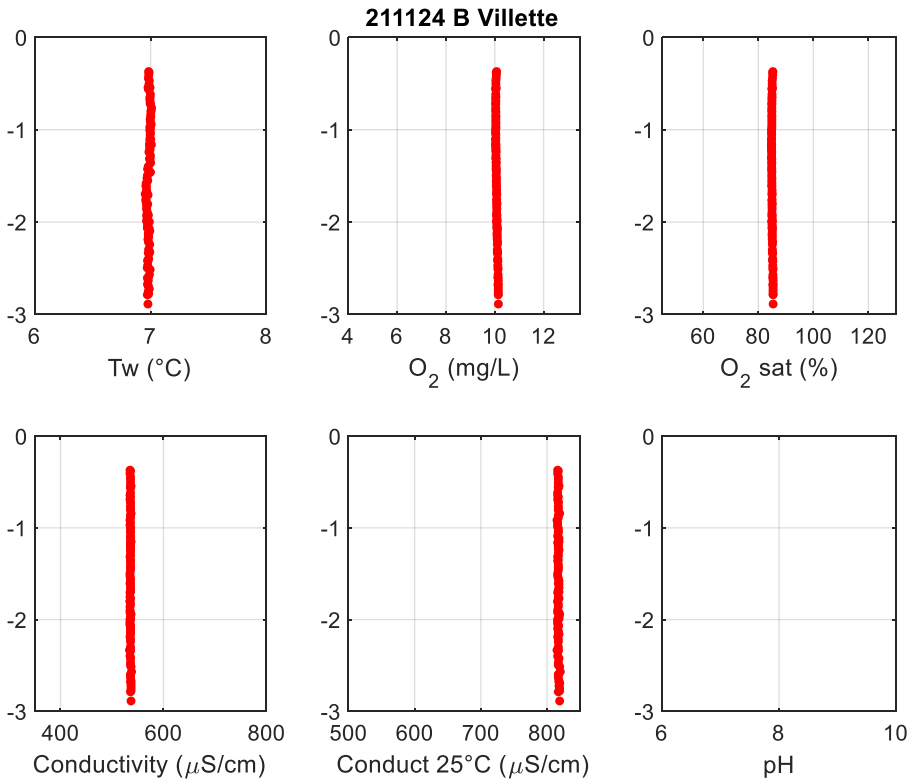


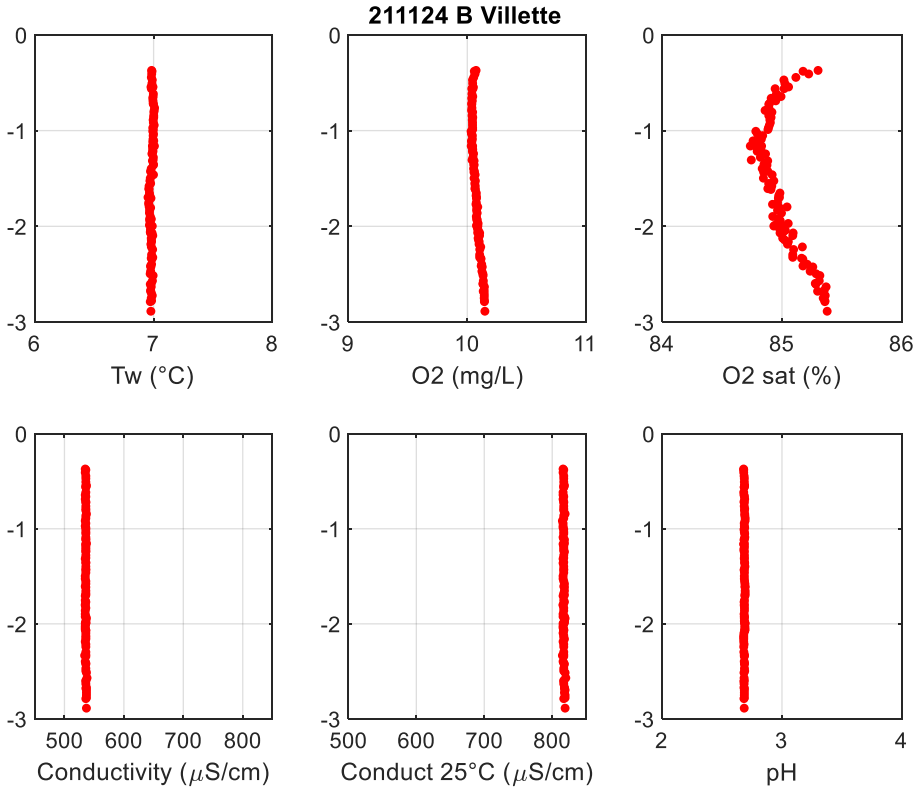
24/11/2021
Point A – 10:05 (UTC)





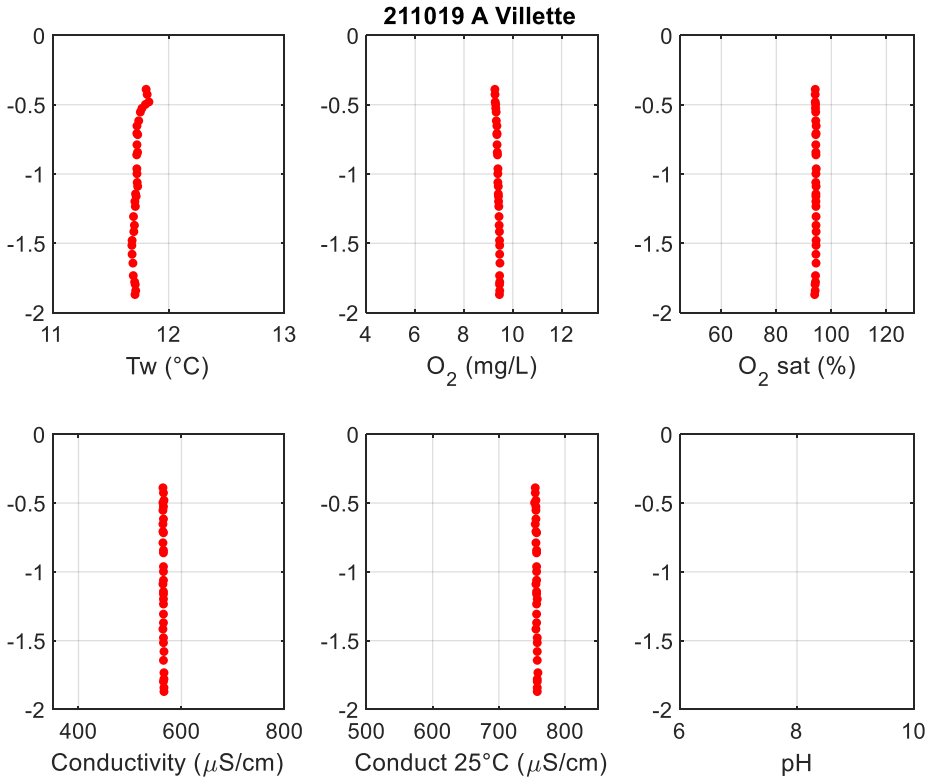
Point B – 10:38 (UTC)

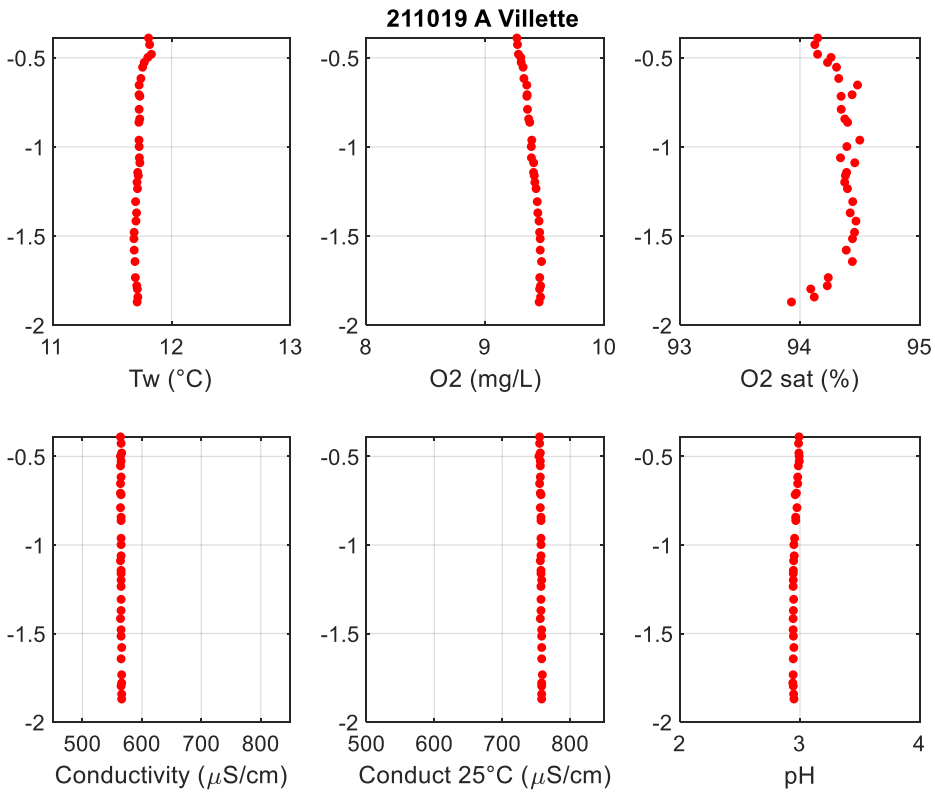




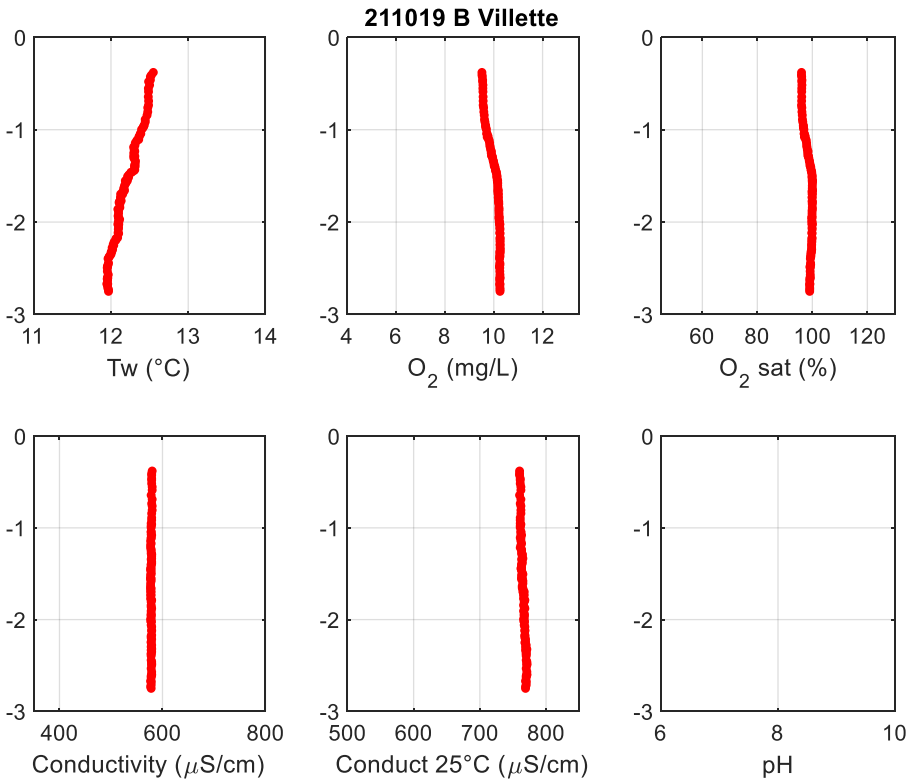
19/10/2021 – Thermal stratification (B)

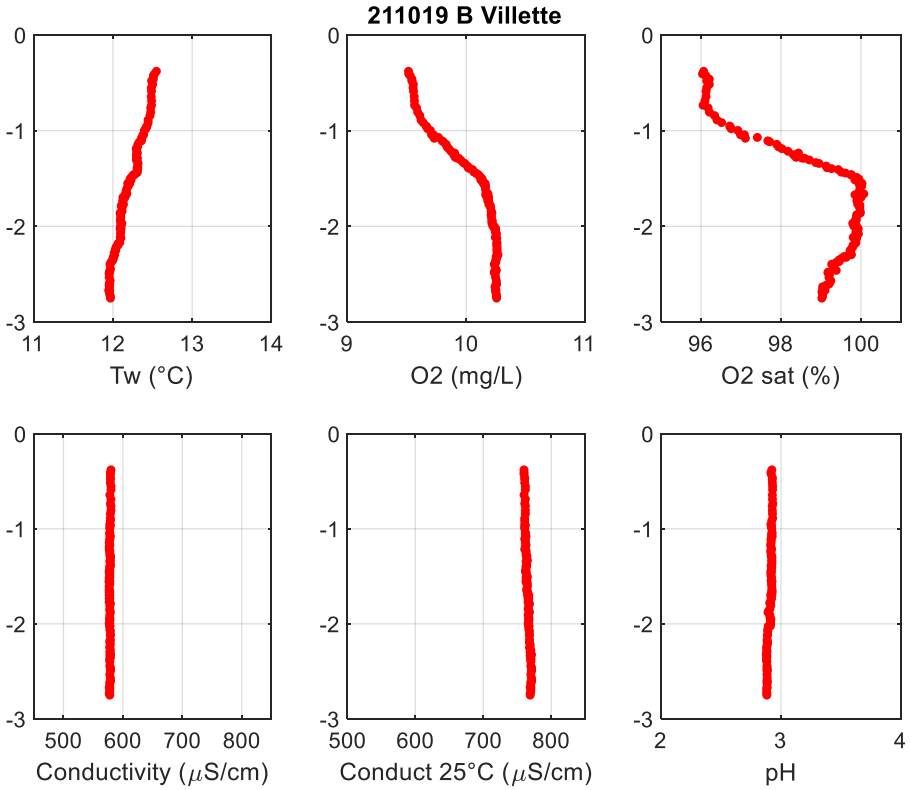
Point A – 09:27 (UTC)



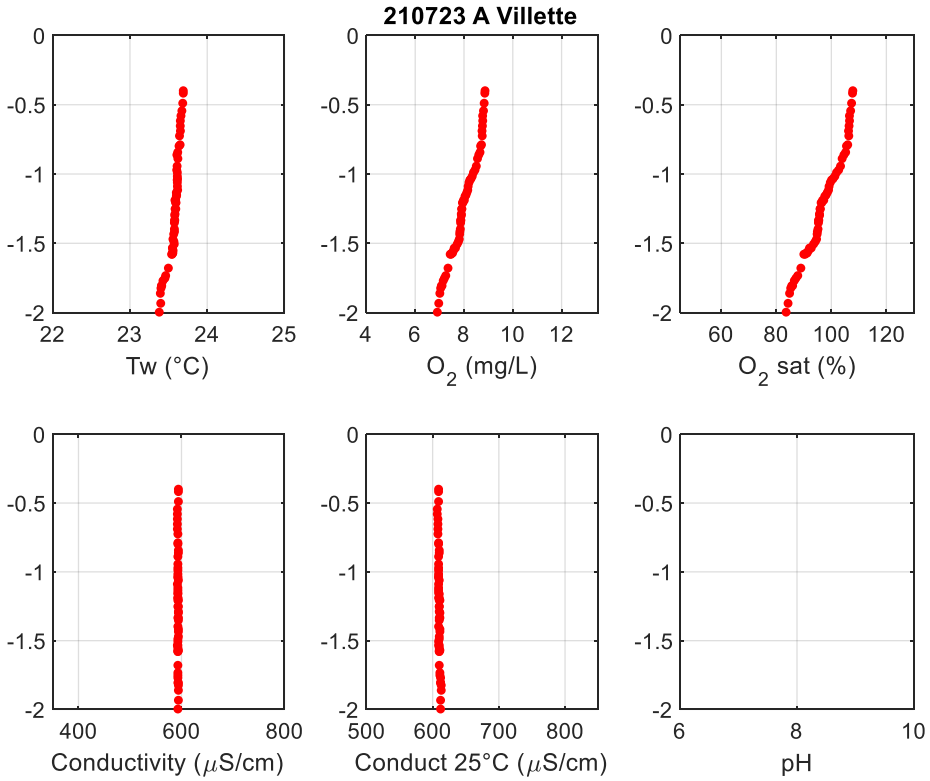


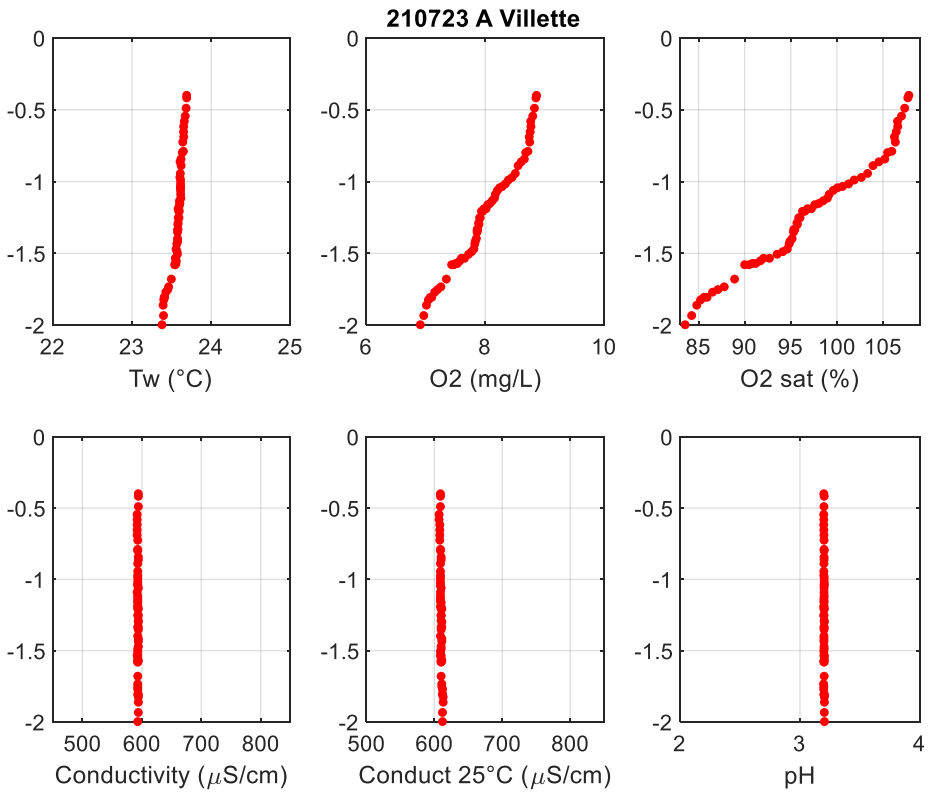
Point B – 10:28 (UTC)



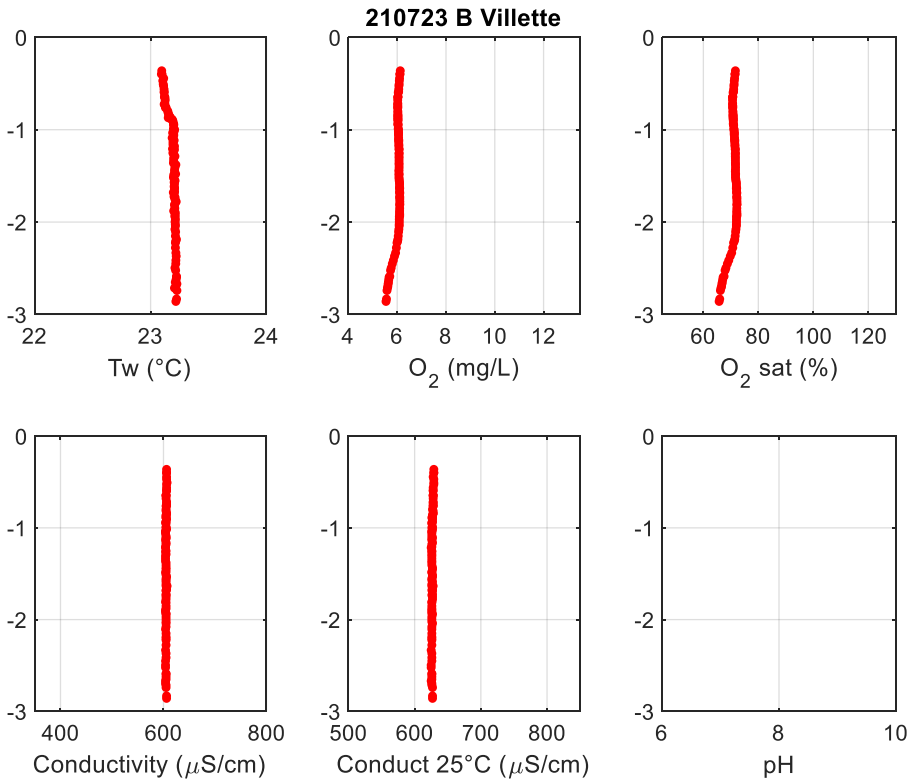


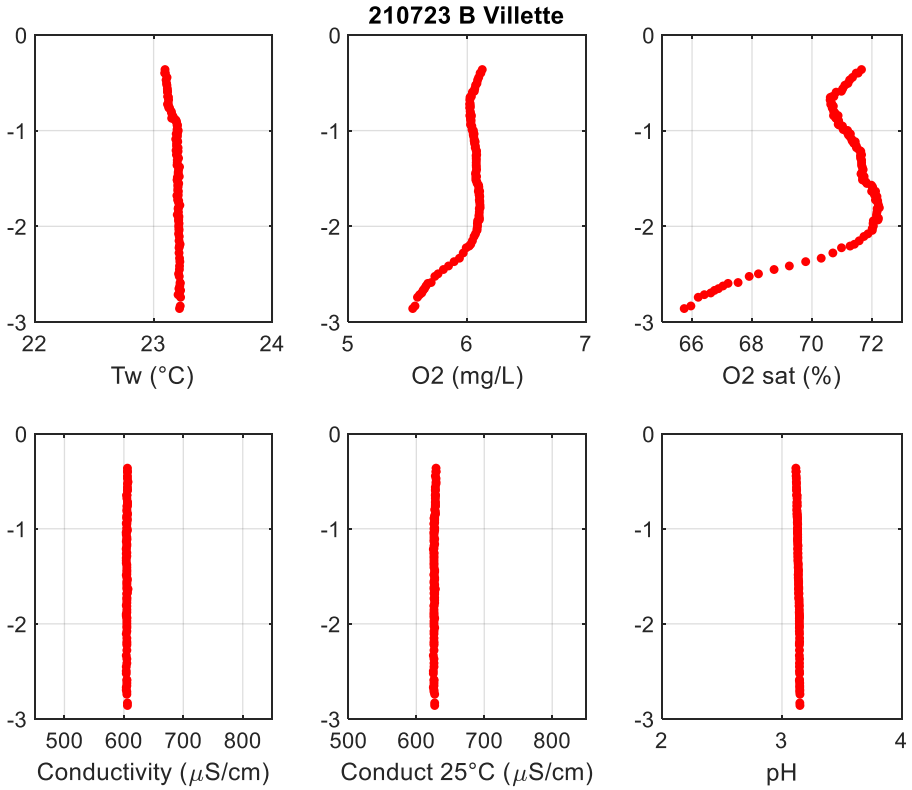
23/07/2021
Point A – 09:38 (UTC)





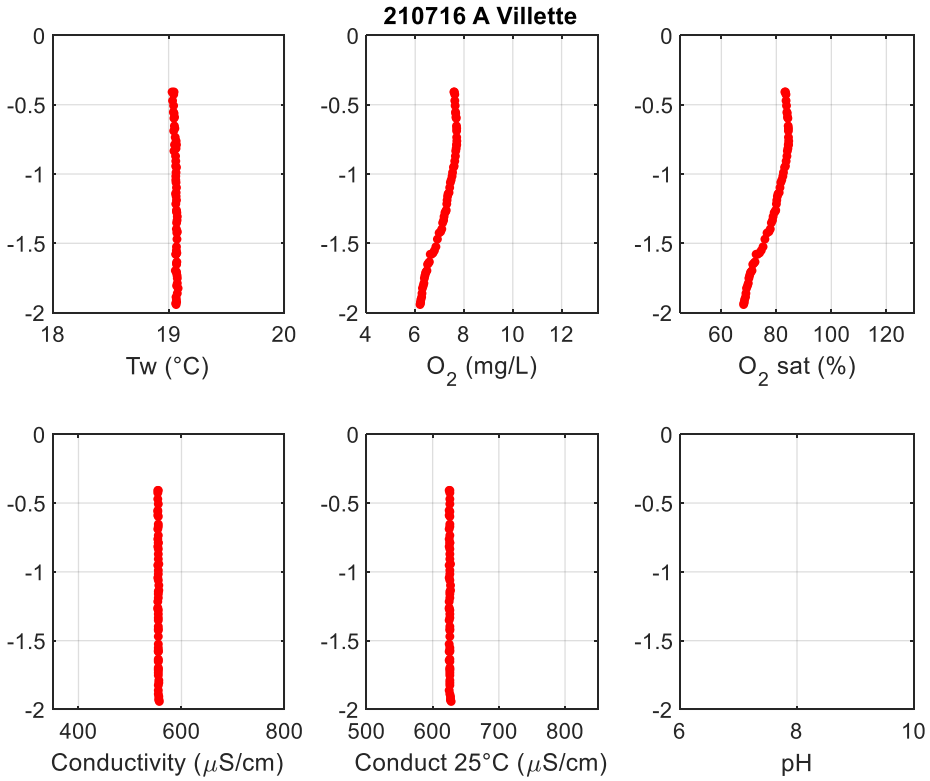
Point B – 08:00 (UTC)



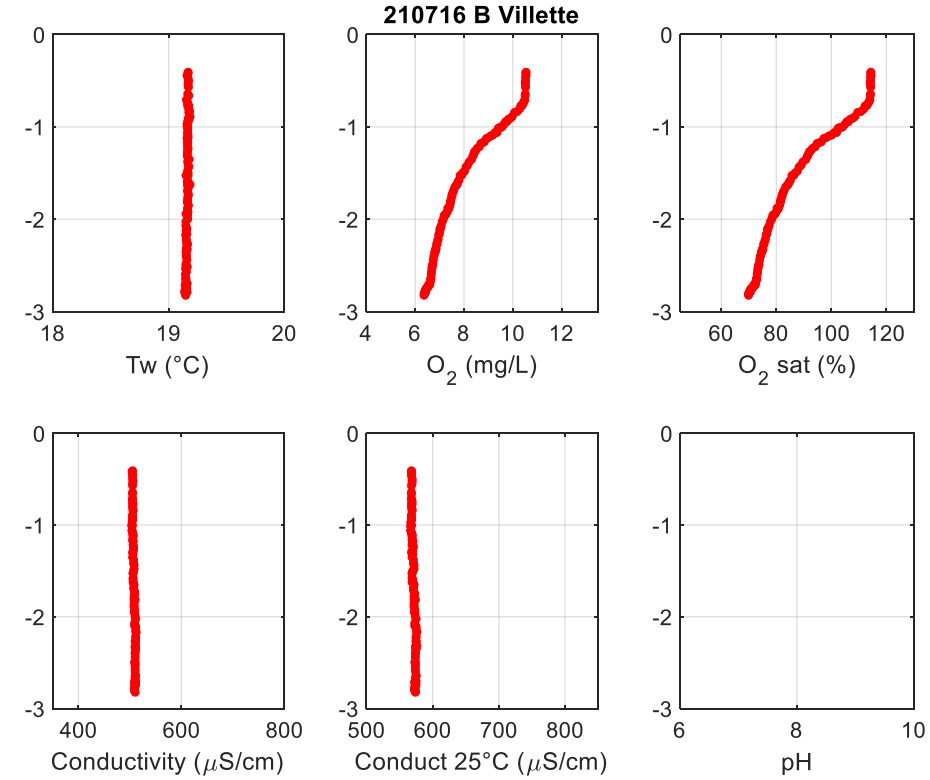


16/07/2021

Point A – 09:30 (UTC)

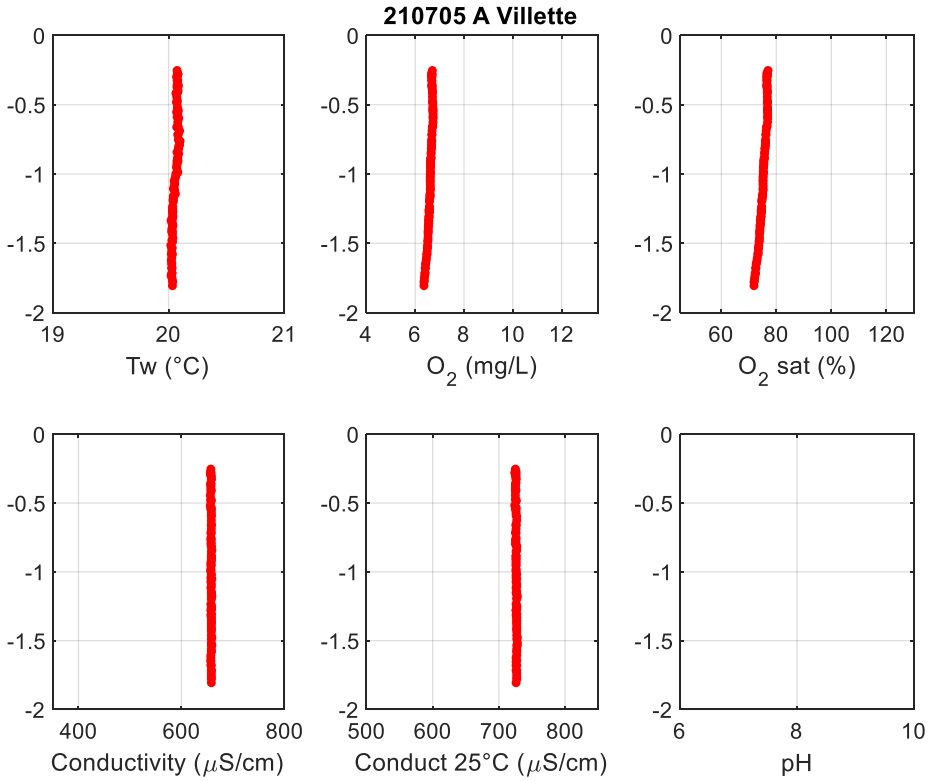


Point B – 08:20 (UTC)

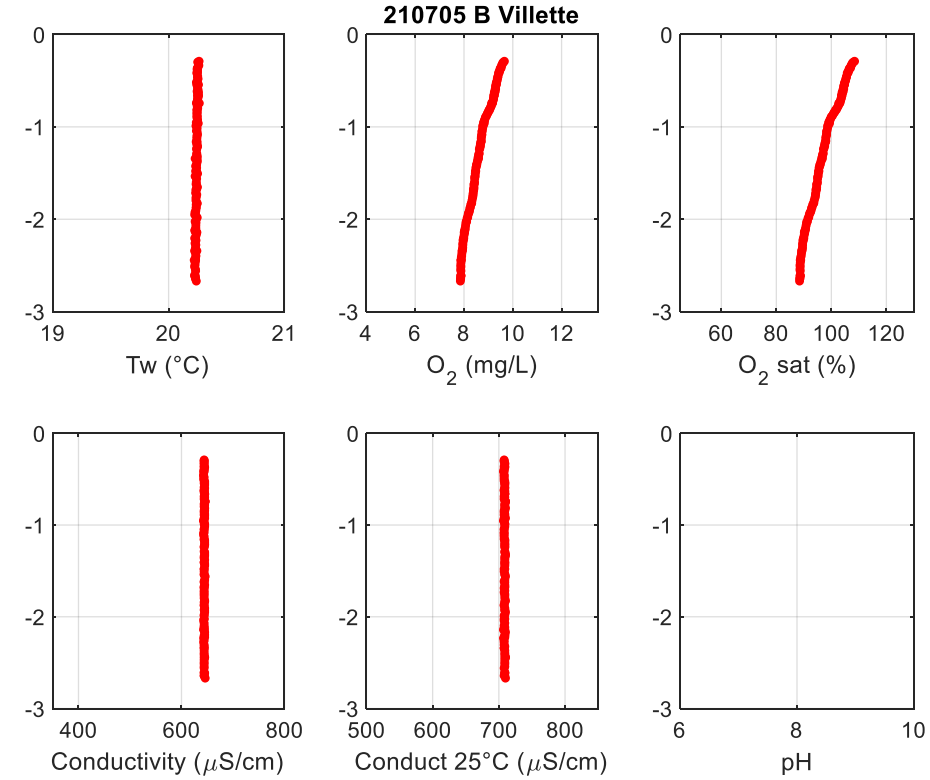


05/07/2021

Point A – 07:52 (UTC)

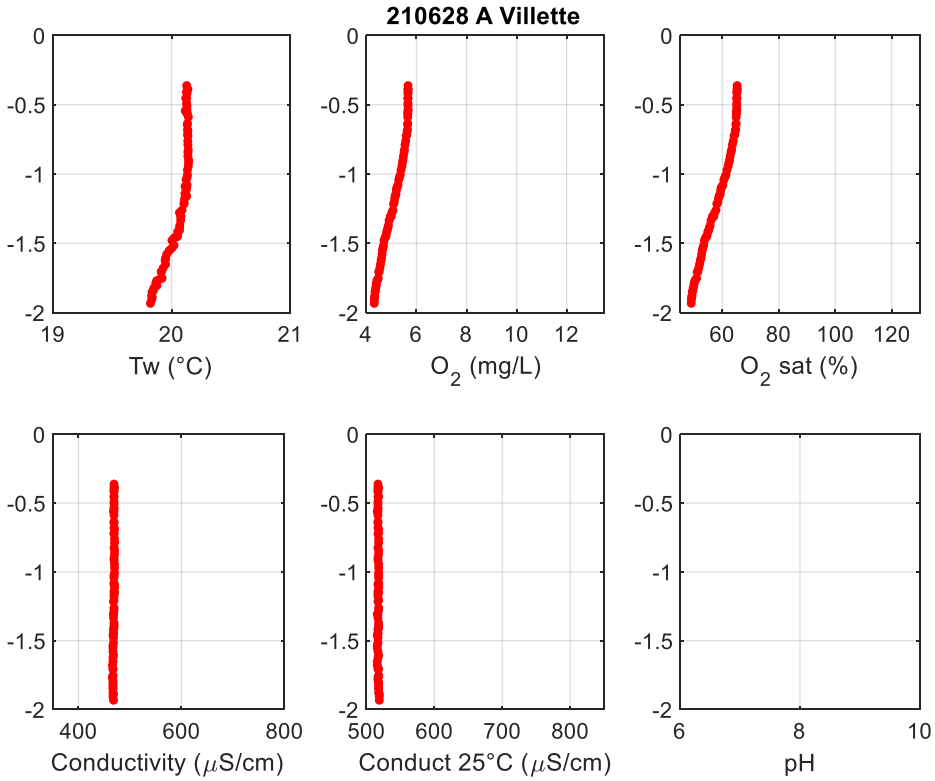


Point B – 08:42 (UTC)

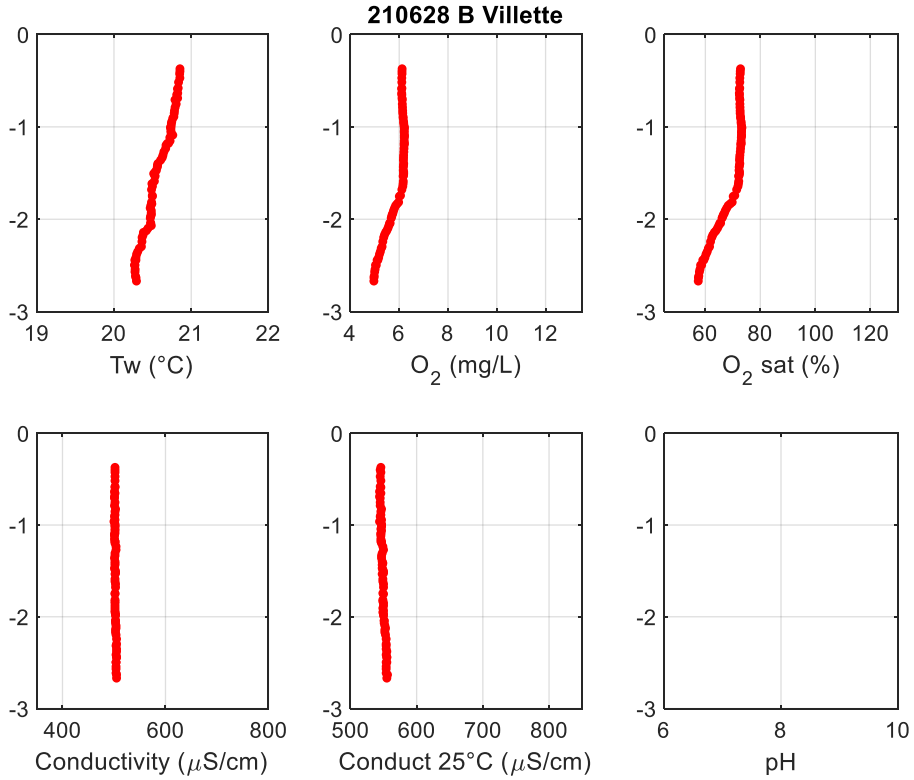


28/06/2021 – Thermal stratification (A and B)

Point A – 08:27 (UTC)

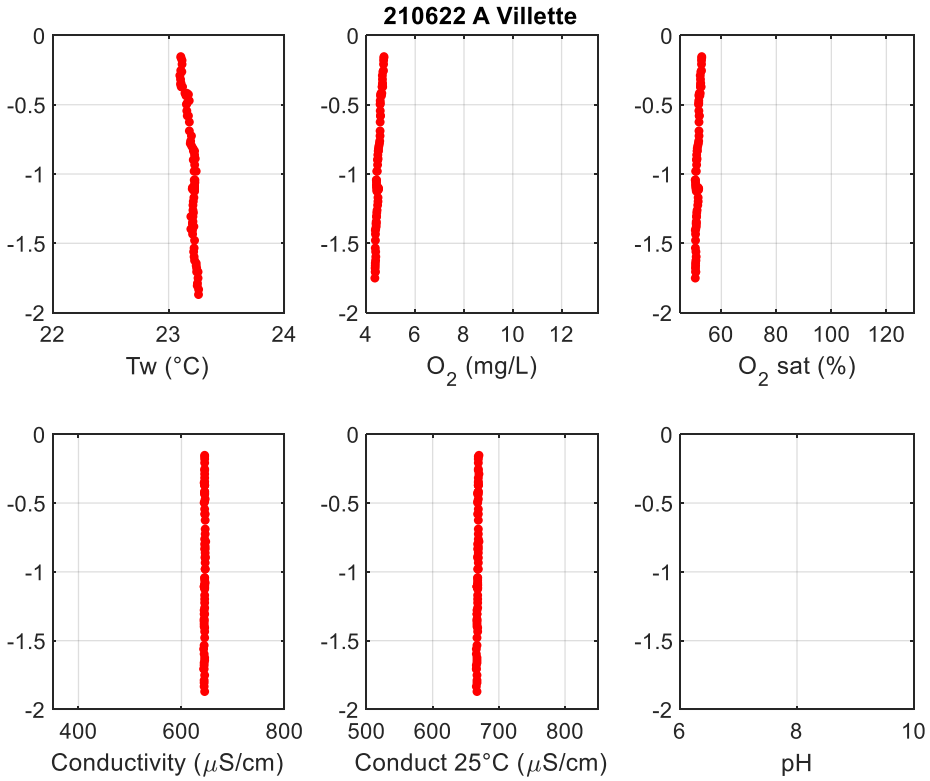


Point B – 09:30 (UTC)

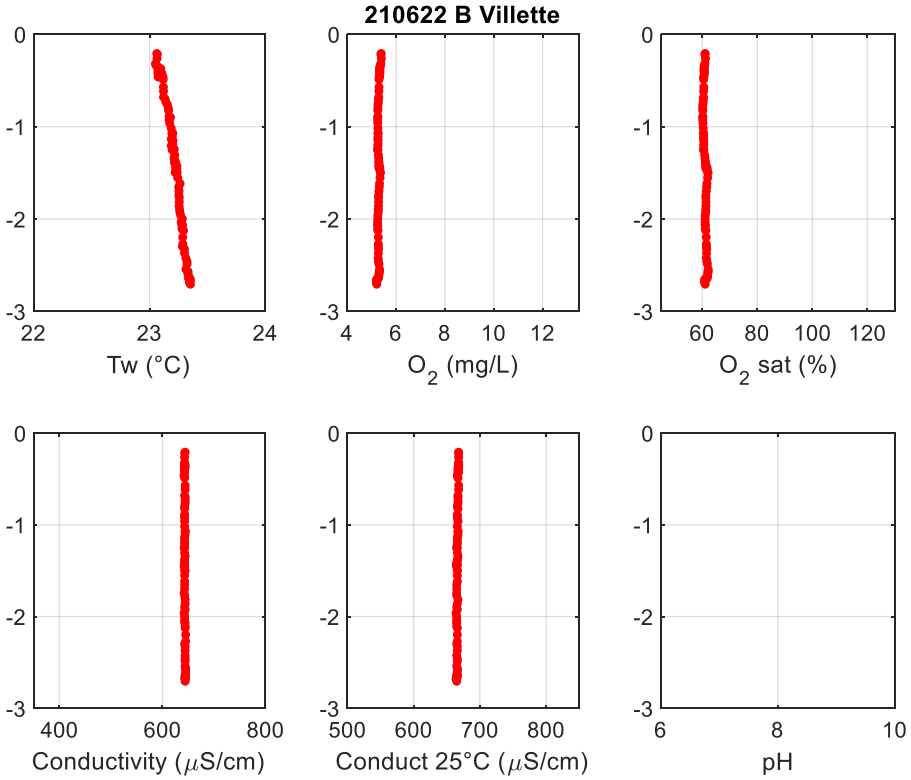


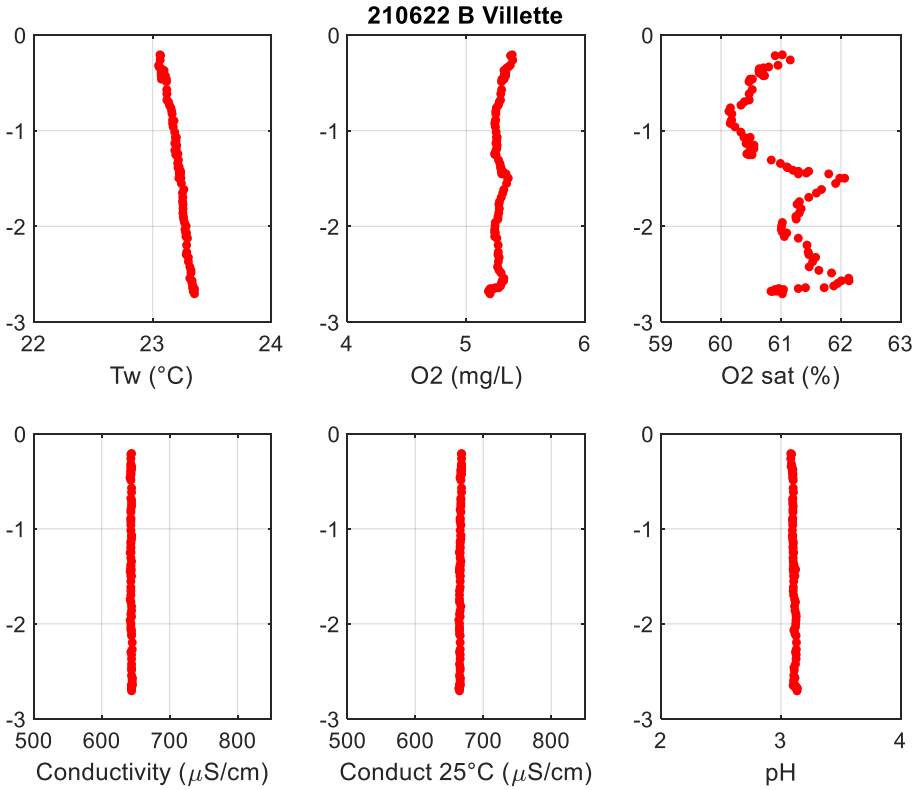
22/06/2021 – Higher Tw at the bottom

Point A – 08:12 (UTC)



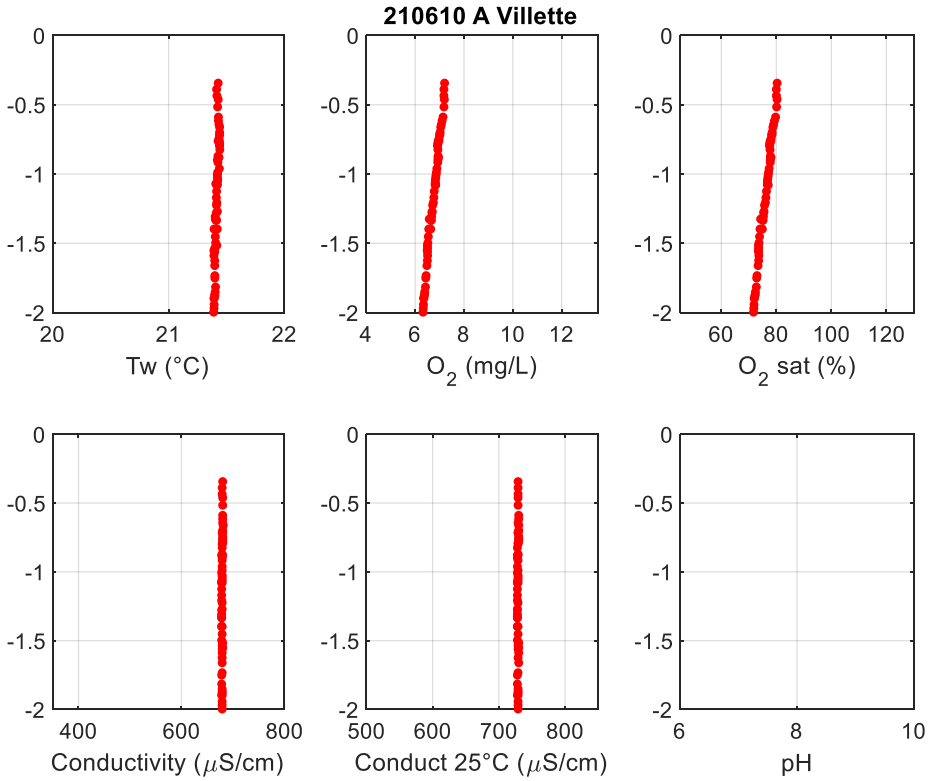
Point B – 09:06 (UTC)



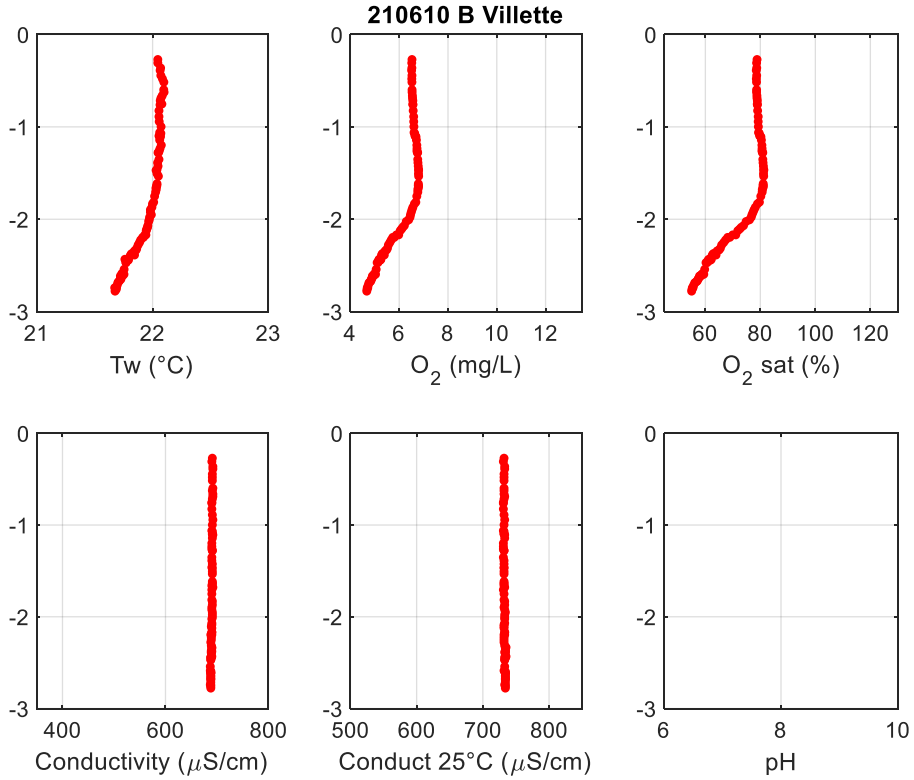


10/06/2021 – Thermal stratification (B)

Point A – 08:03 (UTC)

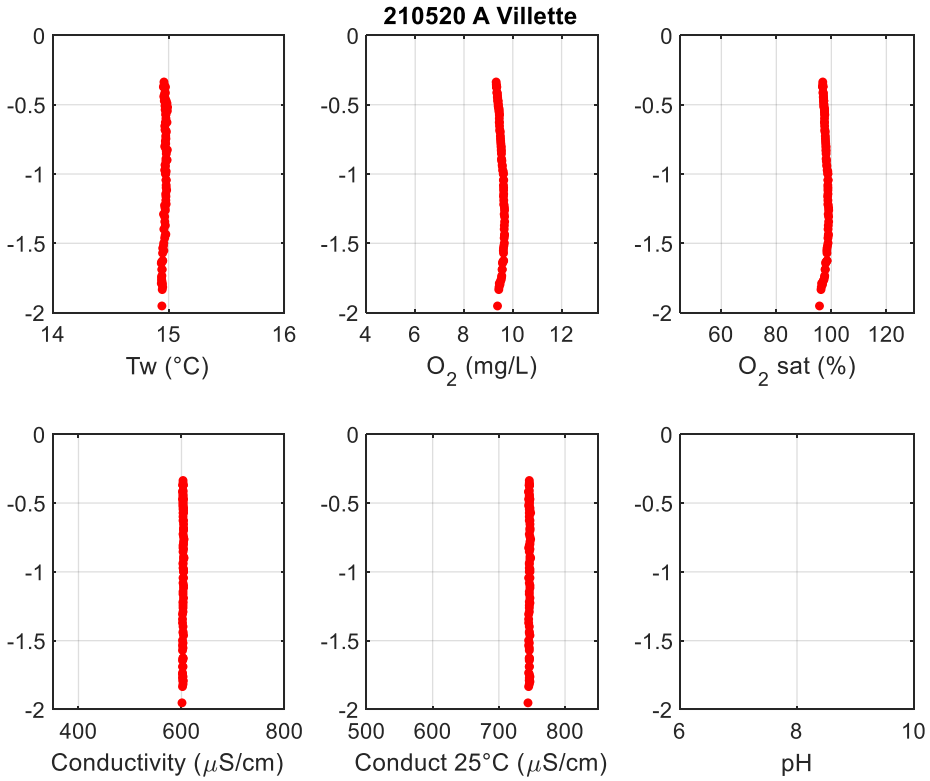


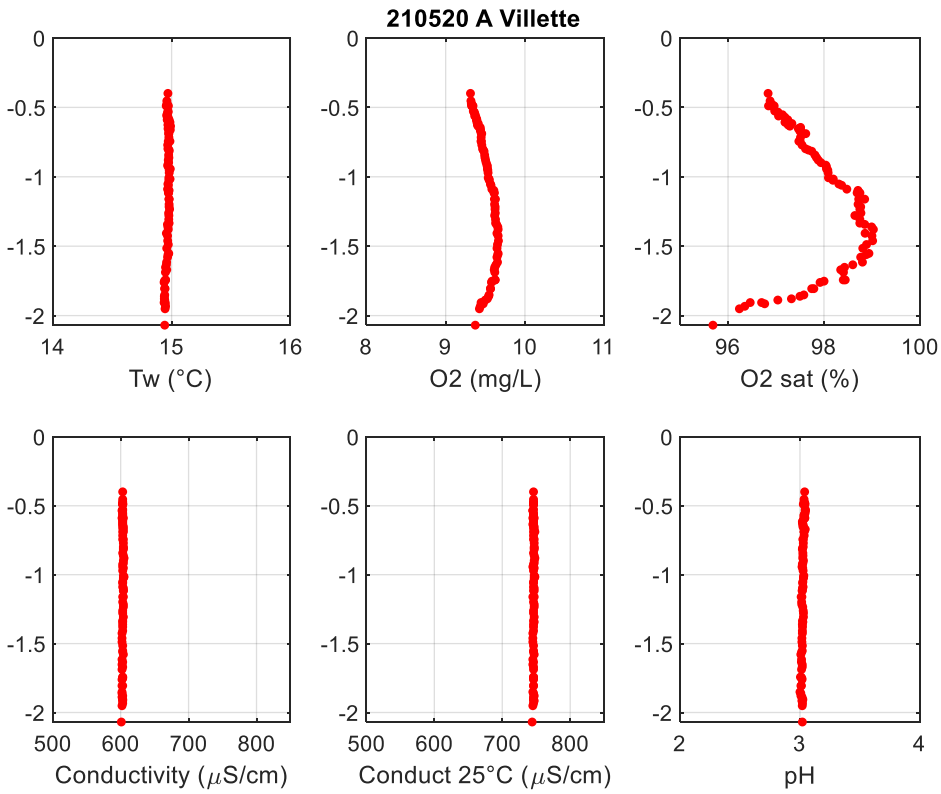
Point B – 09:15 (UTC)



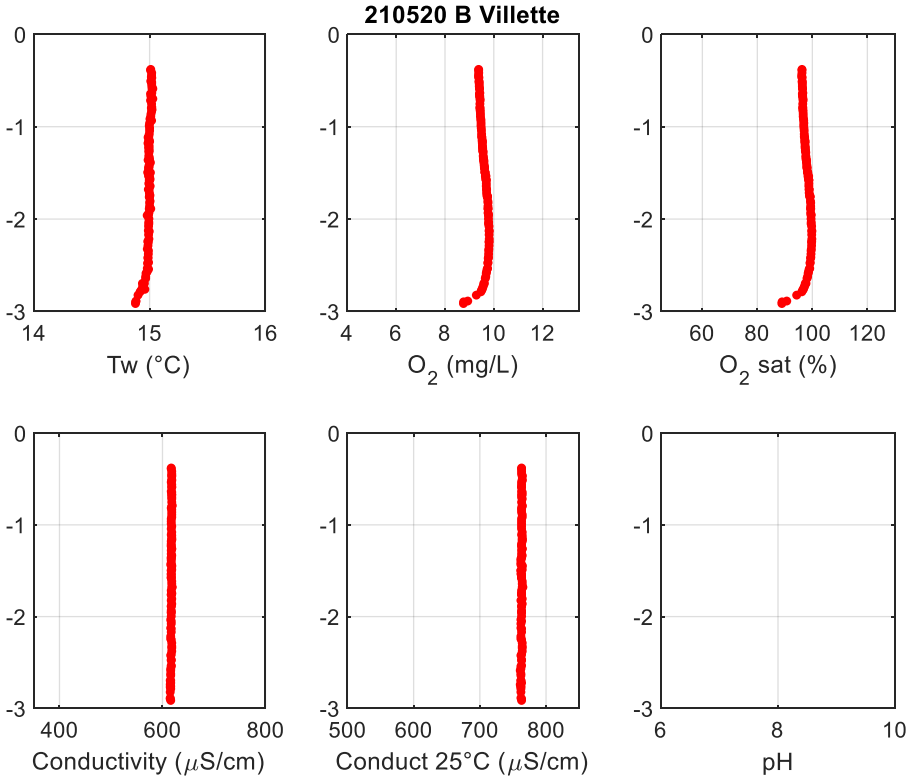
20/05/2021

Point A – 08:02 (UTC)



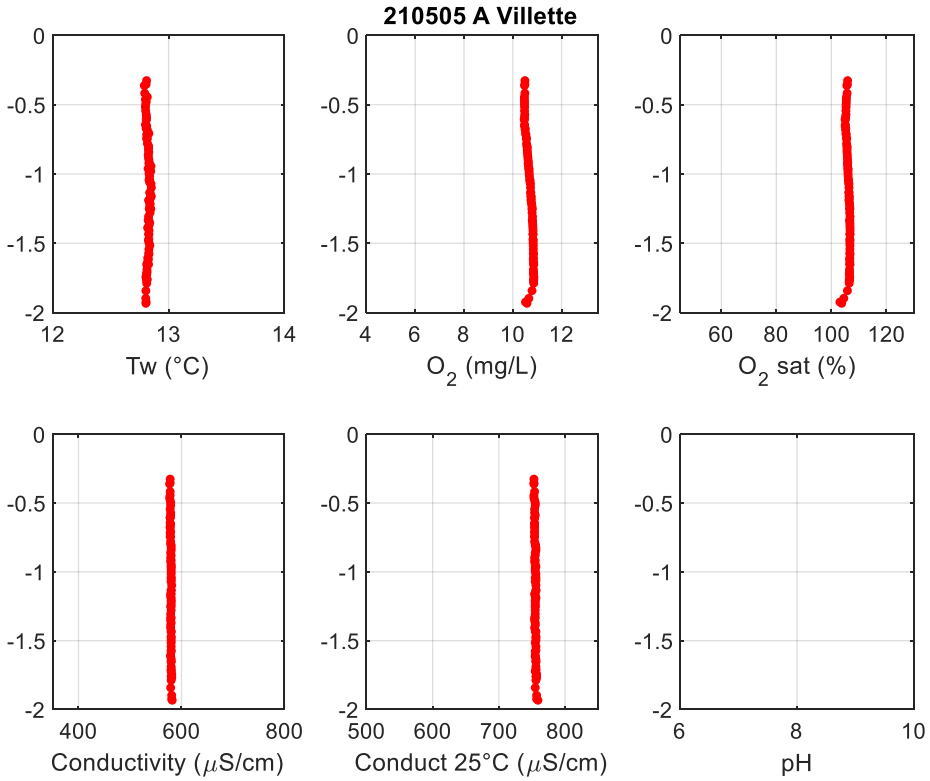


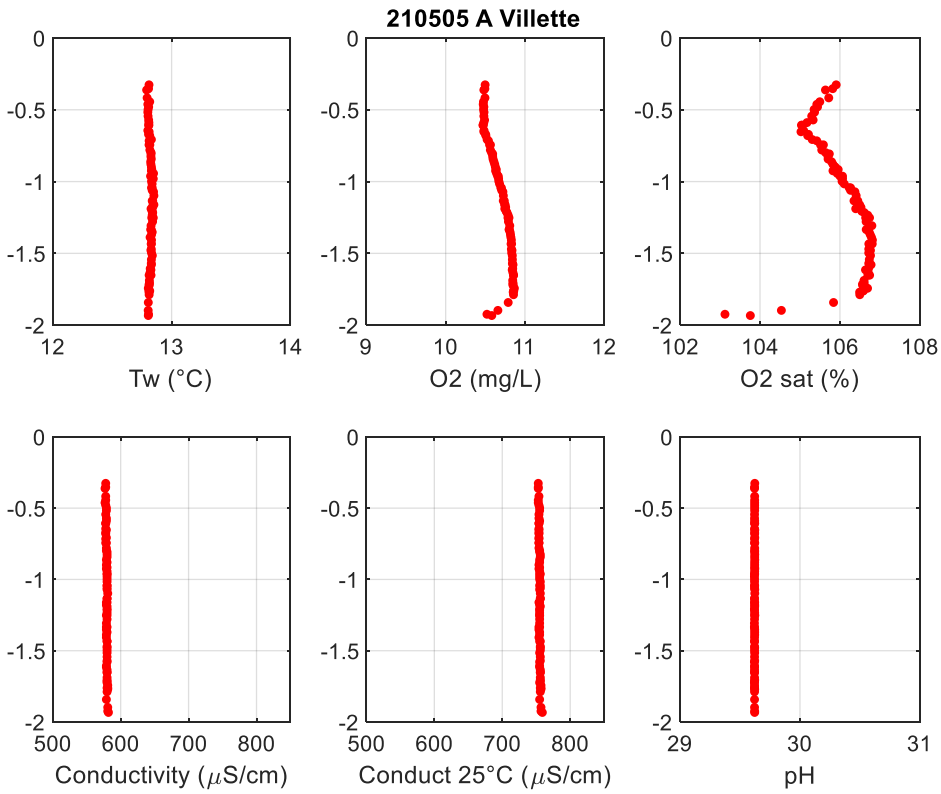
Point B – 09:15 (UTC)



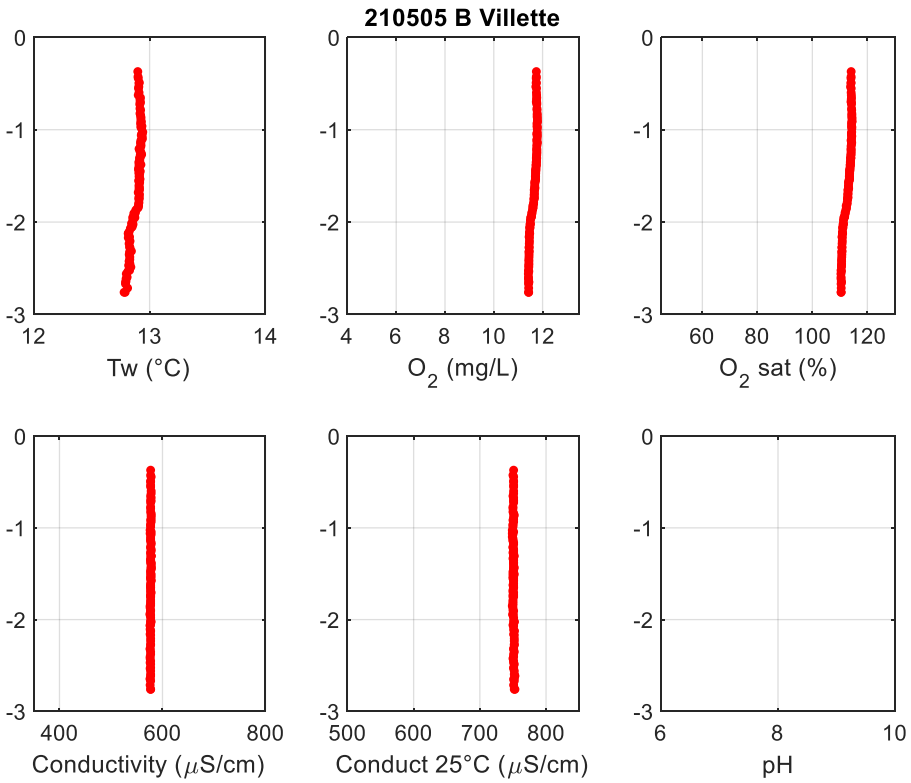
05/05/2021 – pH stops working

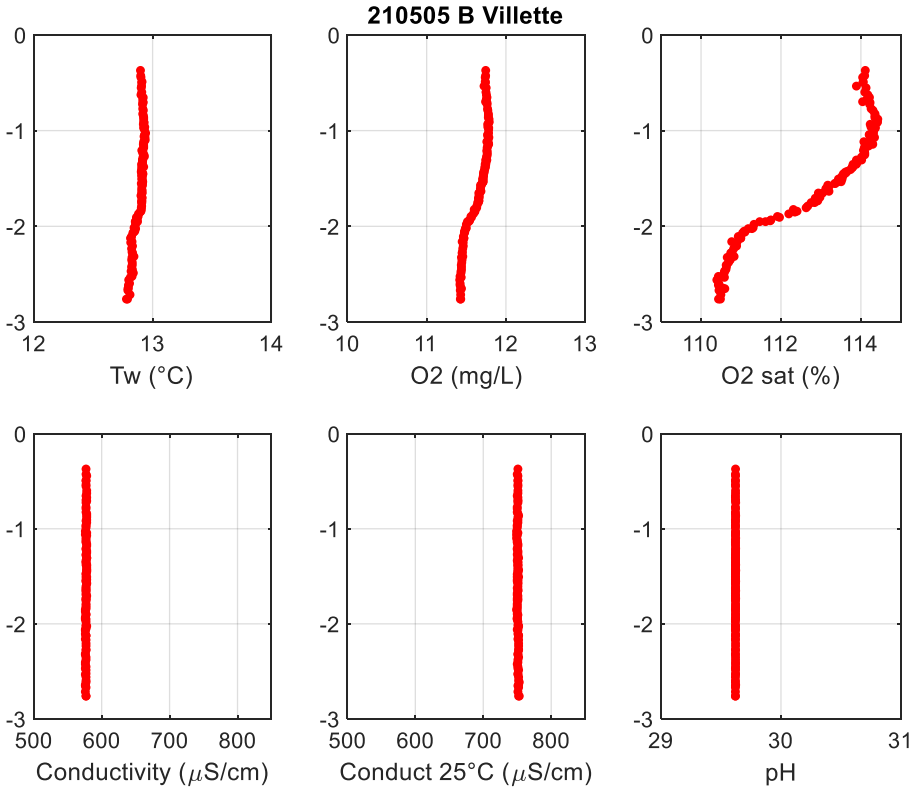
Point A – 11:00 (UTC)





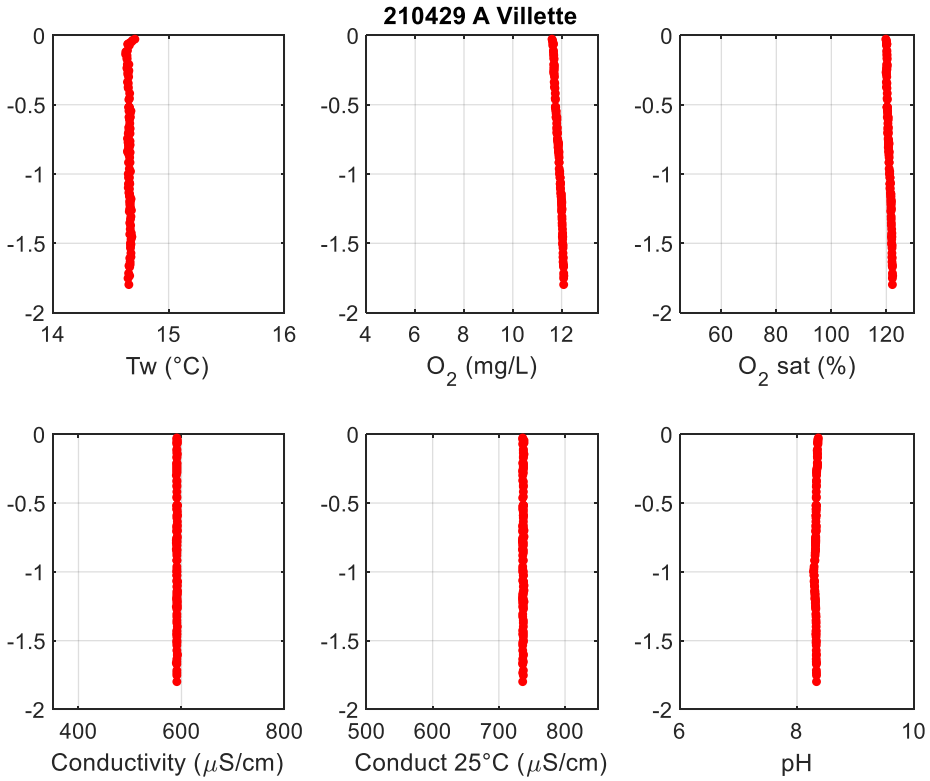
Point B – 14:25 (UTC)



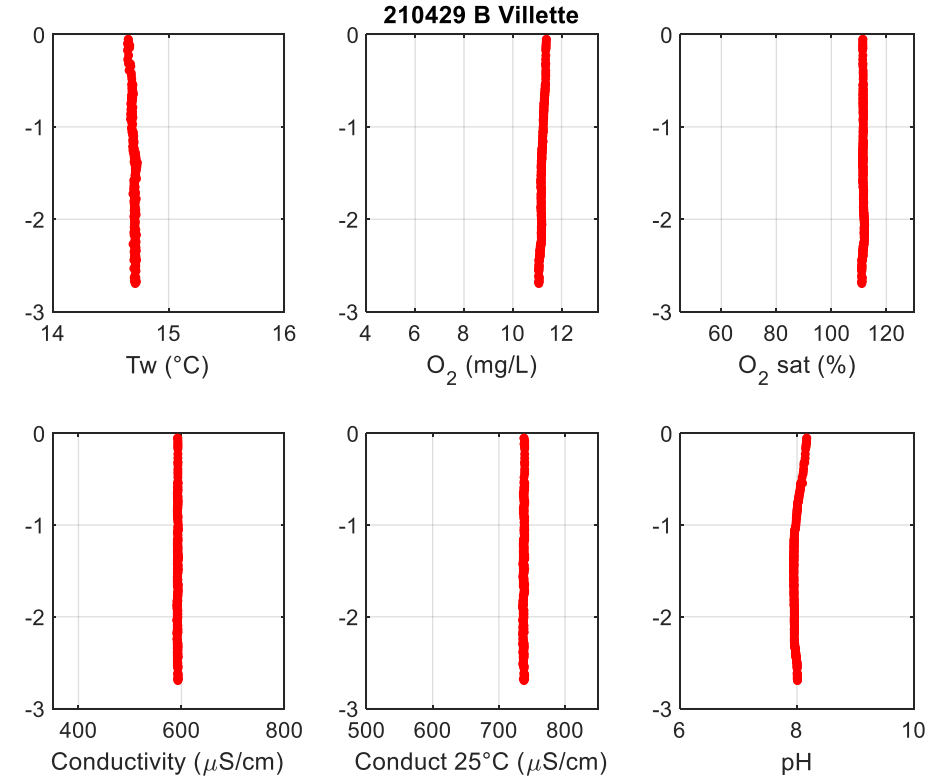


29/04/2021

Point A – 07:55 (UTC)

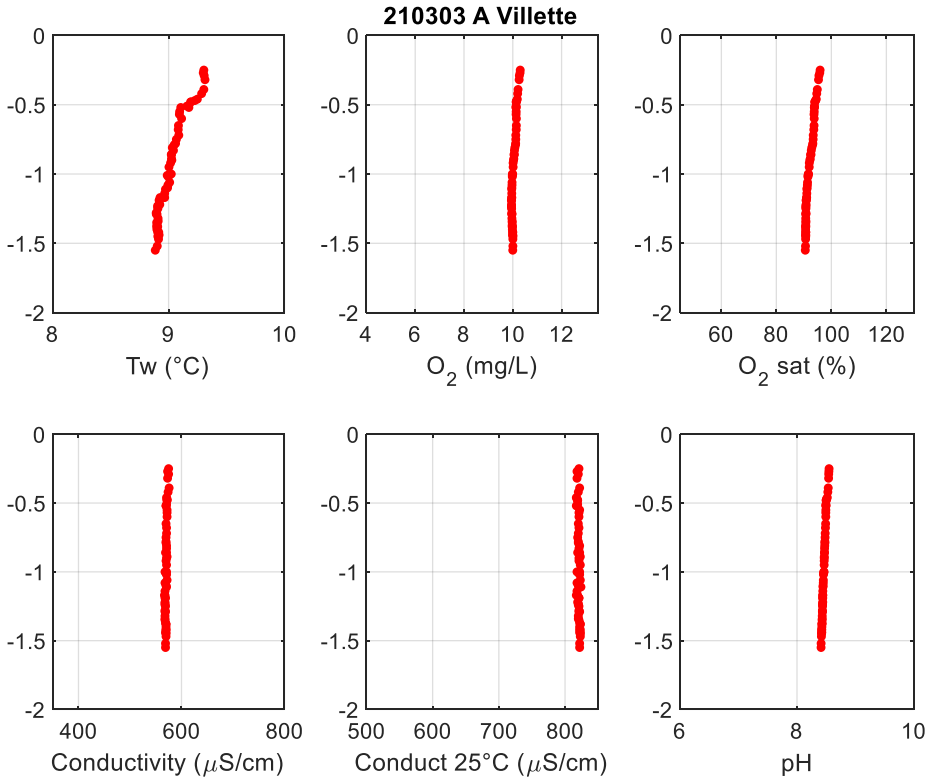


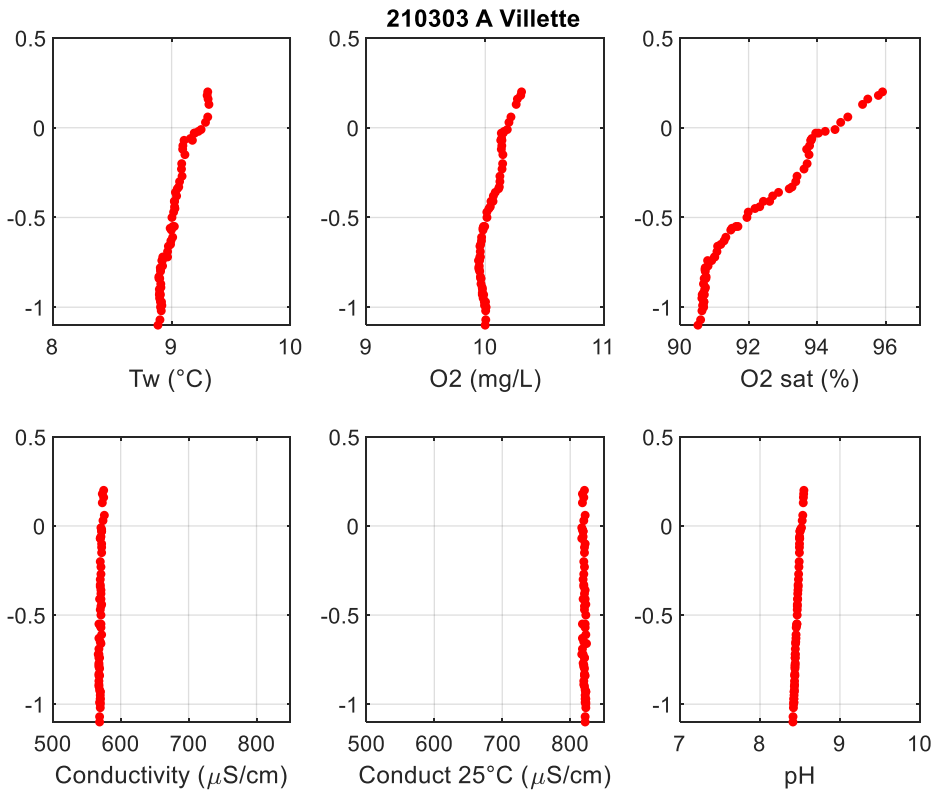
Point B – 09:05 (UTC)



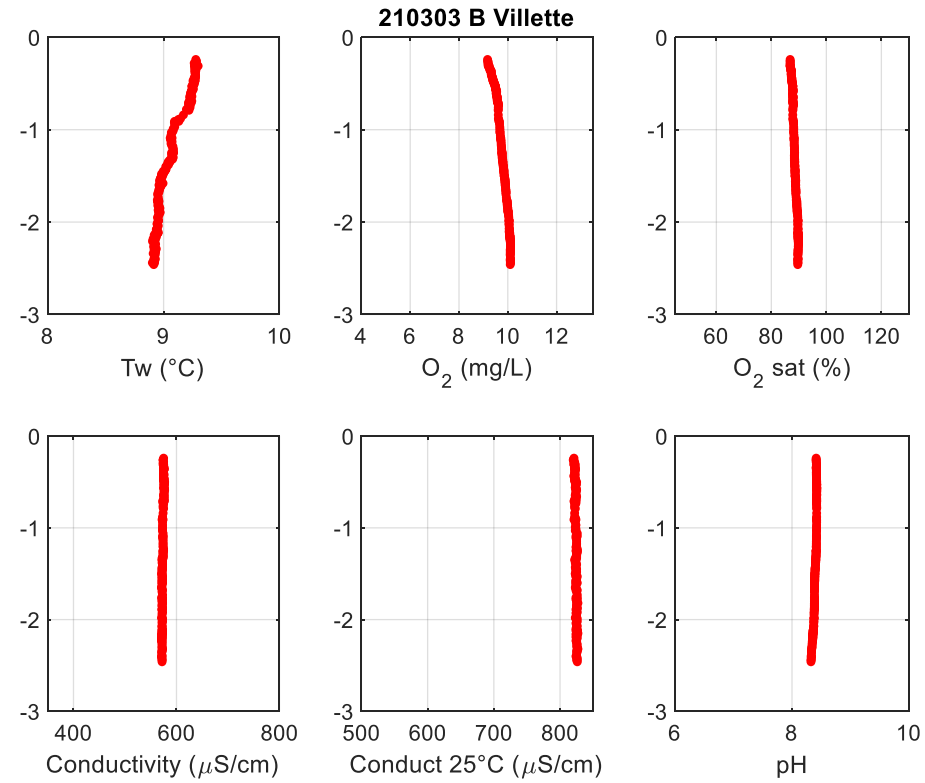
03/03/2021 – Thermal stratification (A and B)

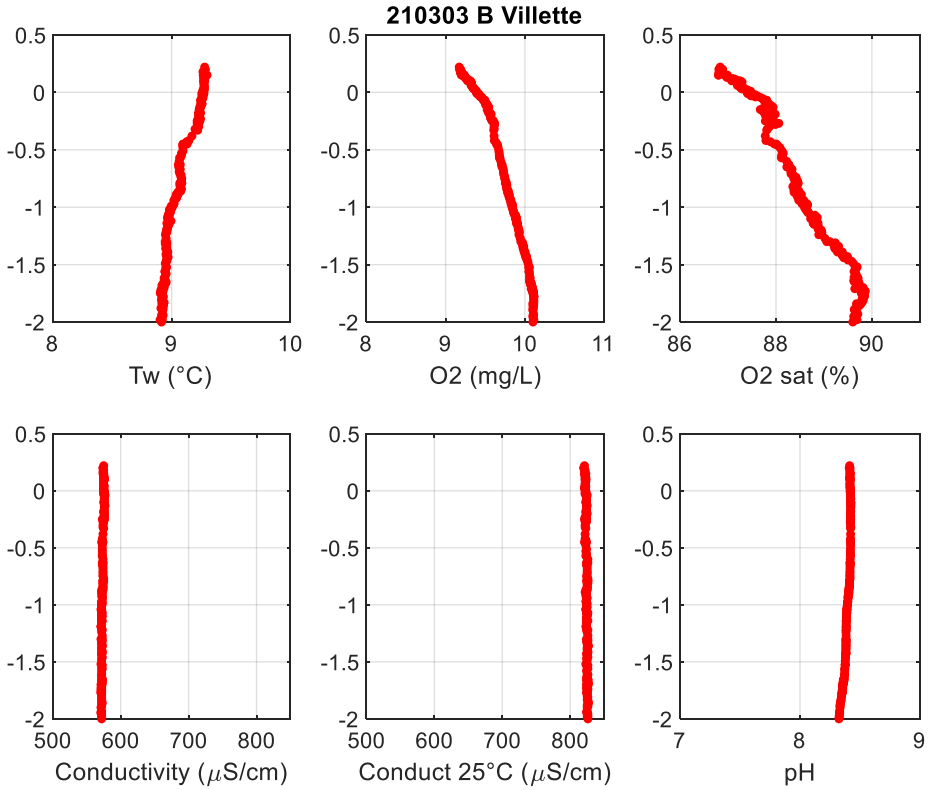
Point A – 09:25 (UTC)





Point B – 10:50 (UTC)

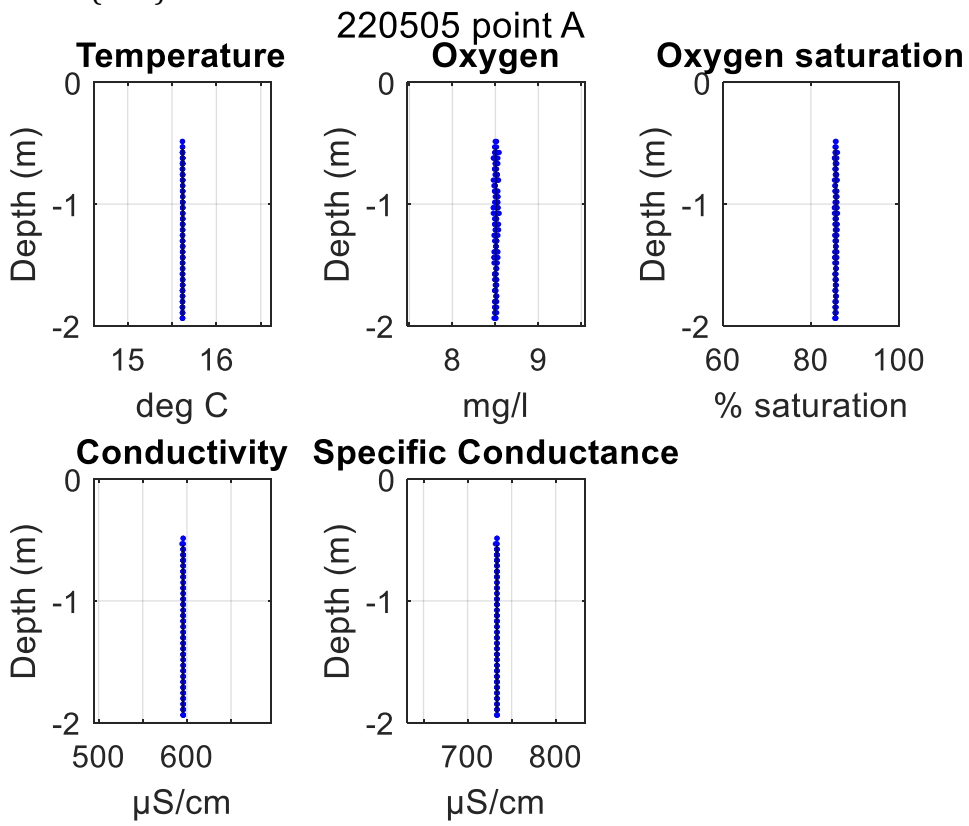




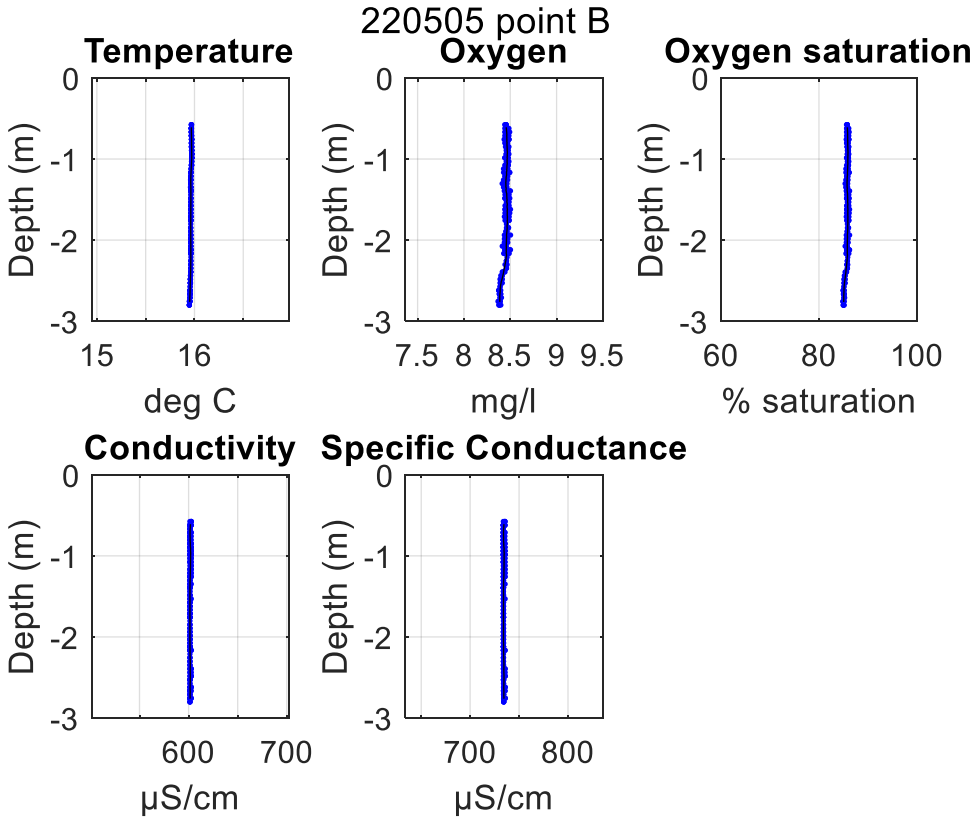
2. SeaBird profiles

05/05/2022

Point A – 07:20 (UTC)

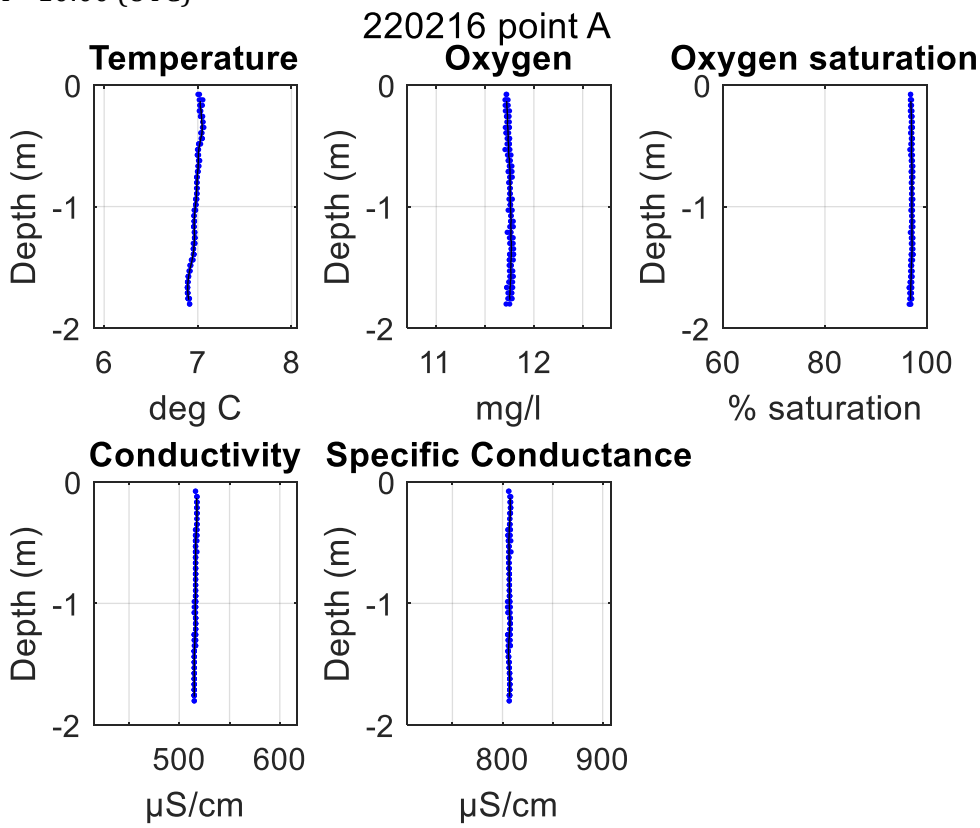


Point B – 08:10 (UTC)

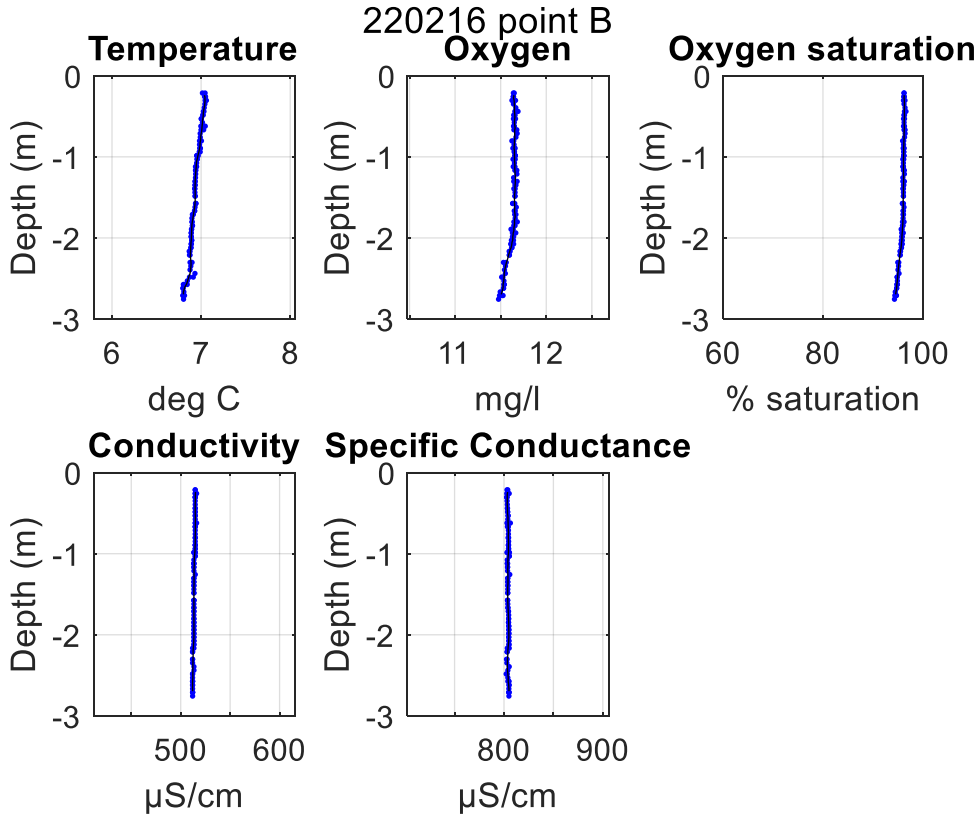


16/02/2022

Point A – 10:00 (UTC)

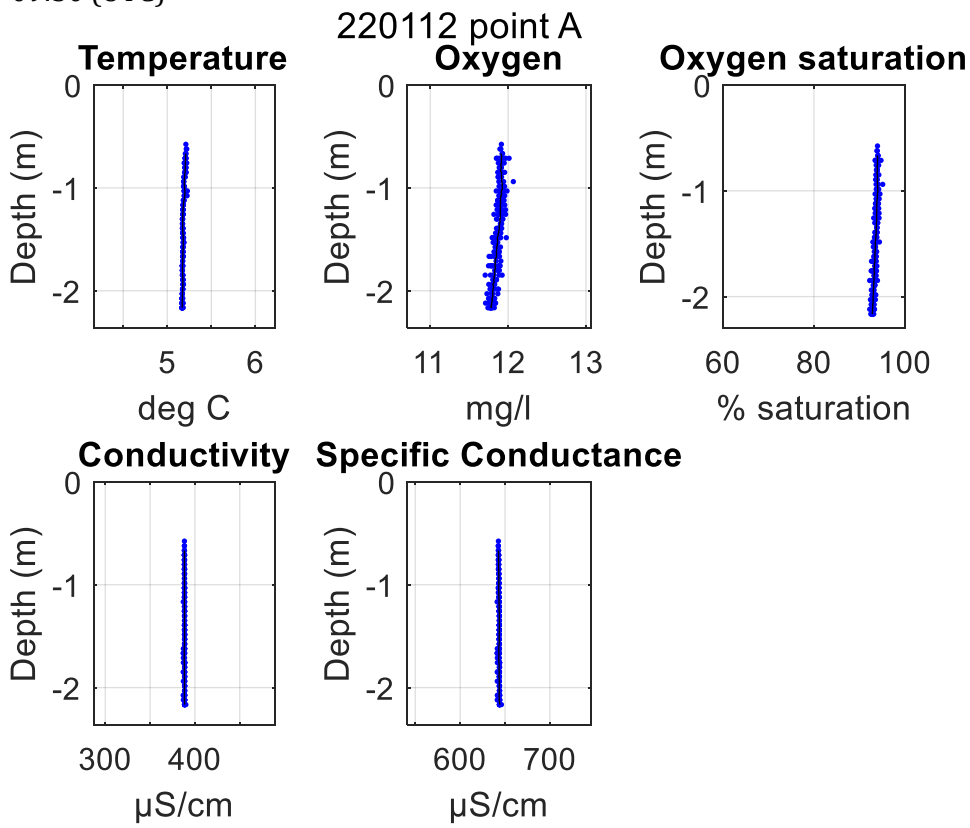


Point B – 10:50 (UTC)

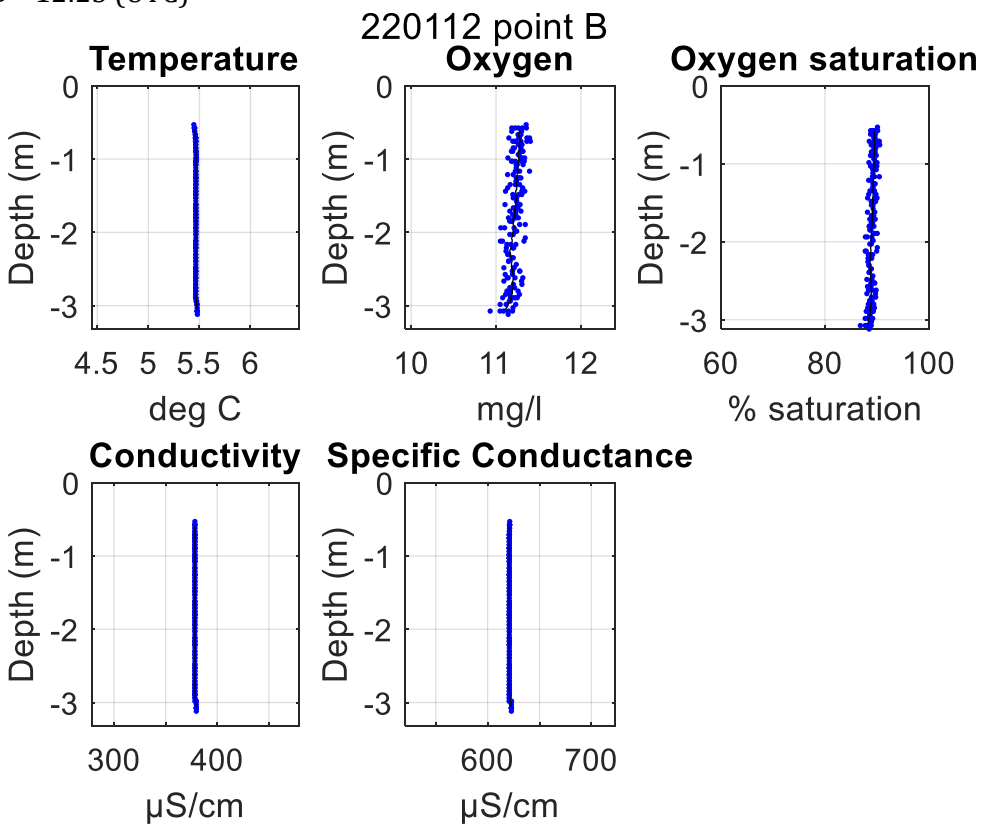


12/01/2022

Point A – 09:50 (UTC)

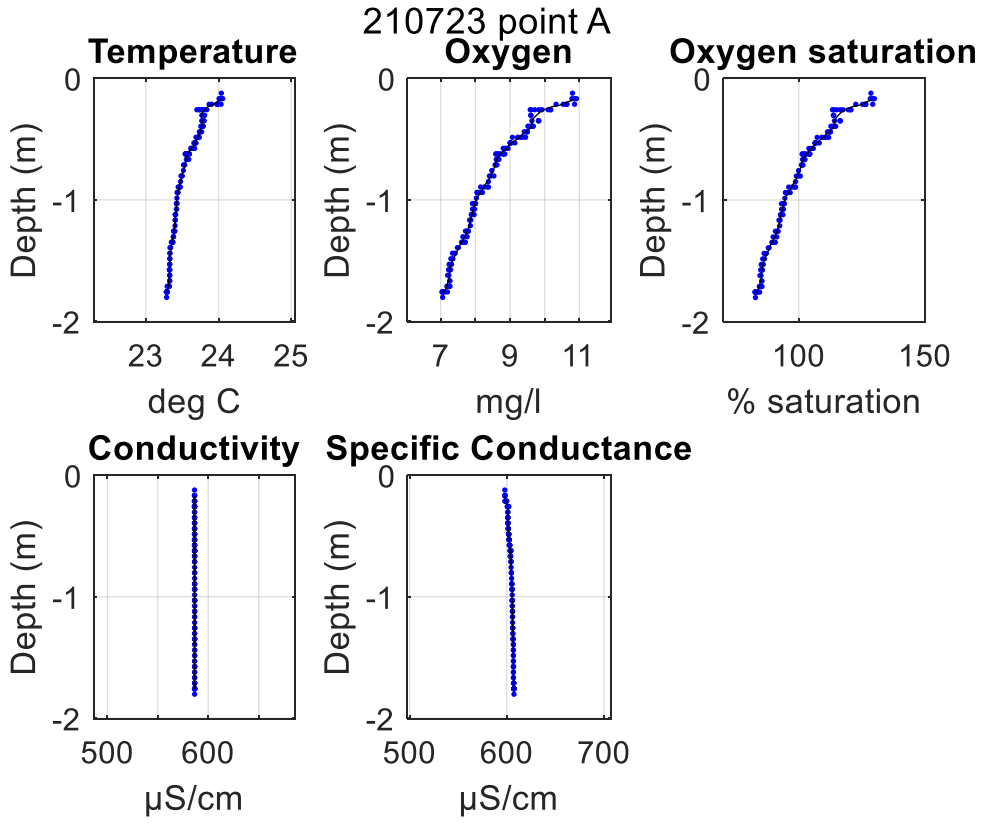


Point B – 12:25 (UTC)

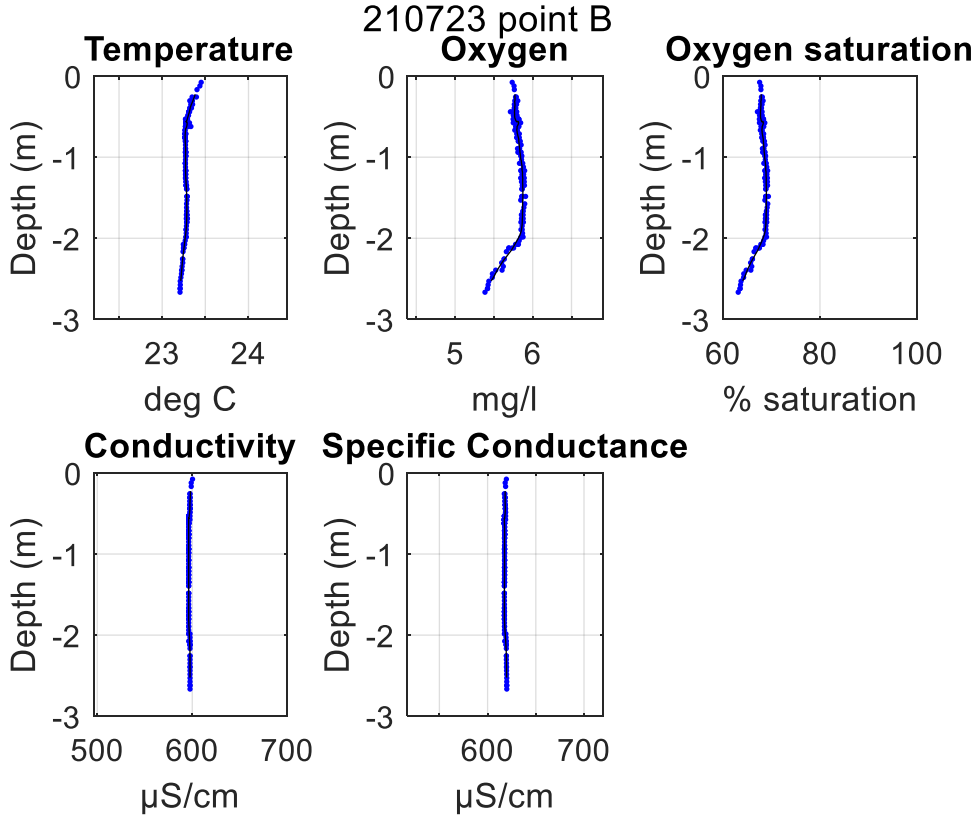


23/07/2021

Point A – 09:36 (UTC)

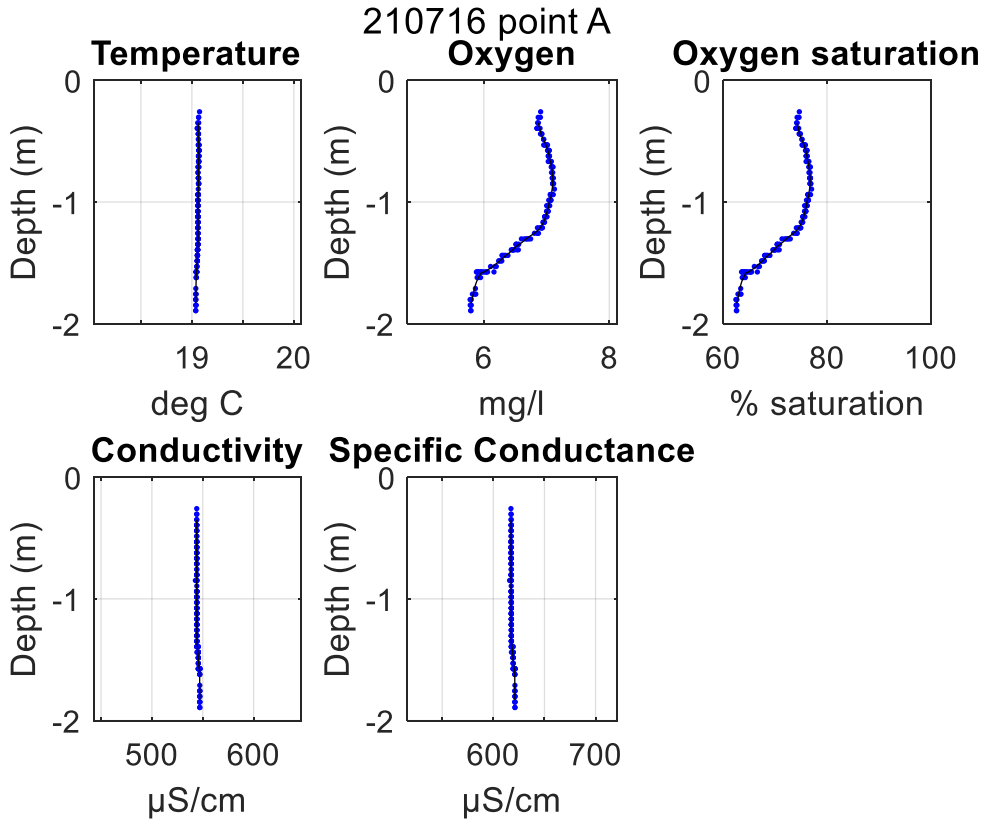


Point B – 08:00 (UTC)

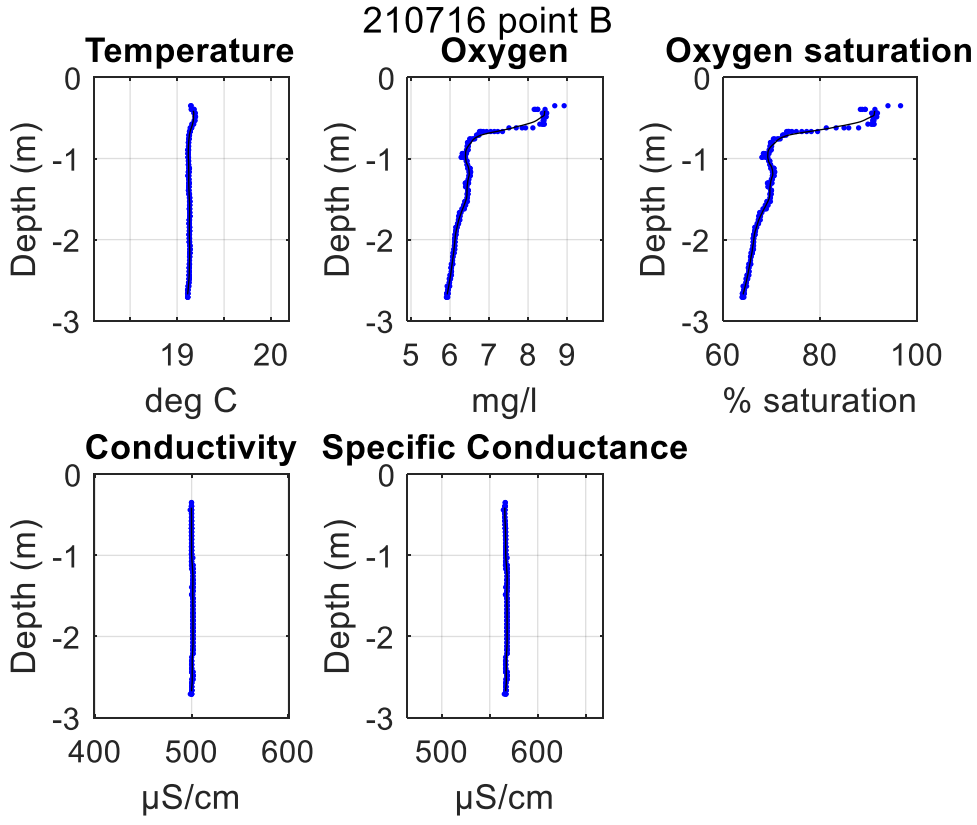


16/07/2021

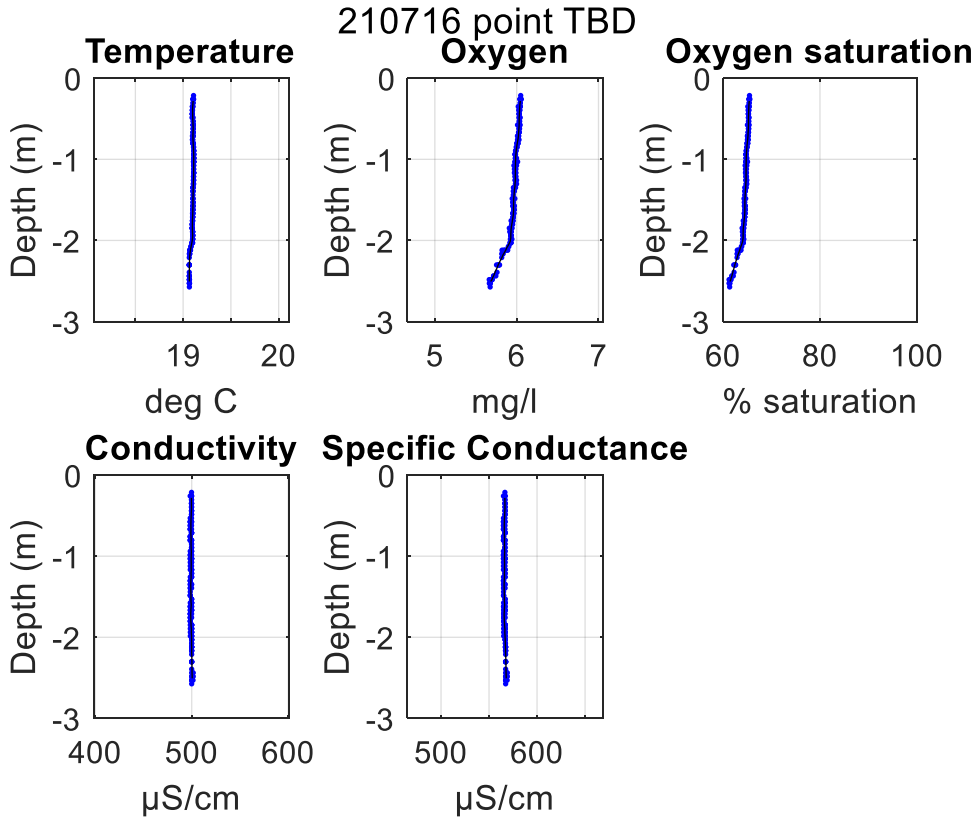
Point A – 09:30 (UTC)



Point B – 8:20 (UTC)

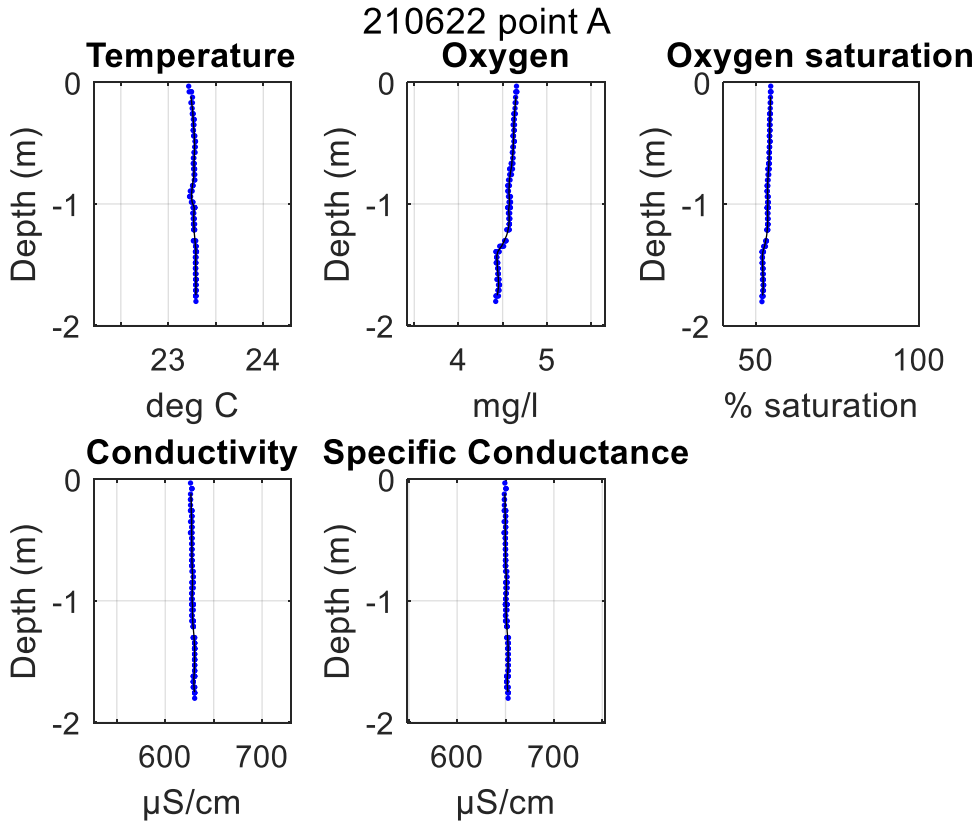


Point TBD – 07:58 (UTC)

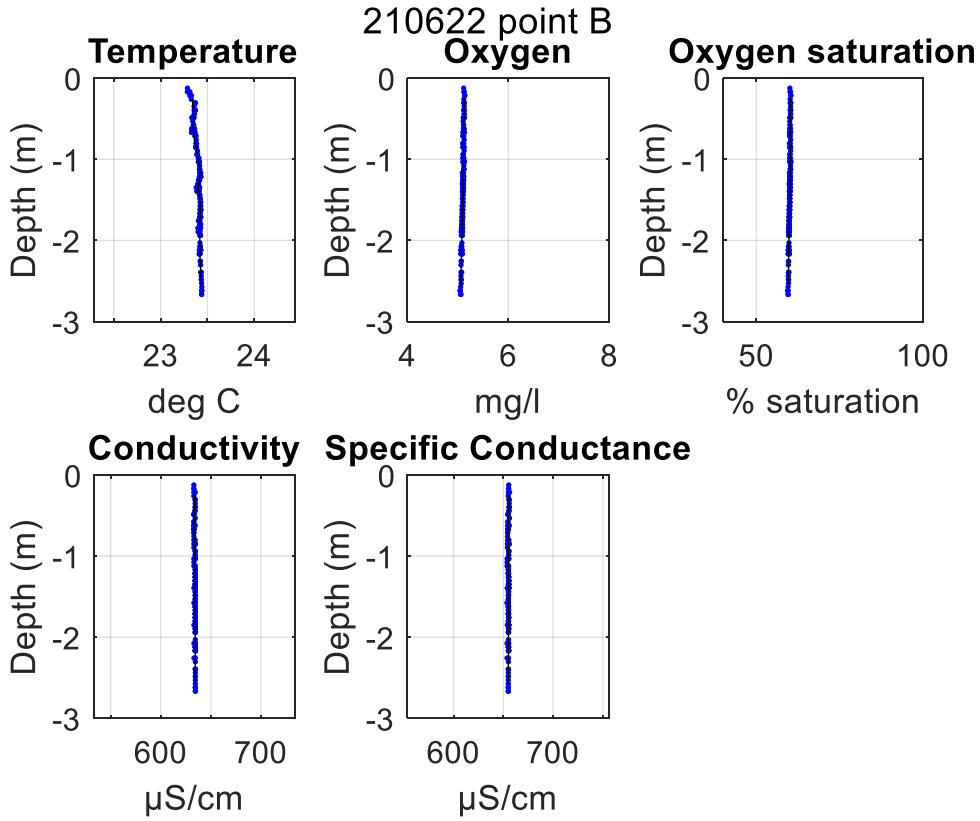


22/06/2021

Point A – 08:22 (UTC)

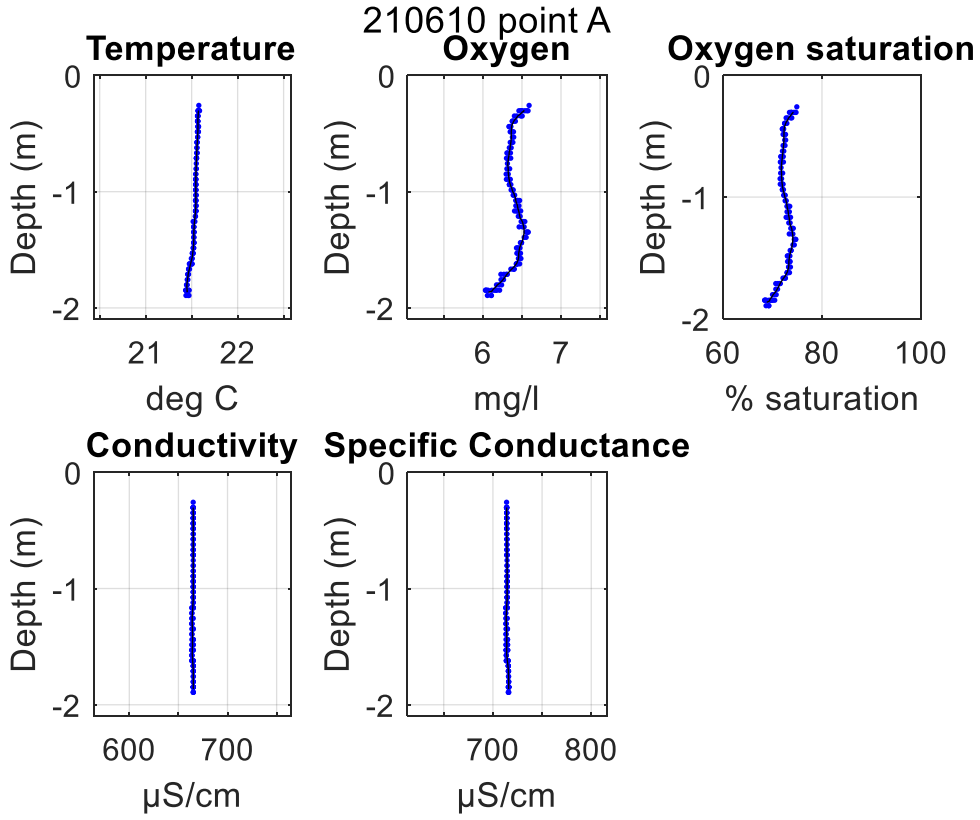


Point B – 09:20 (UTC)

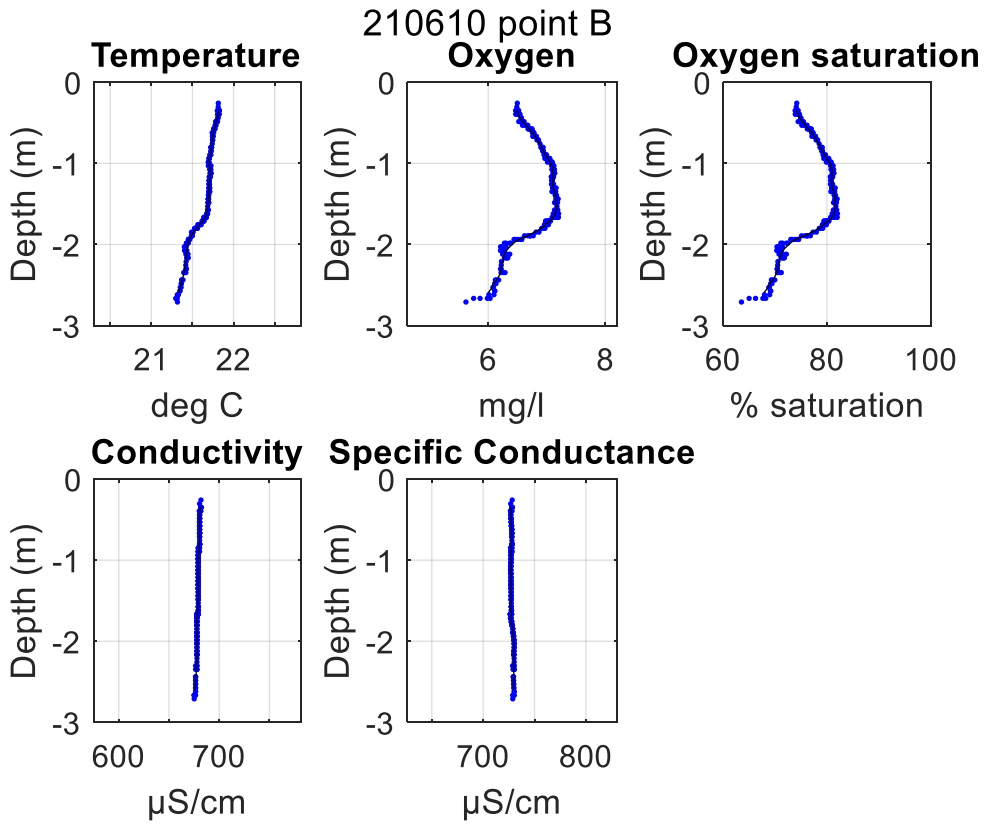


10/06/2021

Point A – 08:11 (UTC)

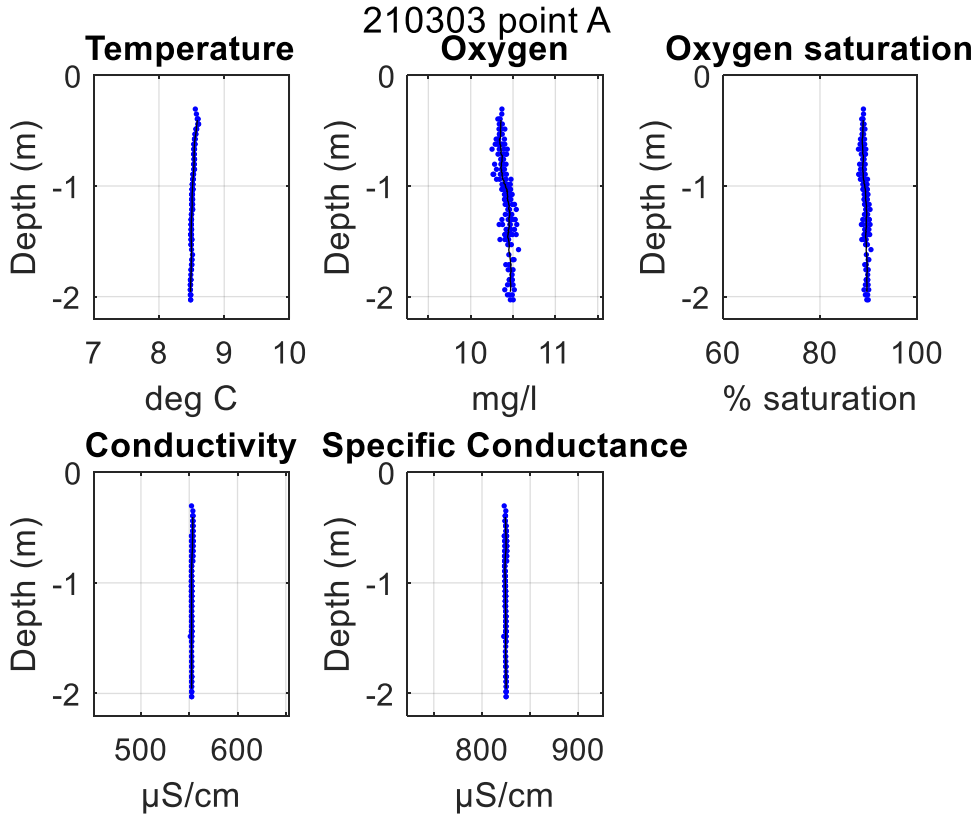


Point B – 09:15 (UTC)

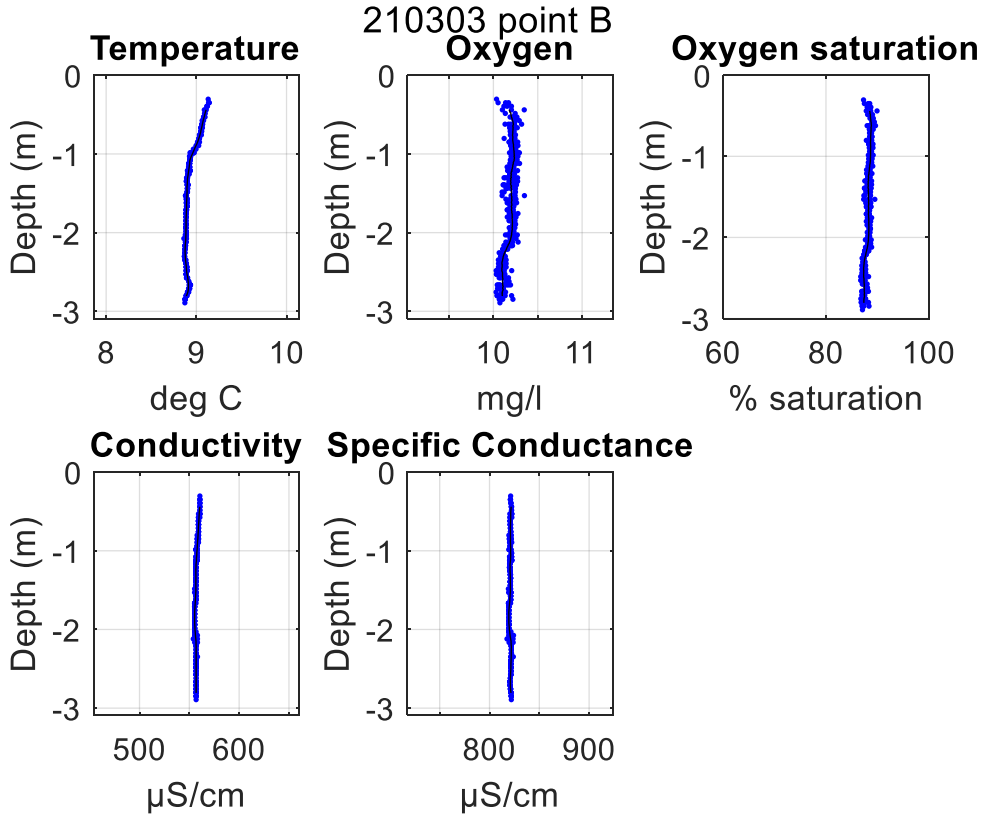


03/03/2021

Point A – 09:08 (UTC)

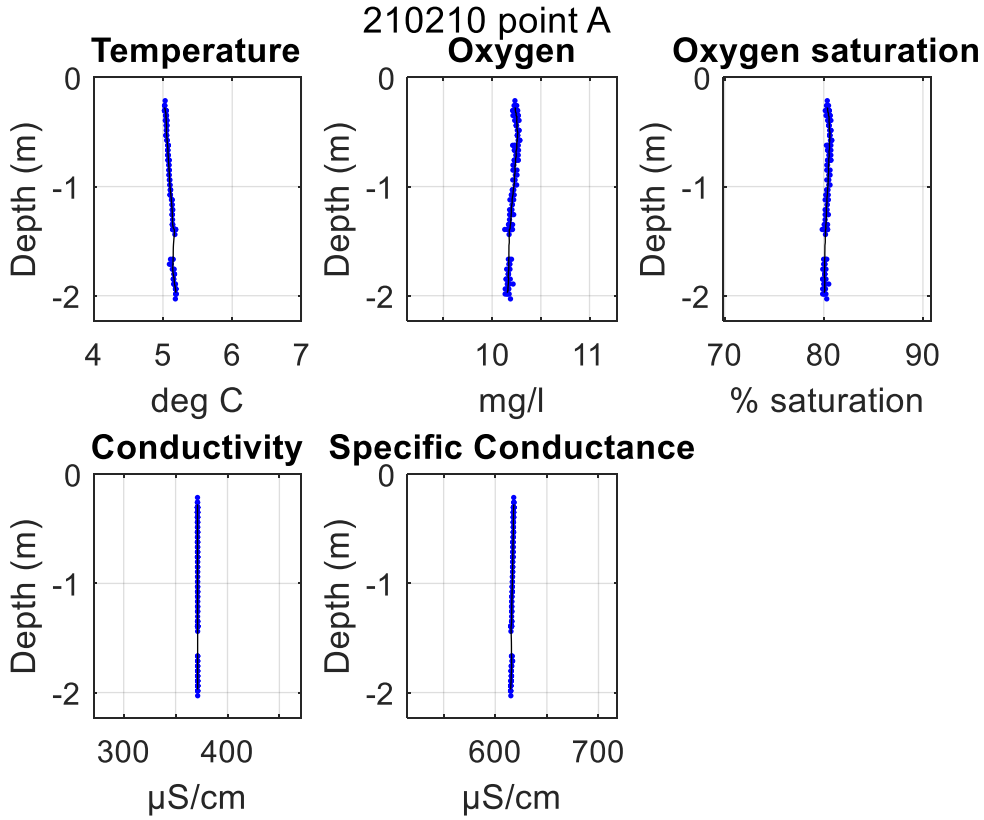


Point B – 10:33 (UTC)

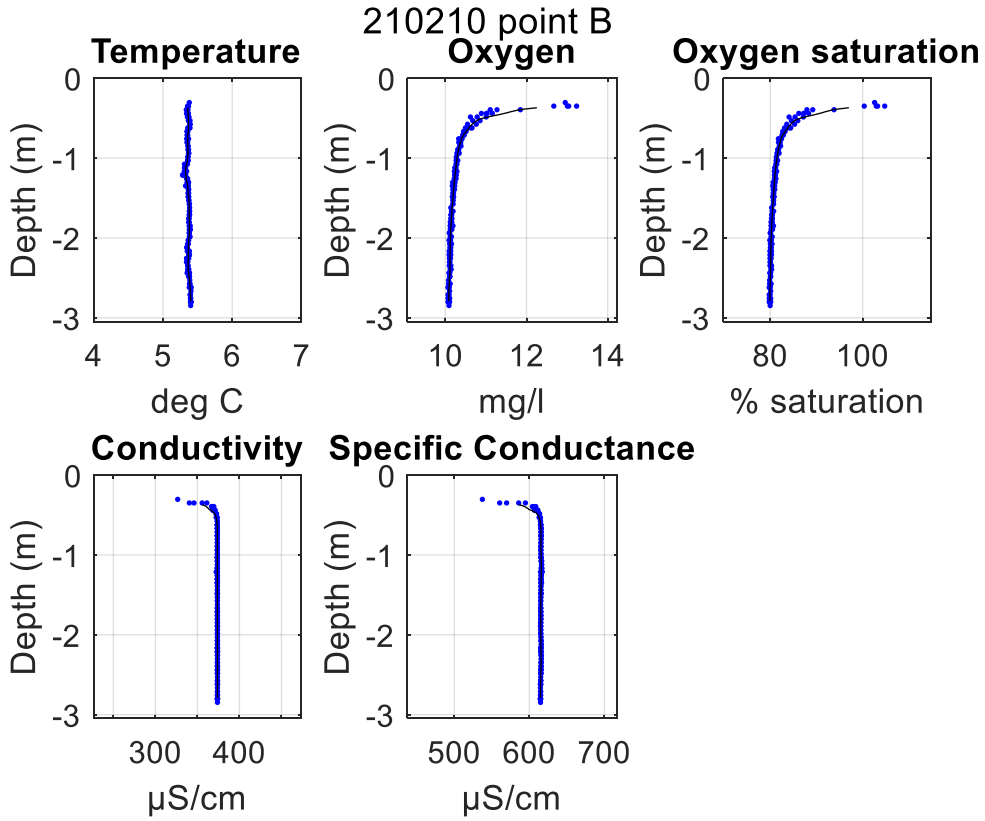


10/02/2021

Point A – 09:40 (UTC)

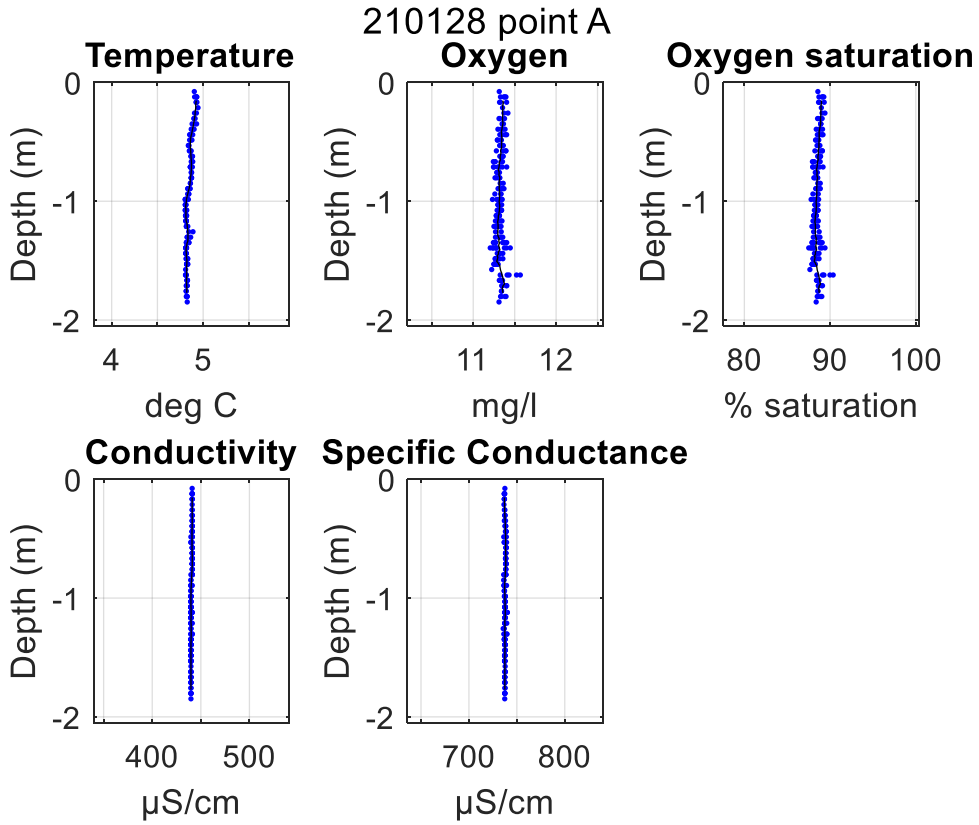


Point B – 10:40 (UTC)

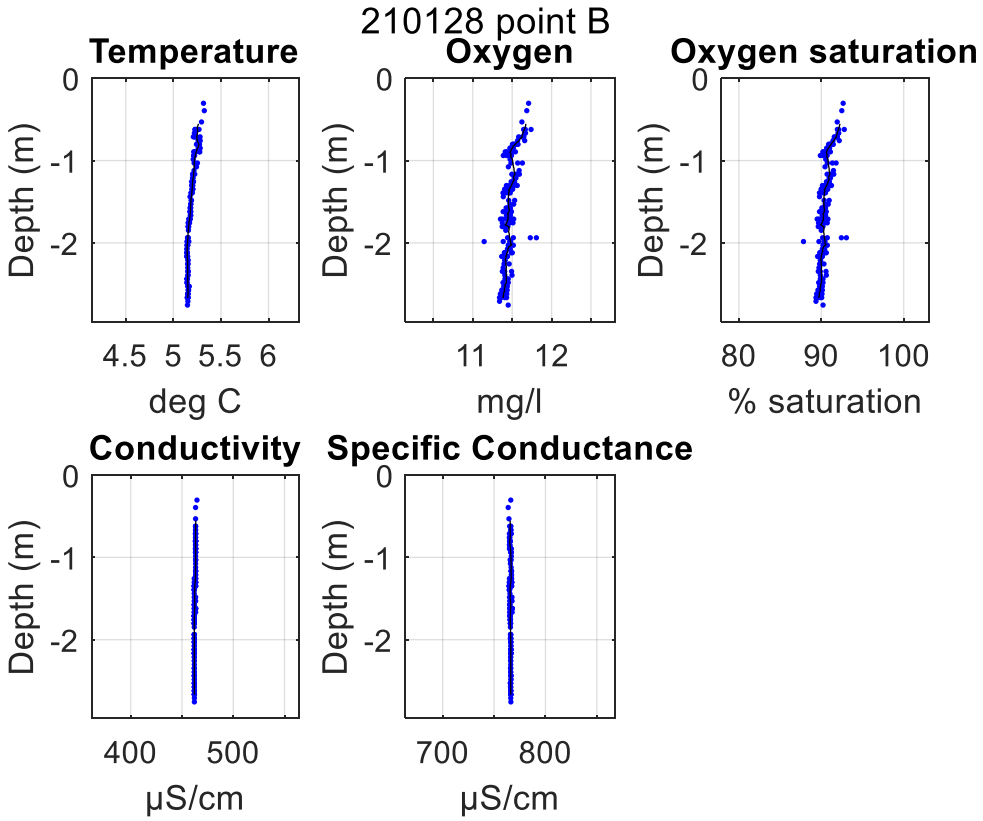


28/01/2021

Point A – 09:32 (UTC)

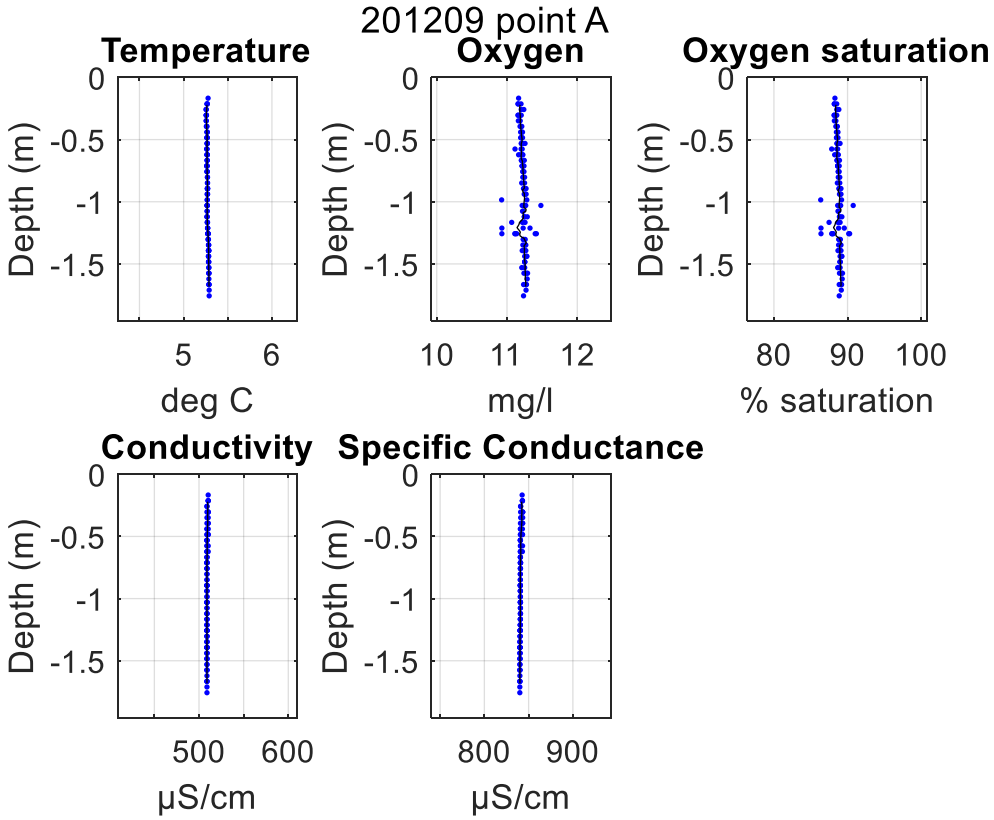


Point B – 10:40 (UTC)

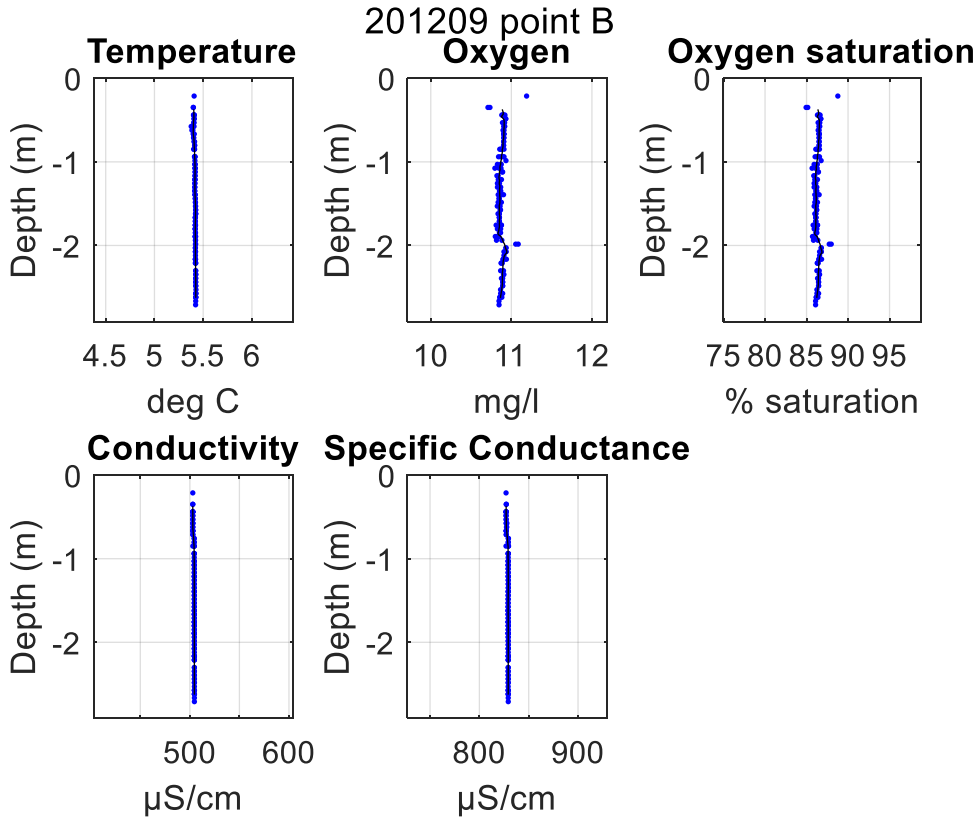


09/12/2020

Point A – 09:51 (UTC)



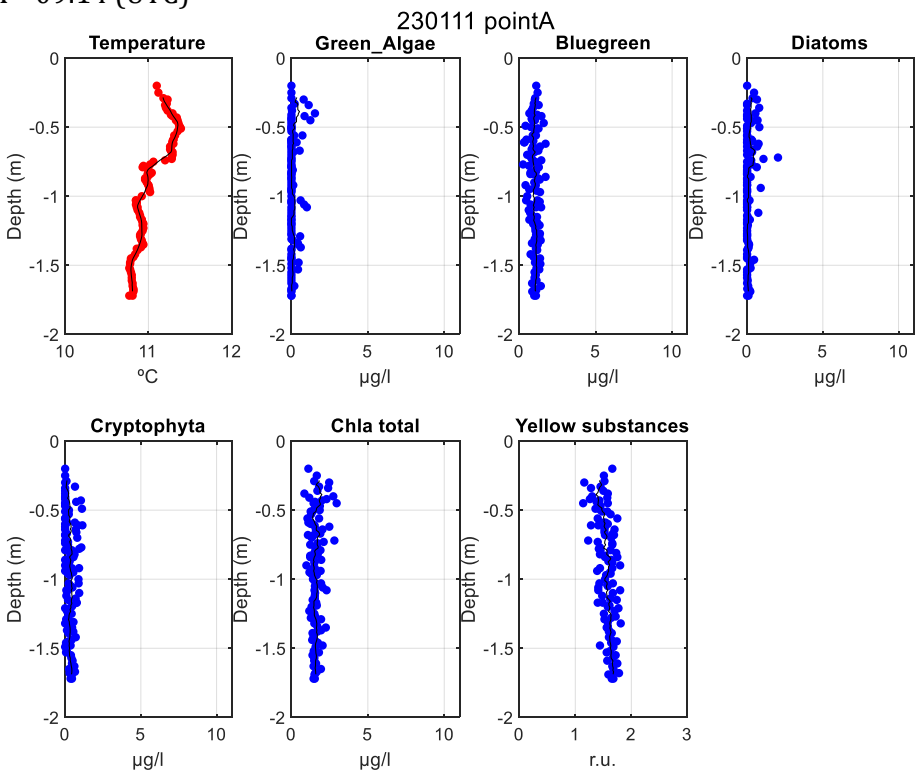
Point B – 10:27 (UTC)



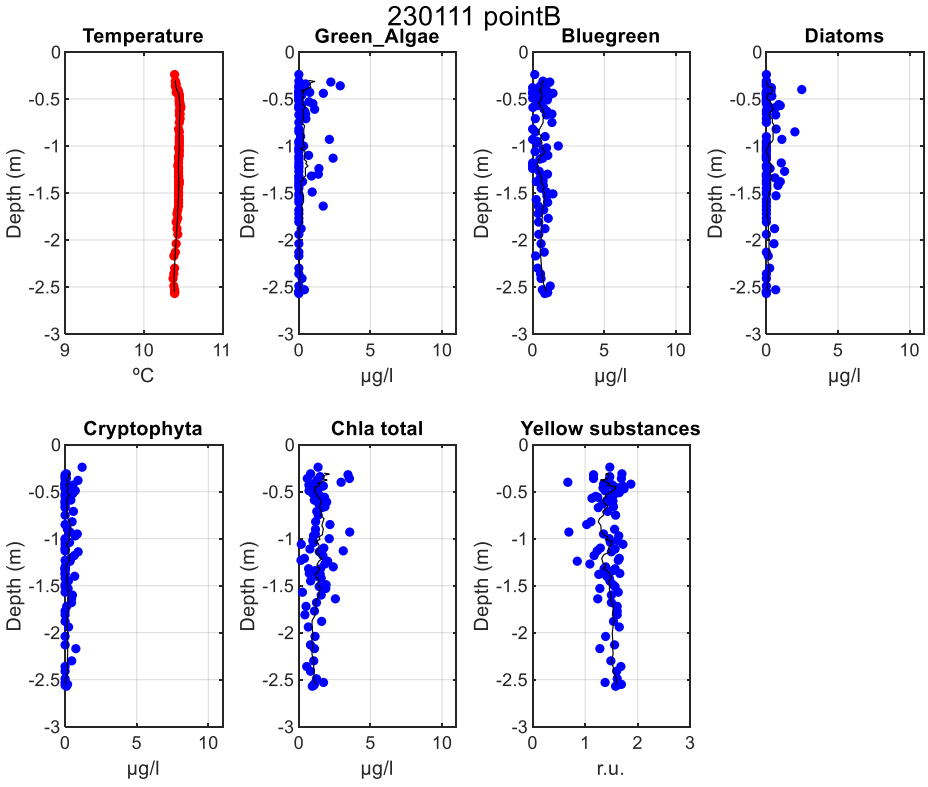
3. Fluoroprobe BBE profiles

11/01/2023

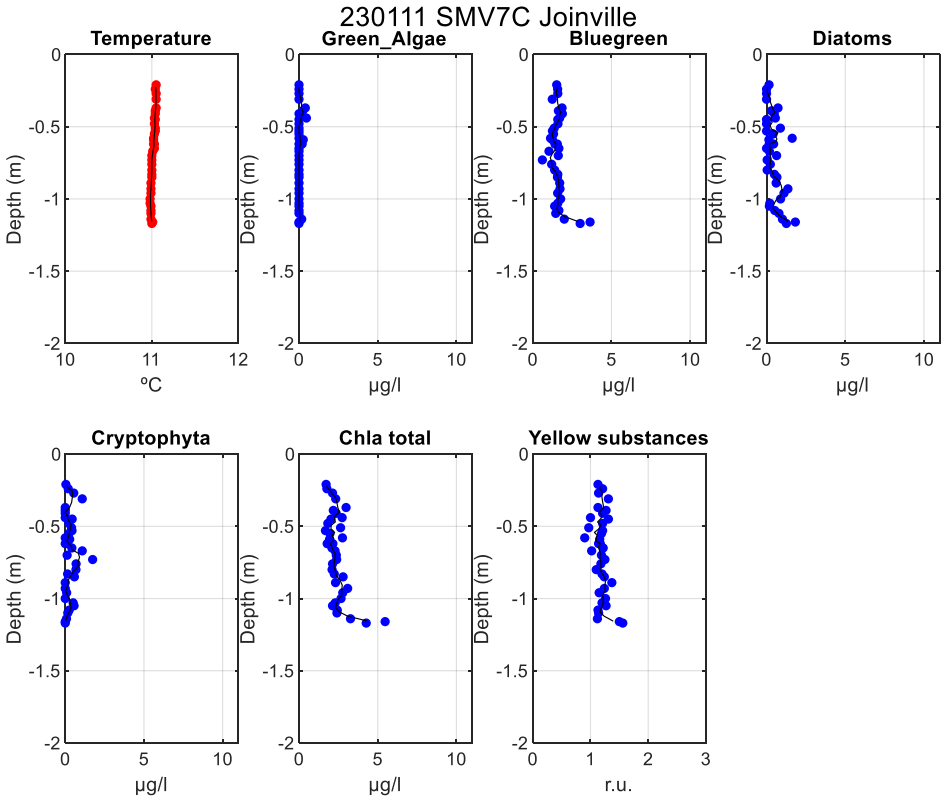
Point A – 09:14 (UTC)



Point B – 07:46 (UTC)

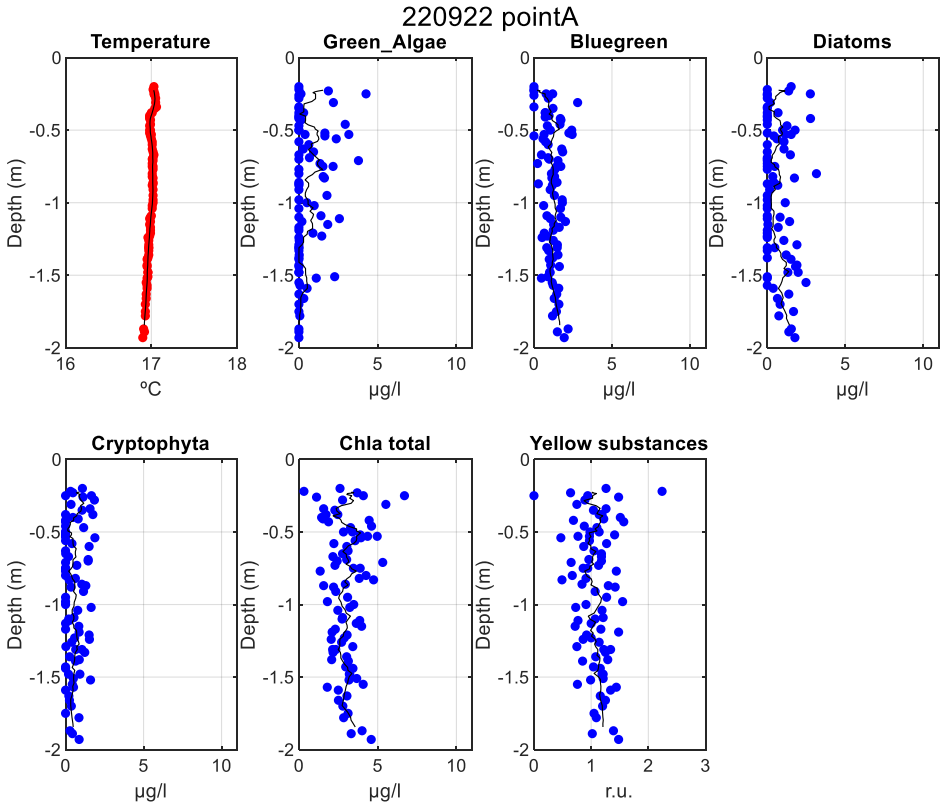


At the same day, measurement at Joinville (Marne), point SMV7C, at 13:25 (UTC):

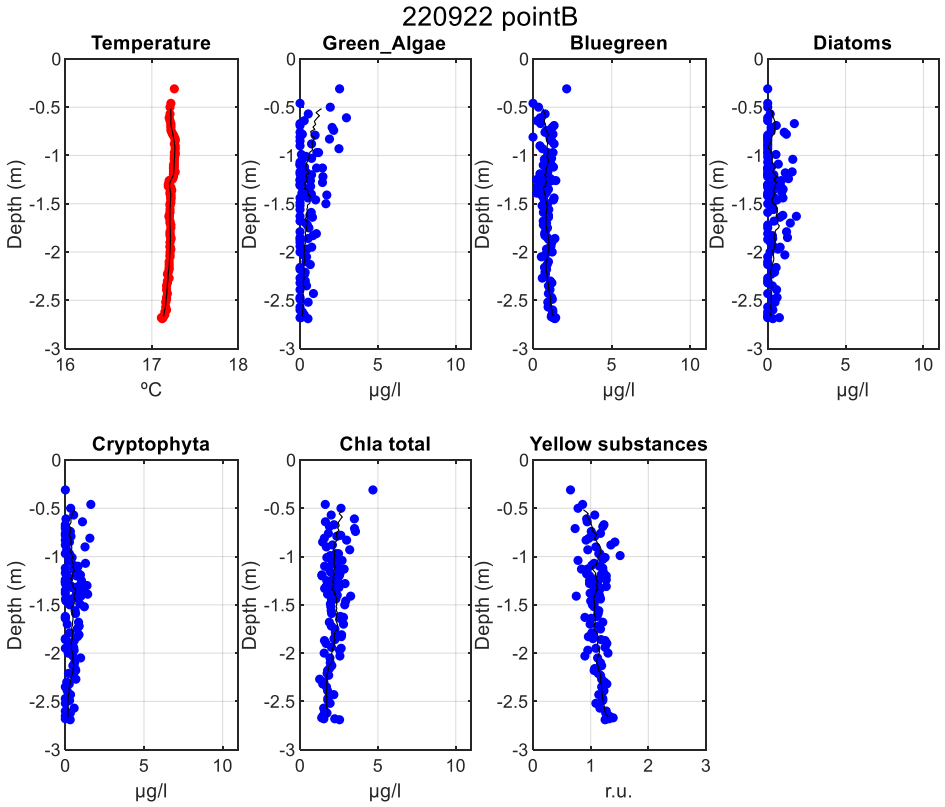


22/09/2022

Point A – 07:46 (UTC)

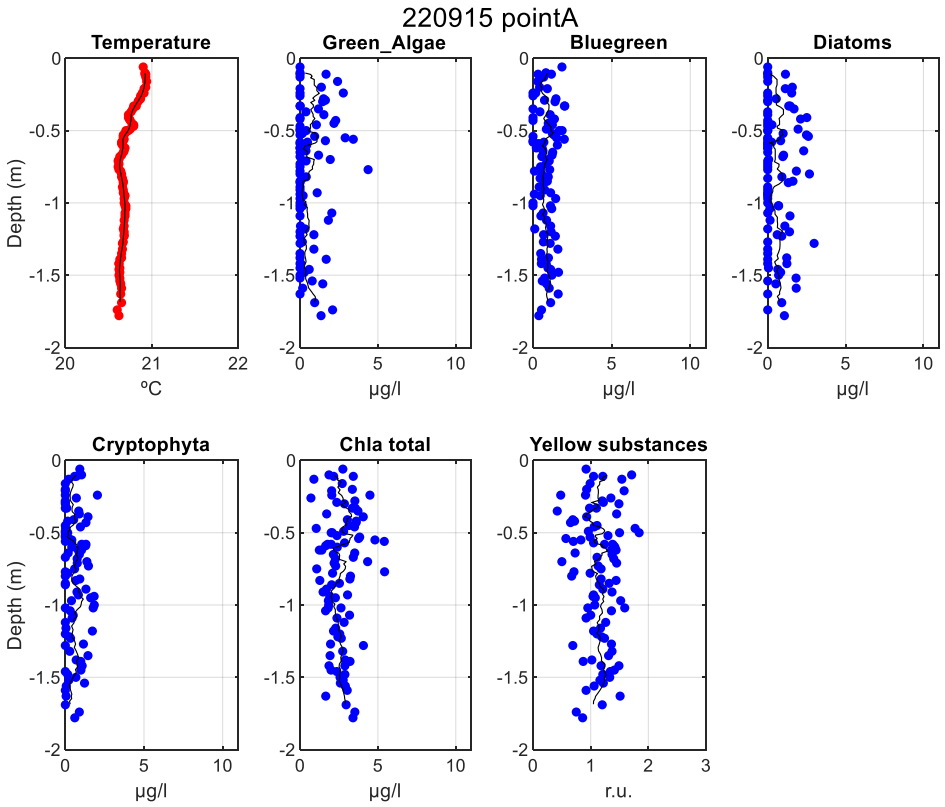


Point B – 08:37 (UTC)

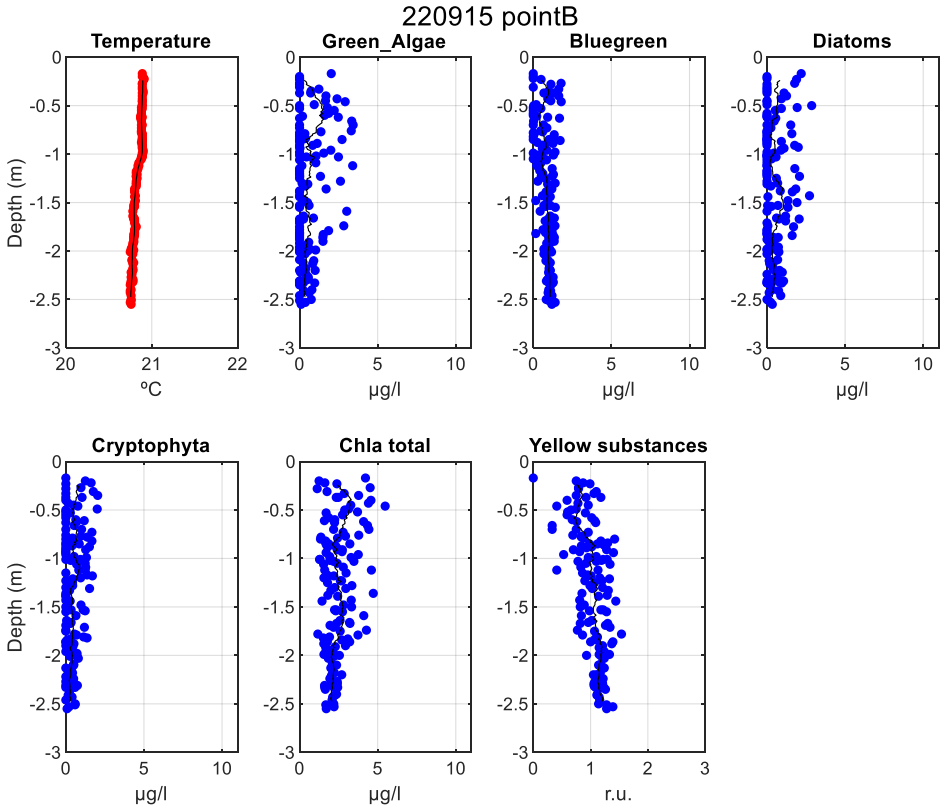


15/09/2022

Point A – 07:30 (UTC)

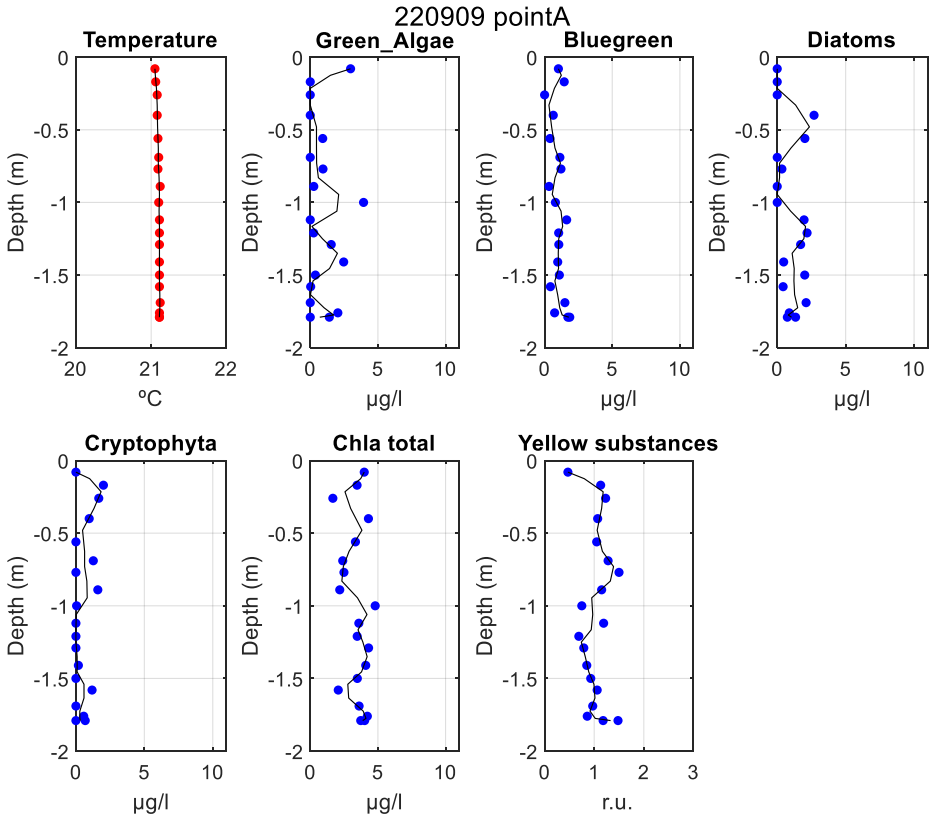


Point B – 08:41 (UTC)

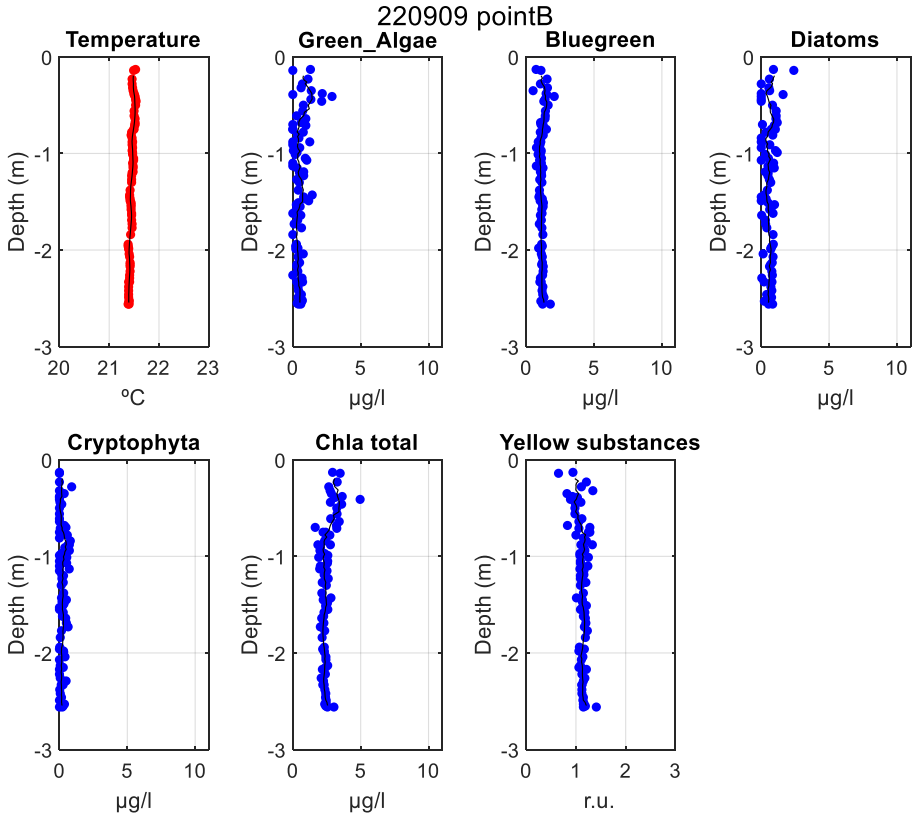


09/09/2022

Point A – 08:34 (UTC)

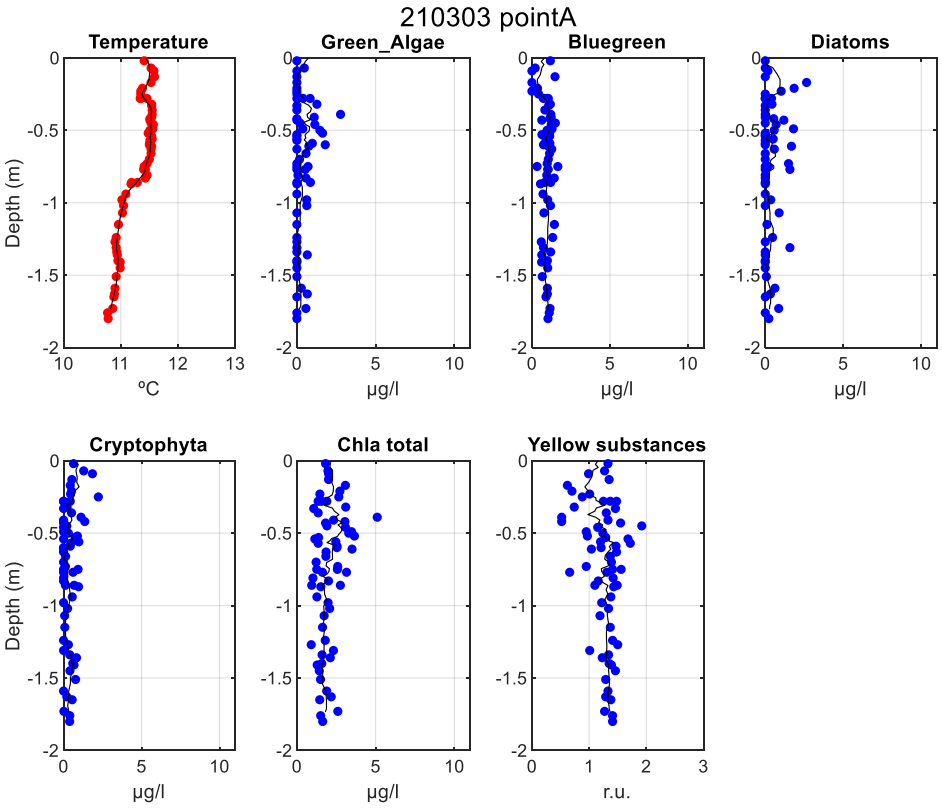


Point B – 07:47 (UTC)

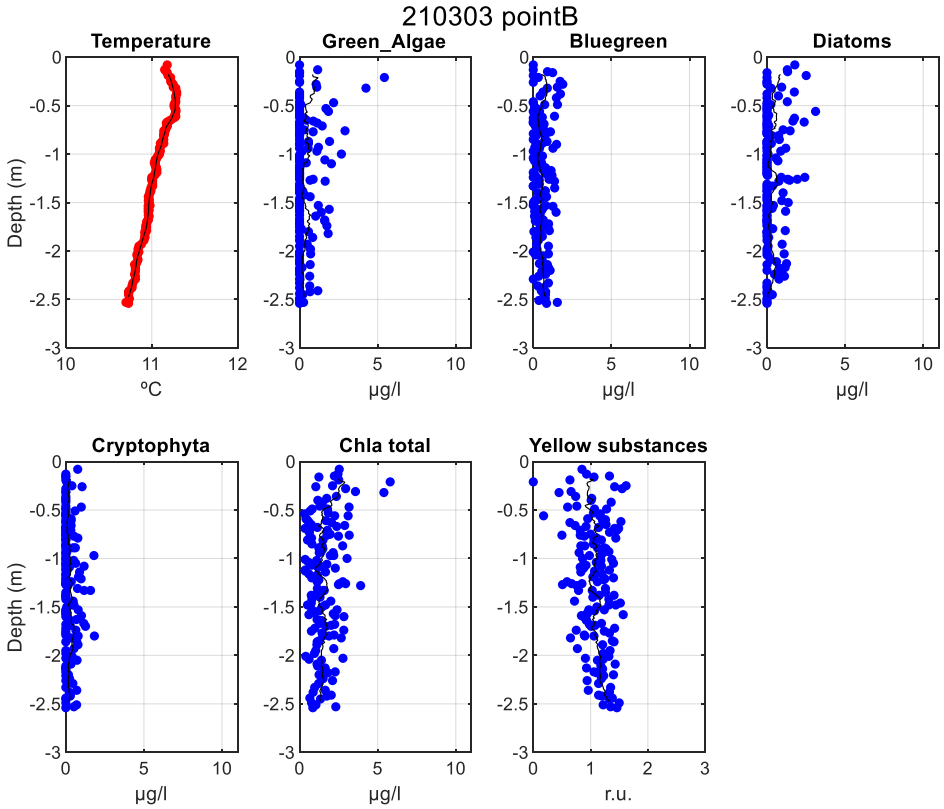


03/03/2021

Point A – 09:17 (UTC)

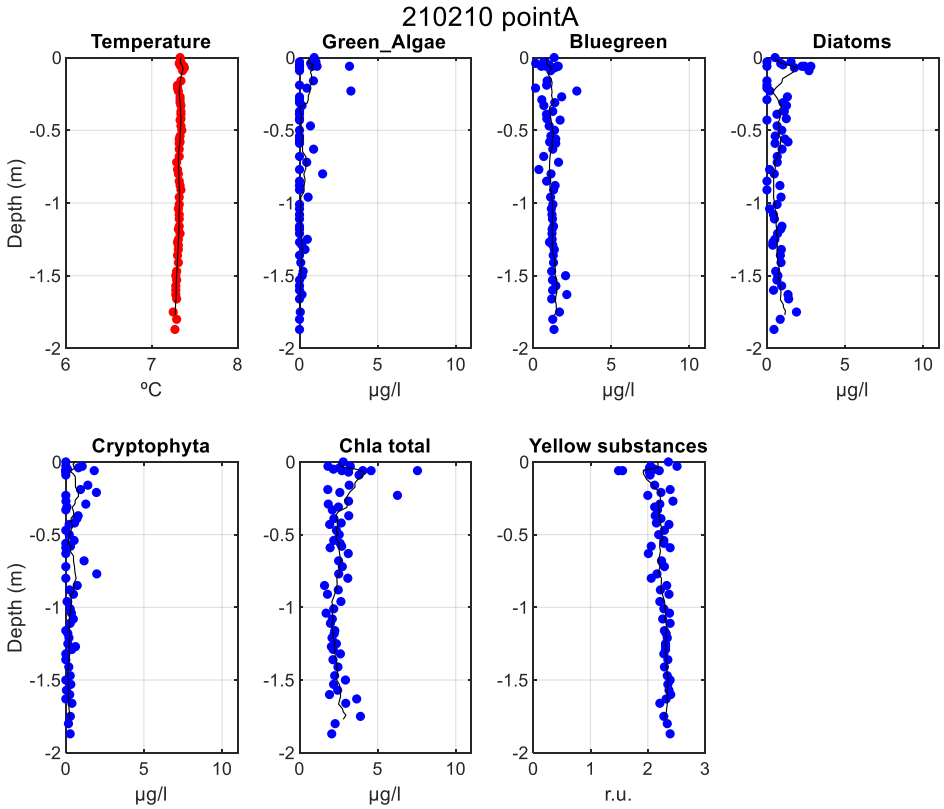


Point B – 10:24 (UTC) (previously point C)

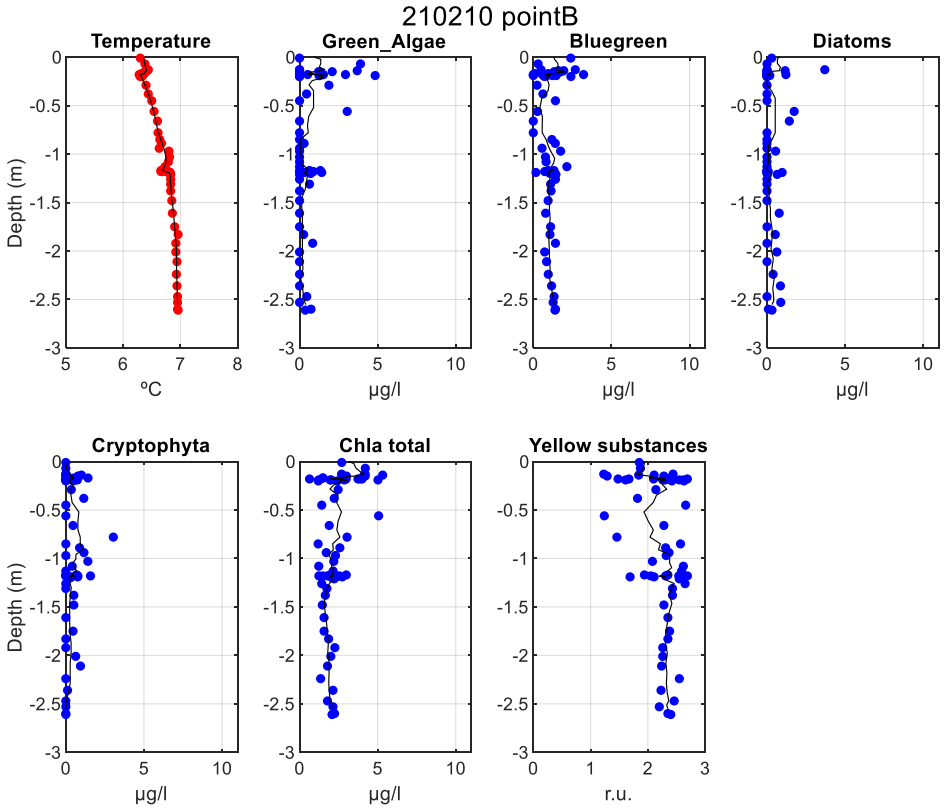


10/02/2021

Point A – 09:40 (UTC)

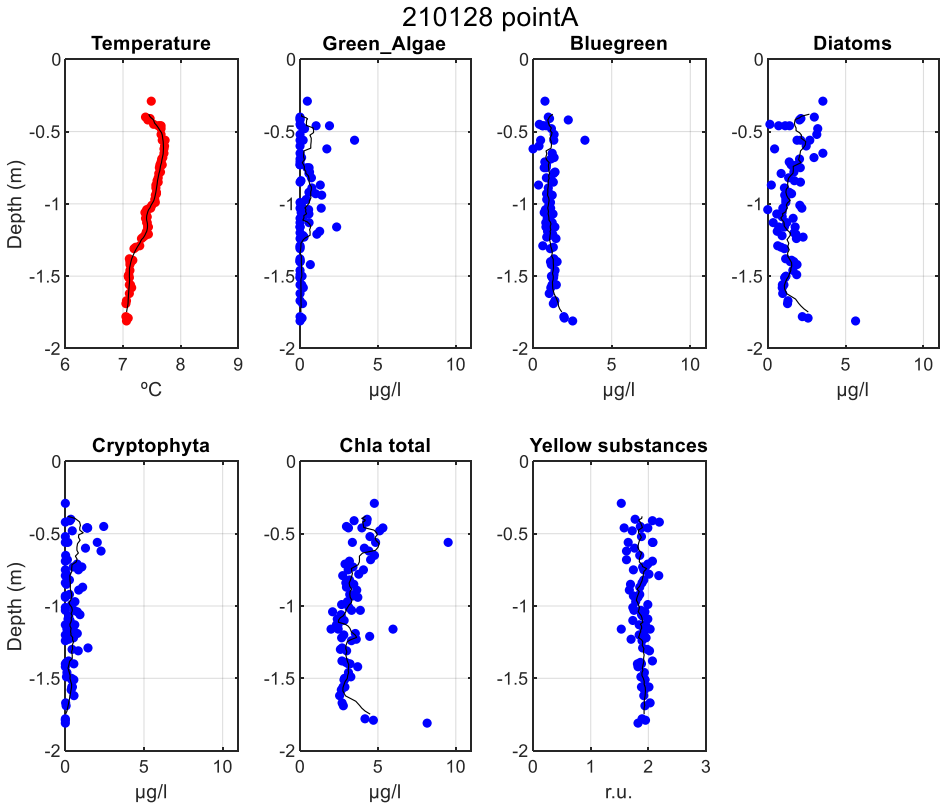


Point B – 10:40 (UTC) (previously point C)

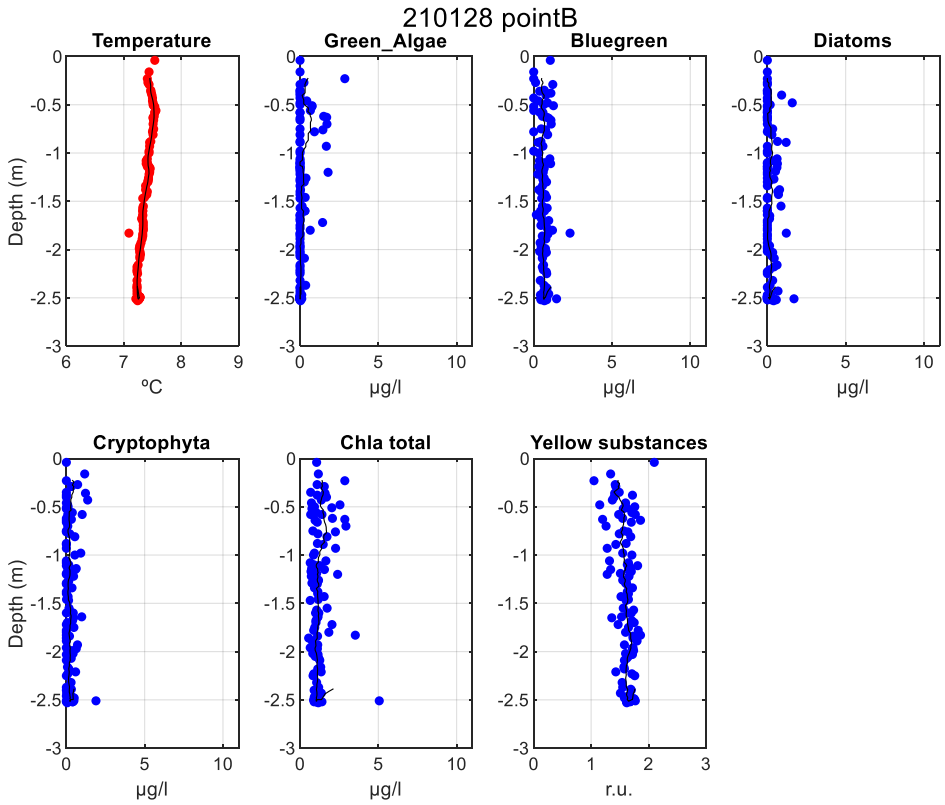


28/01/2021

Point A – 09:31 (UTC)

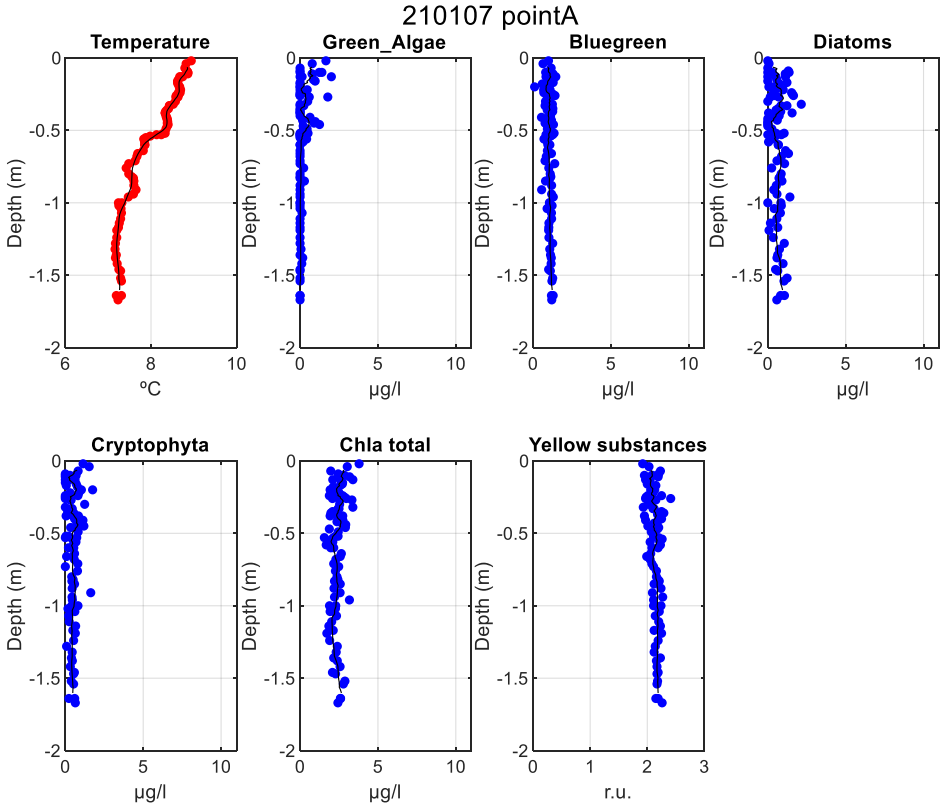


Point B – 10:31 (UTC) (previously point C)

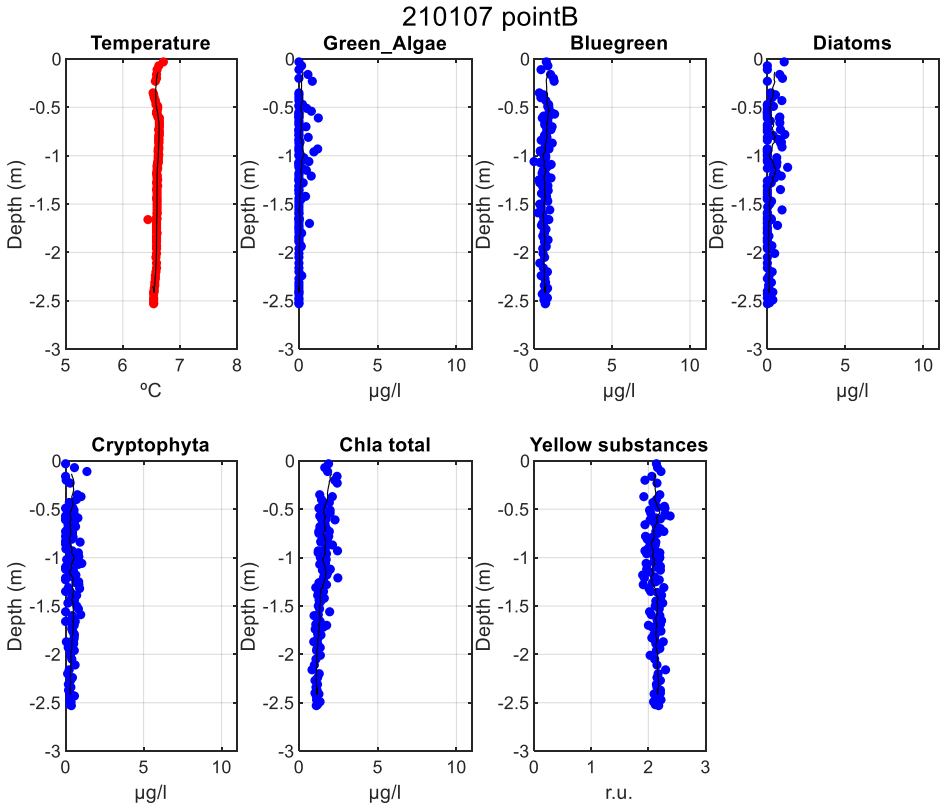


07/01/2021

Point A – 09:10 (UTC)

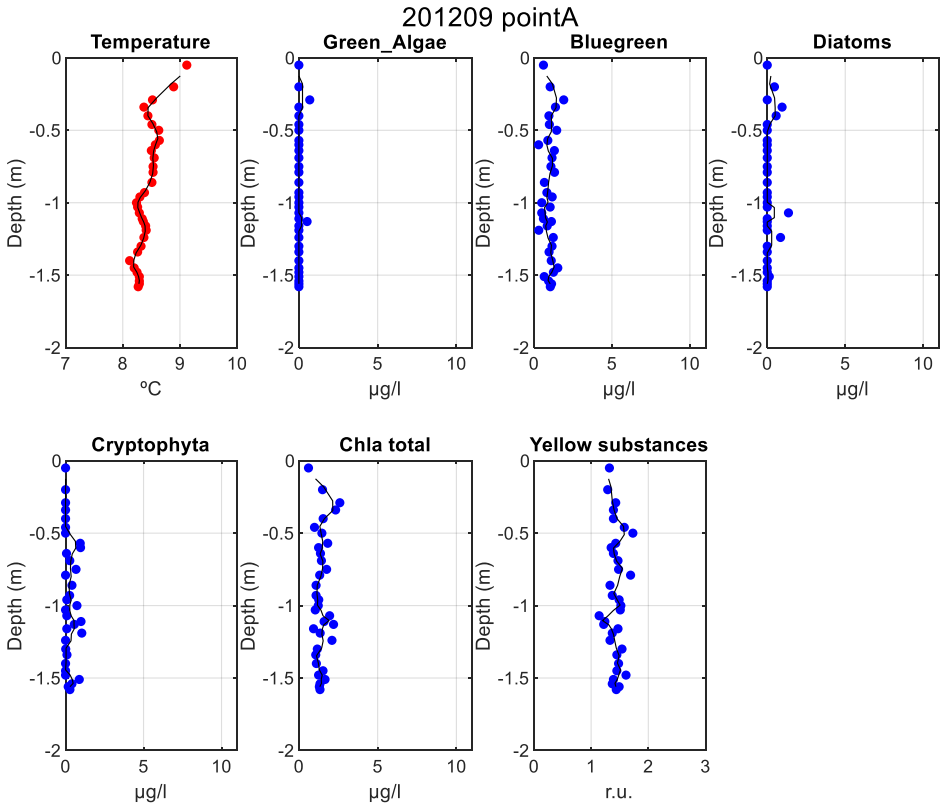


Point B – 09:55 (UTC) (previously point C)

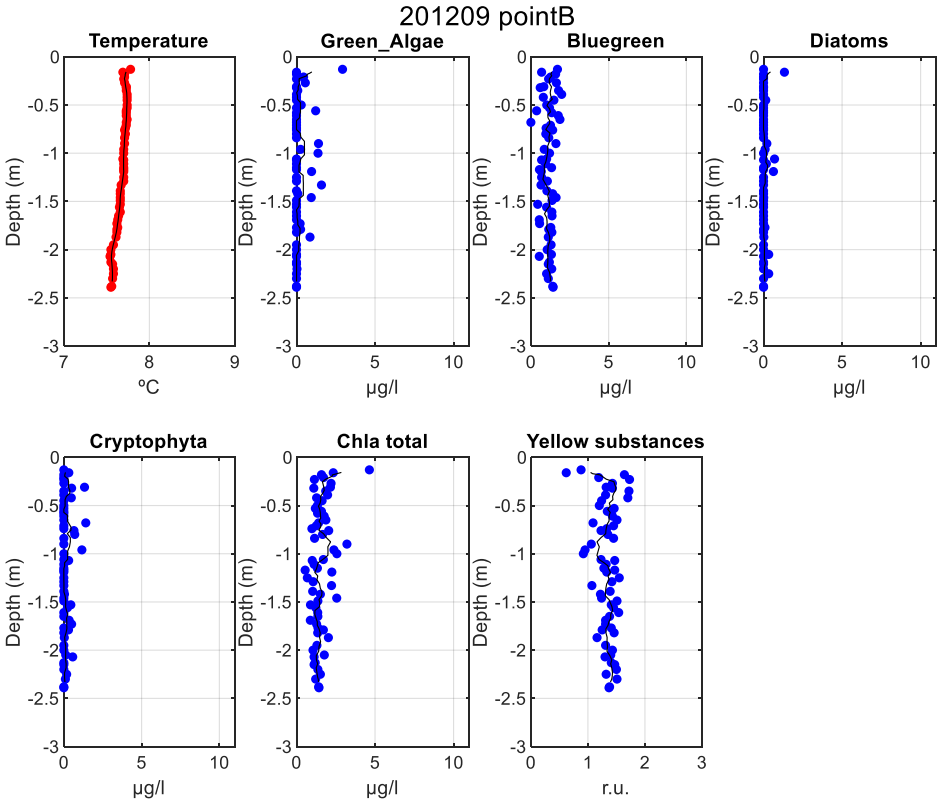


09/12/2020

Point A – 10:00 (UTC)

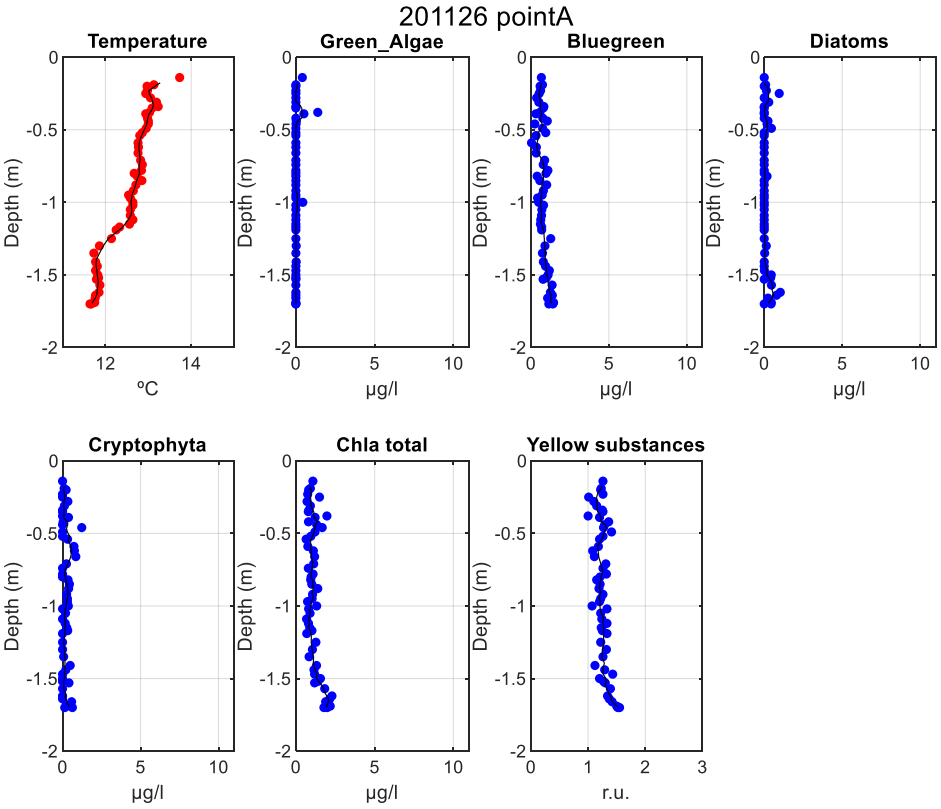


Point B – 10:30 (UTC) (previously point C)

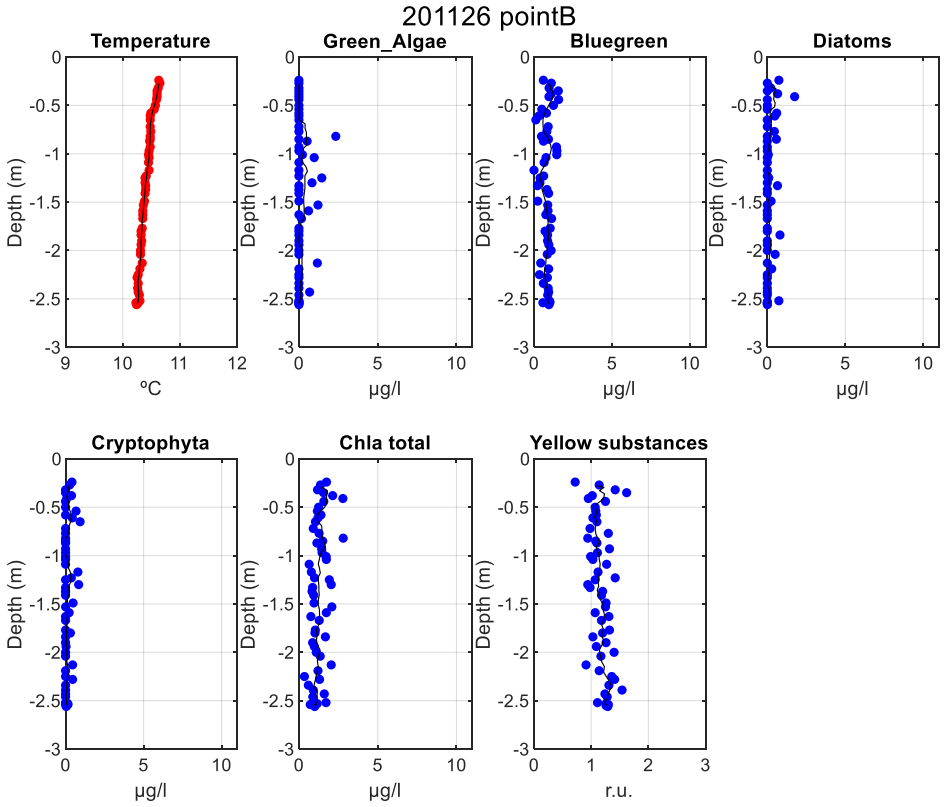


26/11/2020

Point A – 08:55 (UTC)

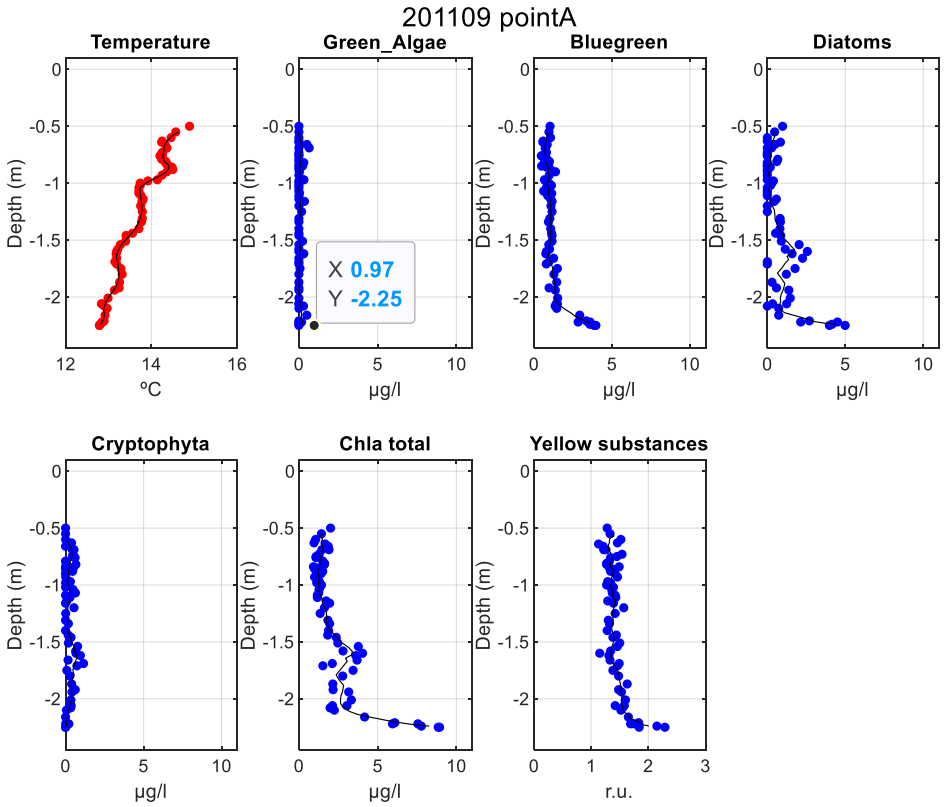


Point B – 11:00 (UTC) (previously point C)

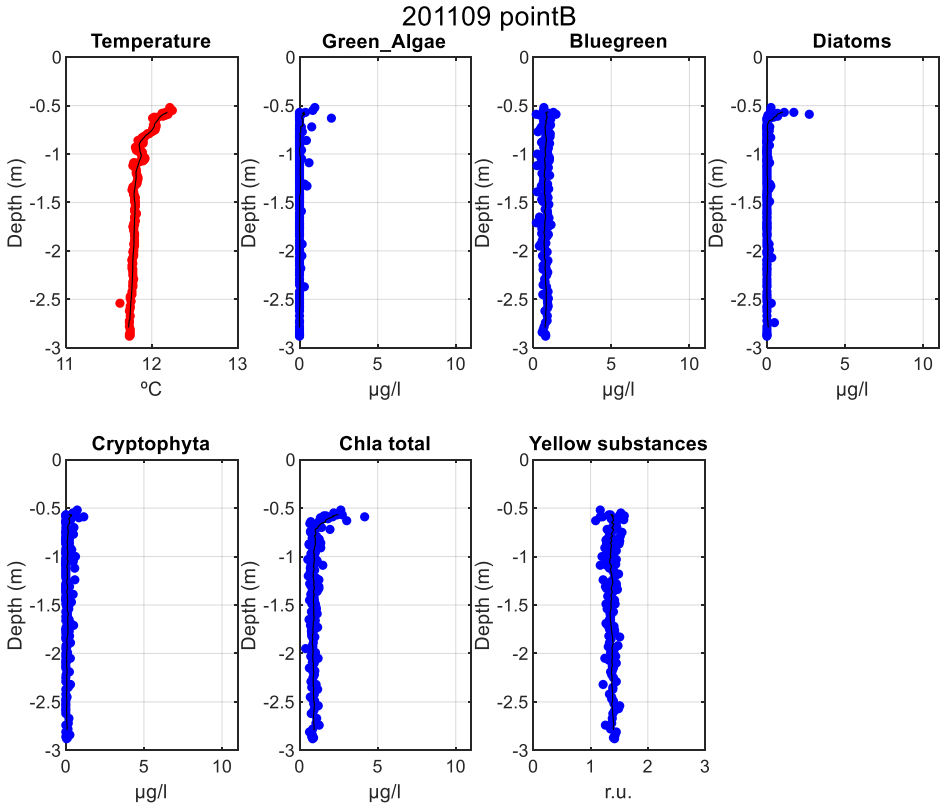


09/11/2020

Point A – 09:00 (UTC) – need of pression correction!

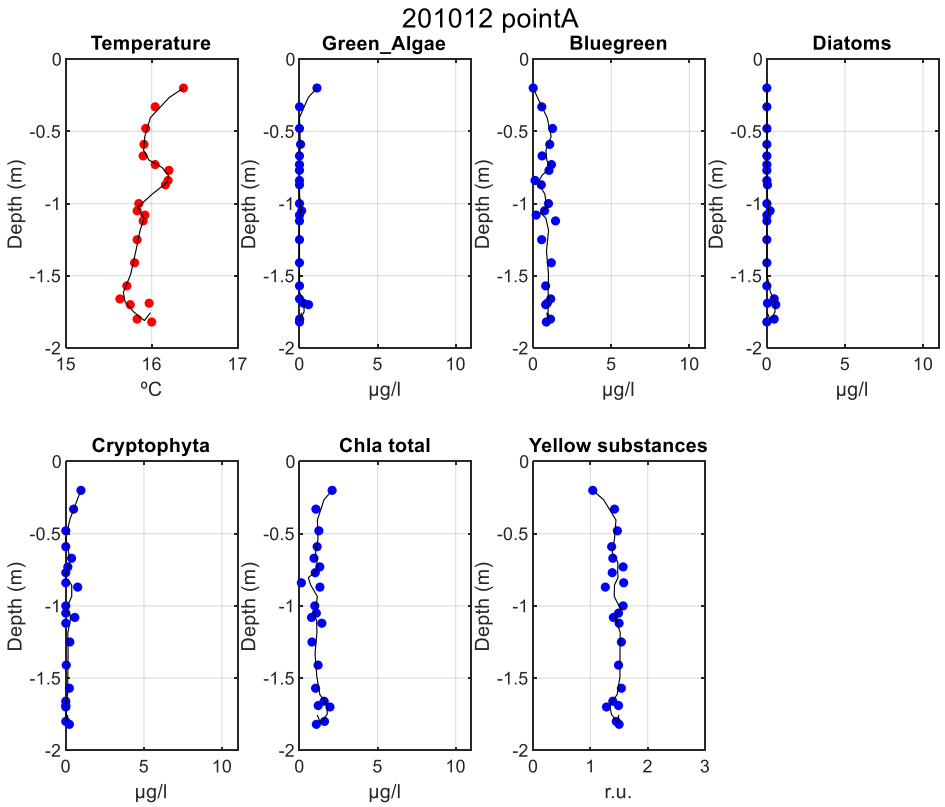


Point B – 09:42 (UTC) (previously point C)

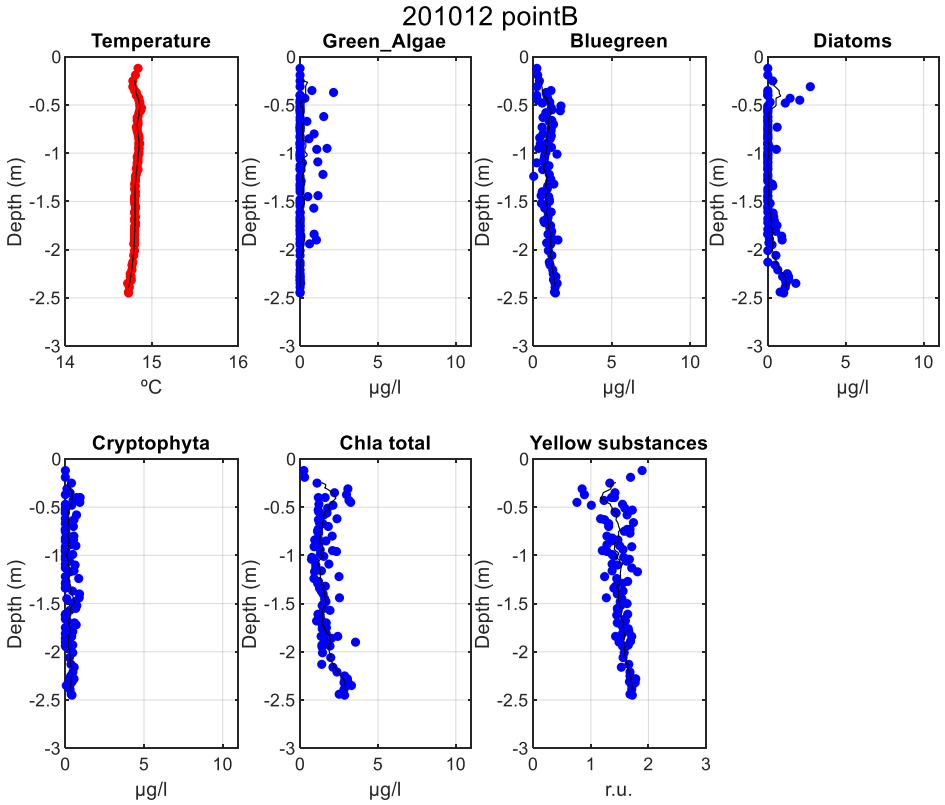


12/10/2020

Point A – 07:08 (UTC)



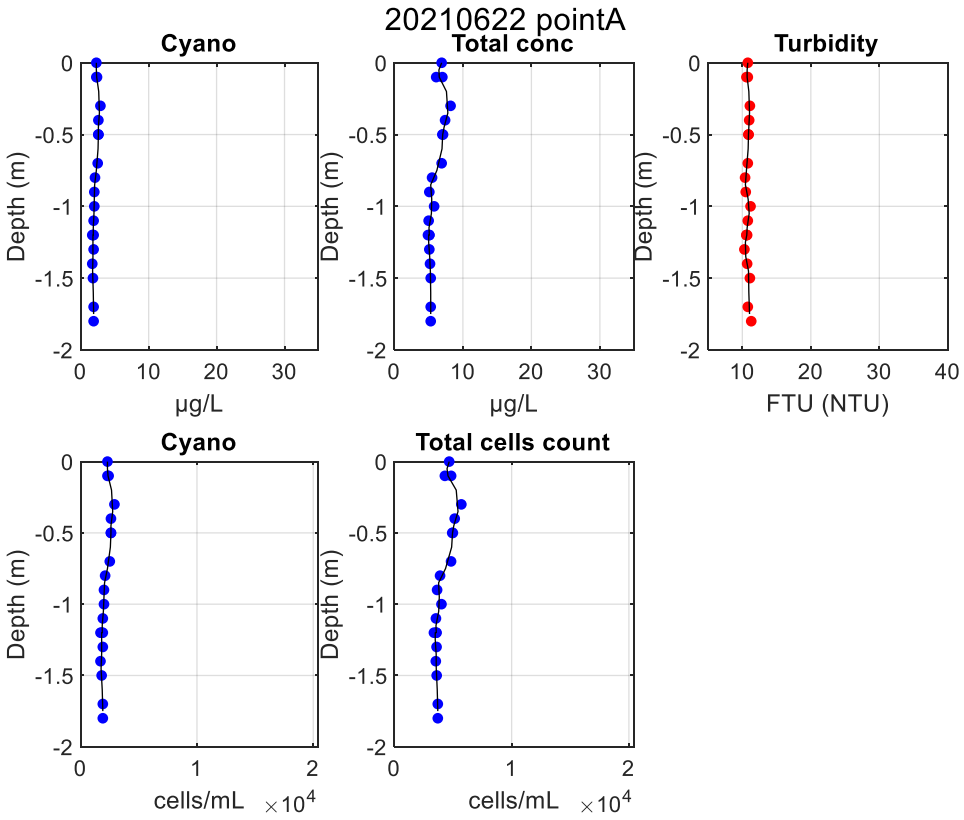
Point B – 08:10 (UTC) (previously point C)



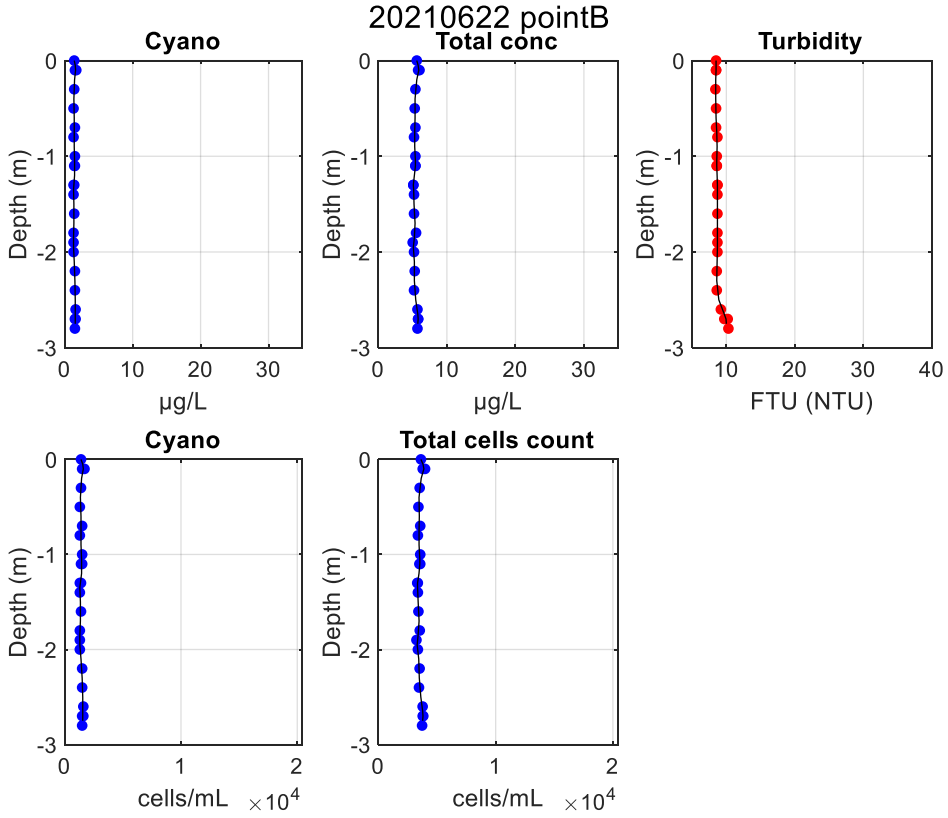
4. AlgaeTorch profiles

22/06/2021

Point A – 08:12 (UTC)

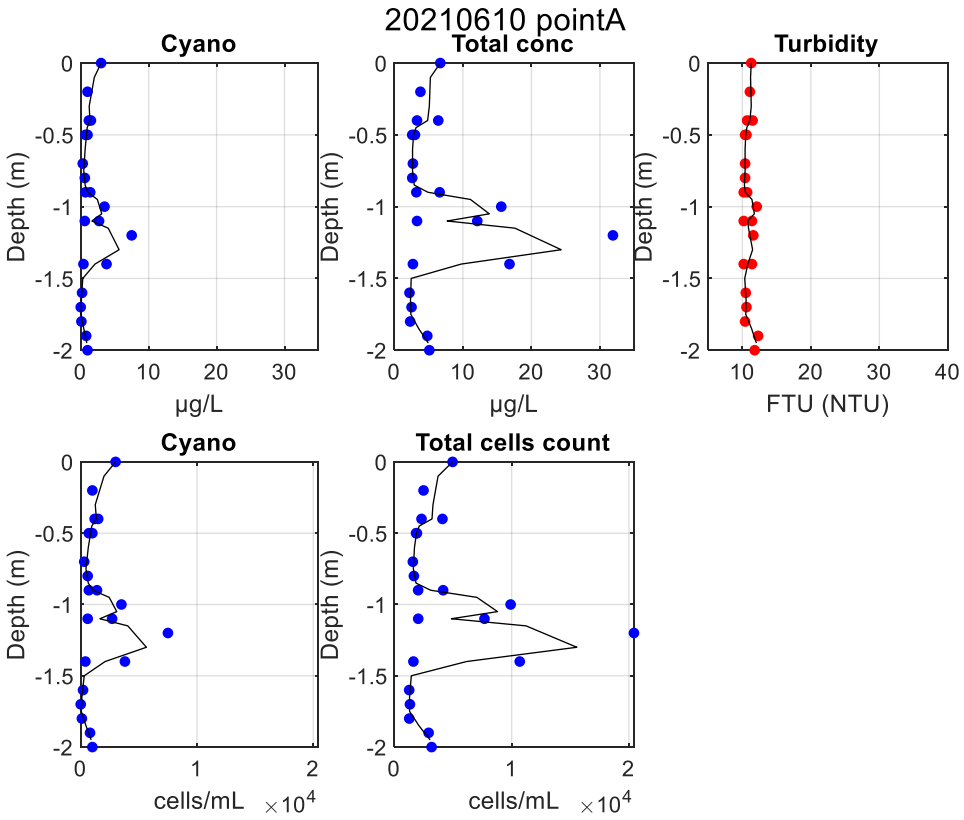


Point B – 09:07 (UTC)

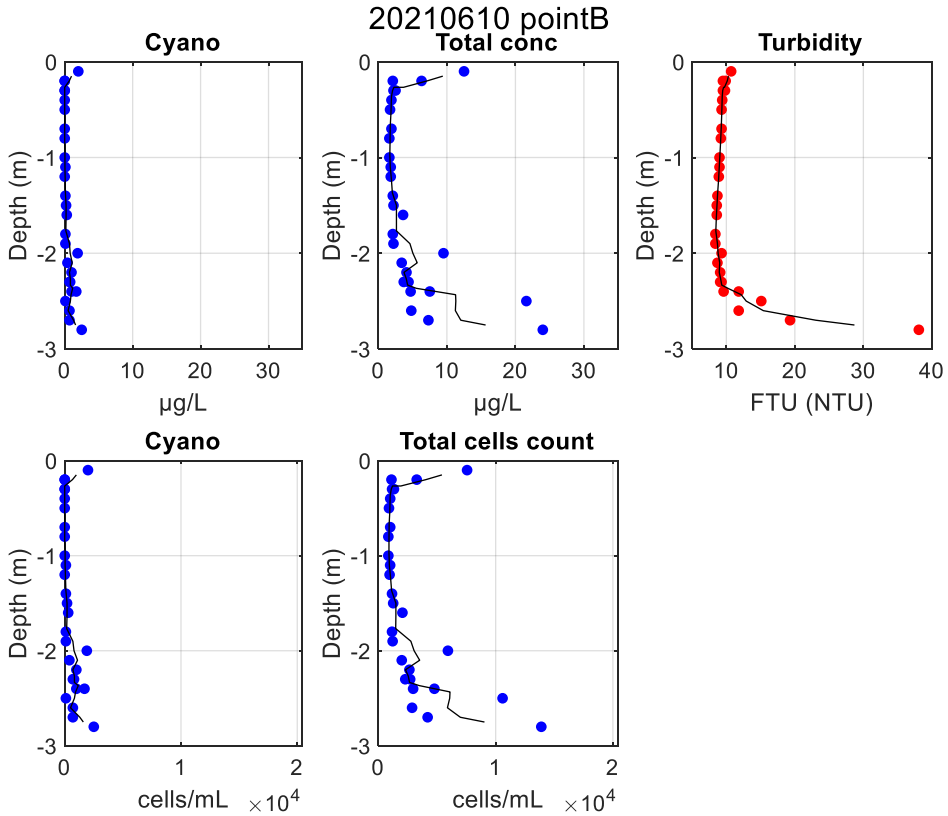


10/06/2021

Point A – 08:04 (UTC)

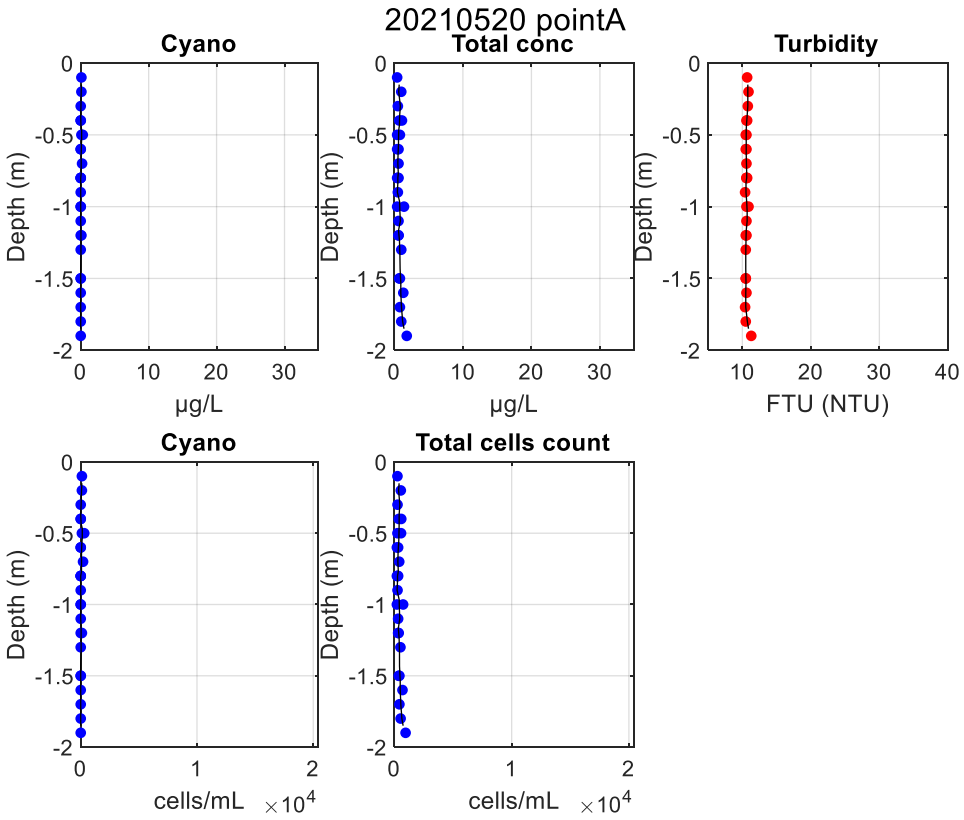


Point B – 09:13 (UTC)

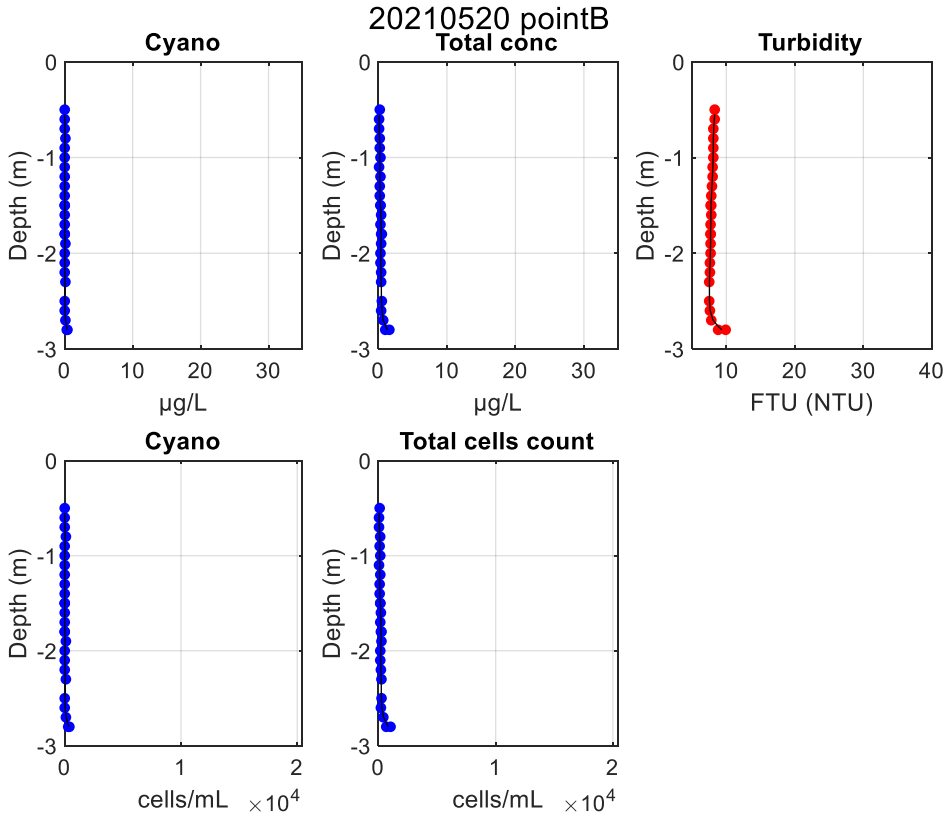


20/05/2021

Point A – 07:43 (UTC)

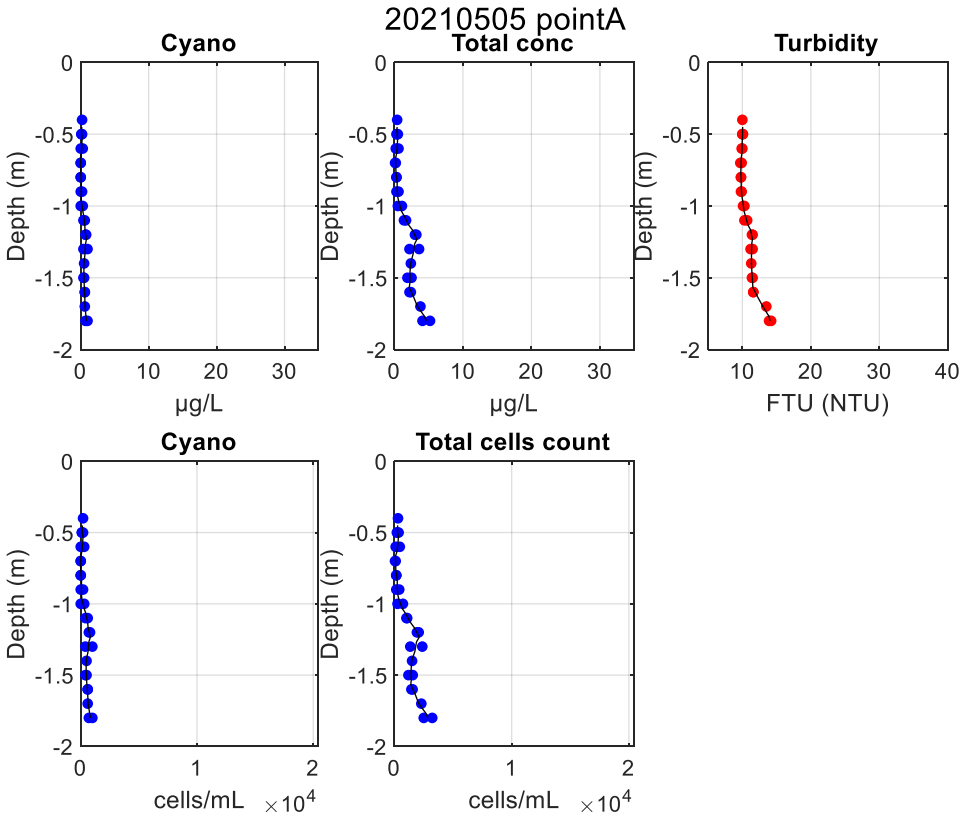


Point B – 09:14 (UTC)



05/05/2021

Point A – 12:12 (UTC)



Point B – not done.

IV-e. Validation between basic and complete versions of the monitoring system

The basic and the complete versions of the monitoring system were installed simultaneously, between May and June 2021 (at point A, from 20/05 to 10/06/2021; at point B, from 05 to 20/05/2021).

At point A, the relationship between water temperature measurements from nke (basic system) and Campbell (complete system) sensors, and conductivity from nke (basic system) and OTT (complete system) sensors measurements were very good, with RMSE=0.06°C and R^2=1.00 for water temperature, and RMSE = 19 µS.cm^{-1} and R^2=0.85 for conductivity (Figure A.9 and Table A.3).

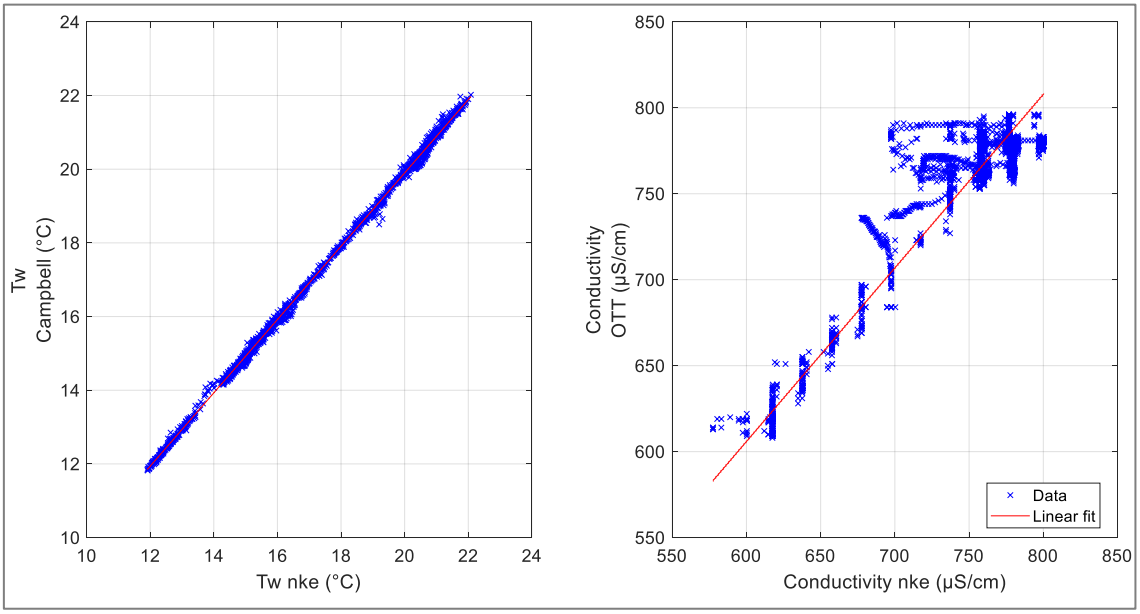


Figure A.9. Water temperature: nke and Campbell sensors measurements at the surface and conductivity measurements of nke (surface) and OTT (middle) sensors at Point A.

Table A.3. Relationship between measurements from nke, Campbell and OTT sensors at point A.

Point	Equipment	Variable	N. of points	Equation	R^2	RMSE
A surface	Nke and Campbell	Water temperature (°C)	5063	Tw_{nke}=1.00*$Tw_{campbell}$	1.00	0.06
A surface and middle	Nke and OTT	Conductivity (µS.cm^{-1})	3002	Cs_{nke} = 1.01 * Cm_{OTT}	0.85	18.80

At point B, the relationship between sensors nke (basic system) and Campbell (complete system) temperature measurements of the three depths were very good, with RMSE=0.02°C and R^2=1.00 for all depths (Figure A.10 and Table A.4).

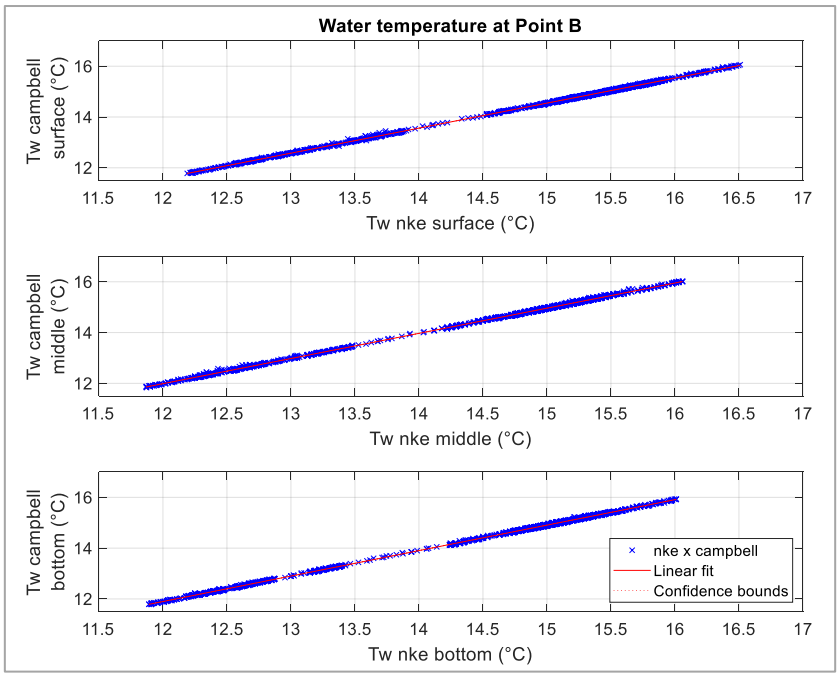


Figure A.10. Water temperature nke and Campbell sensors measurements (surface, middle and bottom) at point B.

Table A.4. Relationship between measurements from nke and Campbell (cb) sensors at point B.

Point	Equip-ment	Variable	N. of points	Equation	R^2	RMSE
B surface	Nke and Campbell	Water temperature (°C)	2113	Tws_{nke}=0.99*Tws_{cb}-0.24	1.00	0.02
B middle	Nke and Campbell	Water temperature (°C)	2113	Twm_{nke}=0.99*Twm_{cb}+ 0.09	1.00	0.02
B bottom	Nke and Campbell	Water temperature (°C)	2113	Twf_{nke}=1.00*Tws_{cb}-0.14	1.00	0.02

V. Submitted article: “Fluorescence spectroscopy for tracking microbiological contamination in urban water bodies”

Fluorescence spectroscopy for tracking microbiological contamination in urban waterbodies

Natália Angelotti de Ponte Rodrigues[1,2*], Rémi Carmigniani[2], Arthur Guillot - Le Goff[1,2], Françoise S. Lucas[1], Claire Therial[1], Manel Naloufi[1], Aurélie Janne[3], Francesco Piccioni[4], Mohamed Saad[1], Philippe Dubois[1], Brigitte Vinçon-Leite[1*]

[1] LEESU, Ecole des Ponts, Univ Paris Est, Creteil, Marne-la-Vallée, France

[2] LHSV, Ecole des Ponts, EDF R&D, Chatou, France

[3] Syndicat Marne Vive, Saint-Maur-des-Fossés, France (currently at Agence Eau Seine Normandie, Courbevoie, France)

[4] PROLOG INGENIERIE, Paris, France

*** Correspondence:**
Corresponding Authors
natalia.angelotti-de-ponte-rodrigues@enpc.fr; b.vincon-leite@enpc.fr

Keywords: Fluorescent dissolved organic matter (FDOM), Parallel factor analysis (PARAFAC), DOM signature, urban waters, urban swimming, Faecal indicator bacteria (FIB).

This paper was written in British English.

Abstract

Dissolved organic matter (DOM) plays a crucial role in freshwater ecosystem function. Monitoring of DOM in aquatic environments can be achieved by using fluorescence spectroscopy. Particularly, DOM fluorescence can constitute a signature of microbiological contamination with a potential for high frequency monitoring. However, limited data are available regarding urban waterbodies. This study considers fluorescence data from field campaigns conducted in the Paris metropolitan region: two watercourses (La Villette basin and the river Marne), two stormwater network outlets (SO), and a wastewater treatment plant effluent (WWTP-O). The objectives of the study were to characterize the major fluorescence components in the studied sites, to investigate the impact of local rainfall in such components and to identify a potential fluorescence signature of local microbiological contamination. The components of a PARAFAC model (C1-C7), corresponding to a couple of excitation (ex) and emission (em) wavelengths, and the fluorescence indices HIX and BIX were used for DOM characterization. In parallel, faecal indicator bacteria (FIB) were measured in selected samples. The PARAFAC protein-like components, C6 (ex/em of 280/352 nm) and C7 (ex/em of 305/340 nm), were identified as markers of microbial contamination in the studied sites. In the La Villette basin, where samplings covered a period of more than two years, which also included similar numbers of wet and dry weather samples, the protein-like components were significantly higher in wet weather in comparison to dry weather. A positive relationship was obtained between C6 and FIB. In urban rivers, the high frequency monitoring of C6 levels would support the faecal contamination detection in rivers. In addition, it could help targeting specific field campaigns to collect comprehensive dataset of microbiological contamination episodes.

1 Introduction

In freshwaters, dissolved organic matter (DOM) plays a key role in ecosystem functioning. Part of the DOM pool is associated to bacterial degradation and nutrient availability (Wetzel, 1992; Boavida and Wetzel, 1998), acting as a supplier of biologically accessible organic substances that sustains heterotrophic aquatic organisms. DOM may also attenuate light intensity (Morris *et al.*, 1995; Markager and Vincent, 2000), mobilize and transport contaminants (Polubesova *et al.*, 2007; Mladenov *et al.*, 2010), and affect the concentration of dissolved oxygen in the water (McCabe *et al.*, 2021). Changes in DOM composition may reflect the variation of the status of an ecosystem caused by land use and climate change (Wilson and Xenopoulos, 2009; Williams *et al.*, 2010).

In urban areas, water sports and recreative activities in open waters are increasingly popular. However, these activities can expose users to waterborne diseases when the water quality is poor due to faecal pollution. The microbiological quality of water bodies is assessed through microorganism indicators of faecal contamination. In the USA, the Clean Water Act (USEPA, 1977), and in Europe, the European Bathing Water Directive (BWD, 2006/7/EC), require the monitoring of faecal contamination through faecal indicator bacteria (FIB). *Escherichia coli* (*E. coli*) and *intestinal enterococci* (IE) are well established indicators and are recommended for freshwater microbiological quality assessment (USEPA, 1986). In faecal contaminated waters, *E. coli* is present in greater concentrations than IE. Edberg *et al.* (2000) recommended to use *E. coli* as the best indicator, with IE as a complementary one. However, the standard methods of FIB enumeration require more than 24 hours to provide results. For a reactive contamination detection, a fast-monitoring technique is required.

Fluorescence spectrophotometry has been widely used for characterizing DOM in surface waters (Cumberland et al., 2012; Romero González-Quijano et al., 2022), ground waters (Frank *et al.*, 2018; Sorensen *et al.*, 2020), marine waters (Parlanti *et al.*, 2000; Kim *et al.*, 2020), and for microbial quality assessment (Baker *et al.*, 2015; Sciscenko *et al.*, 2022). It can be used to assess DOM composition, DOM source, type, and microbiological activity (Coble, 1996).

Fluorescent dissolved organic matter (FDOM) contains molecular groups called fluorophores which, when excited by specific UV wavelengths emit fluorescence (Fellman *et al.* 2010). The fluorophores identified in freshwaters are associated with different types of dissolved organic matter and can be mostly attributed to humic-like (peaks A, C and M) and protein-like (peak T) components. Peak A is associated with terrestrial components. Peak C is often linked to anthropogenic activities such as agriculture. Peak M is associated with marine humic-like components. Peak T may indicate microbial activity (Coble, 1996).

Fluorescence intensity ratios can be used to assess the origin and transformation degree of DOM. Three fluorescence indices are commonly used: BIX, HIX and FI. The biological index (BIX) was introduced by Huguet *et al.* (2009), based on the freshness index (ratio β:α) from Parlanti *et al.* (2000). BIX is an indicator of the contribution of recently autochthonous produced DOM. It represents more recently derived DOM, including those freshly transferred from algae to the DOM pool, and the more decomposed DOM.

The humification index (HIX) is related to the extent of humification of organic matter. It was introduced by Zsolnay *et al.* (1999), denoted $HIX_{Zsolnay}$, who assumed that humification is associated with an increase of molecular weight. The emission spectra of these more condensed fluorescing molecules tend to shift to longer emission wavelengths. It was used to estimate the degree of DOM

aromatization in soil, and firstly applied to aquatic systems by Huguet *et al.* (2009). It is the ratio of the sum of the fluorescence intensities between the emission wavelengths of 435 nm and 480 nm and the sum of the intensities emitted between 300 nm and 345 nm, at excitation wavelength of 254 nm.

Ohno (2002) proposed a modified HIX calculation, denoted HIX_{Ohno}, introducing a second inner-filter effect correction. With this new ratio, HIX ranges from 0 to 1, also increasing with the increase of DOM aromatization degree. Since HIX values from Ohno's expression are independent of the concentrations of humic-like substances, it facilitates the comparison of results across different studies. In this paper, both HIX expressions were compared.

The fluorescence index (FI) was first introduced by McKnight *et al.* (2001). It is also known as f_{450}/f_{500} index, used by Huguet *et al.* (2009), and later modified by Cory and McKnight (2005). It allows to determine the source of DOM either microbial (≥1.8) or terrestrial (≤1.4) (Fellman *et al.*, 2010).

Independent fluorescence components, which can be associated to different origins of DOM, can be obtained through the application of the parallel factor (PARAFAC) method to fluorescence excitation-emission matrices (EEMs). PARAFAC is a powerful tool on the characterization and quantification of DOM in different aquatic environments (Stedmon and Bro, 2008).

In urban waterbodies, during rain events, wastewater discharge and runoff affect the water quality and change the DOM composition (Xenopoulos *et al.*, 2021). Humic-like components are associated with terrestrial organic matter such as run-off water, and protein-like components are associated with a high level of microbial activity, such as wastewater. However, few field data are available in urban waters.

The main objective of this study was to use the fluorescence spectroscopy for the characterisation of DOM in urban waters with contrasting origins and FIB levels. The variation of fluorescence indices and PARAFAC components was analysed in an EEM dataset of 414 samples. The components from PARAFAC model, specific to our dataset, were associated to the reference fluorophores from Coble (1996), peaks T, M, C and A, for characterization.

Fluorescence data were collected in waterbodies located in the Paris metropolitan region, France. The sampling sites are as follows: two watercourses, La Villette basin and the river Marne; two stormwater network outlets (SO); one wastewater treatment plant outlet (WWTP-O). Samples were collected during dry and wet weather. In the La Villette basin, regular field campaigns were conducted for 29 months (October 2020-March 2023), in three sampling points. In the river Marne, samples were collected in 19 different points along a stretch of around 30 km for one summer (2022).

In the samples from SO and WWTP-O, and in the samples from one sampling point from La Villette basin and one from the river Marne, the enumeration of faecal indicator bacteria (FIB) *E. coli* and intestinal enterococci (IE) were conducted. The fluorescence intensities of all samples were normalized in Raman Units and the fluorescence indices were computed.

The specific objectives of the study were:

i. Identifying the major fluorescence components in different investigated urban waters.
ii. Examining spatial variation of FDOM components and the impact of local rainfall episodes in these components.

iii. Exploring the information provided by FDOM about microbiological water quality for monitoring purposes and identifying a potential fluorescence signature of local contamination due to SO and WWTP effluents or runoff during rainfall events.

2 Material and methods

2.1 Sampling sites

The study sites are located in the metropolitan area of Paris, France. Two watercourses, the river Marne and La Villette basin, two SO and one WWTP-O were sampled (Figure 1, Supplementary Table S1).

2.1.1 Watercourses

La Villette waterbody, Paris (48.8918 N, 2.3855 E), is 1500 m long and 3 m deep with a mean flowrate of 2 m^3s^{-1}. A bathing area is located downstream during summer since 2017. Three sampling points, A, B and C were located at the inlet of the canal (A), 1000 m downstream, in the bathing area (B) and C between points A and B (Figure 1). The point A is 9.7 km downstream a SO and 20 km downstream the WWTP of Villeparisis, which has a mean discharge of 0.13 m^3s^{-1} (*Portail assainissement collectif*, 2021). The samples were collected from October 2020 to March 2023, twice a month in average (85 campaigns in total).

The studied stretch of the river Marne is around 30 km long (48.8758 N 2.6721 E upstream, 48.8163 N 2.4414 E downstream). During the studied period (15/06 to 20/10/2022) the mean flowrate was around 36 m^3s^{-1} (HydroPortail, 2022). The samples were collected, in 19 sampling points along the river stretch. The point VGA is located in Saint-Maur-des-Fossés (VGA sailing club). The other 18 points are part of *Syndicat Marne Vive* monitoring program (points SMV). Among the SMV points, 6 correspond to cities candidates to open bathing sites as a legacy of the Paris Olympic Games in 2024. They were considered as "main points" due to a three times higher sampling frequency than the other 12 minor sampling points (Figure 1).

In this river stretch, two outlets of wastewater treatment plants (WWTP) are located. The most upstream outlet is the WWTP St-Thibault-des-Vignes (hereafter WWTP-O), with 44 000 m^3 wastewater treated per day (0.46 m^3s^{-1}) (*Portail assainissement collectif*, 2021). It discharges at 7.8 km upstream point SMV1. The second outlet is WWTP Paris Marne Aval (WWTP-PMA), with 35 000 m^3 wastewater treated per day (0.41 m^3s^{-1}) (*Portail assainissement collectif*, 2021). It is located in Noisy-le-Grand, but it discharges at 1.8 km upstream point SMV10, in Saint-Maur-des-Fossés, on the second half of the studied stretch of the river Marne (Figure 1).

The land use data of the river Marne sampling sites was obtained from the open data platform from *Institut Paris Region* (2022). An area of 1 km^2 immediately upstream the sampling points was considered to study the influence of their immediate surroundings. Due to the lack of samples from the SMV minor sampling points (n=4 at each point), only the 6 major points were explored (Supplementary Figure S1). In the Marne SMV main sampling points, the main land use was impervious area, with average of 65%.

The distance between the sampling points (Supplementary Table S1) and the proportional land-use areas (Supplementary Figure S2, Table S2) were calculated using an open-source geographic information system software (QGIS 3.30).

The sampling weather conditions were obtained from Météo-France stations: for La Villette, it was the Le Bourget weather station (48.96722 N, 2.42778 E, 9 km to the north), and for the Marne, the Torcy weather station (48.86361 N, 2.65111 E, 5 km to the east from SMV1).

2.1.2 Stormwater network and WWTP outlets

Two stormwater outlets (points SO-N and SO-S, Figure 1) were sampled. In both catchments, with a surface area respectively around 500 ha and 1400 ha, the drainage network is separated and discharges into the river Marne. The sampling was conducted at the outlet of their respective catchment. At SO-N, the water is directly released into the river Marne. At SO-S, downstream the sampling point, the water is screened before going through two settling tanks and being released into the river Marne.

Samples of the treated water of WWTP St-Thibault were collected before its release into the river Marne (point WWTP-O in Figure 1). This WWTP is approximately 8 km upstream the SMV1 point with no intermediate WWTP.

2.2 Water sampling

Samples were collected during dry and wet weather. An overview of the water sampling is summarized in Supplementary Table S3. Wet weather is defined as a rainfall episode with precipitation height higher than 5 $mm.day^{-1}$ over 3 days. Two types of sample collection were used: grab samples and 24-hour integrated samples. The grab samples were collected at a punctual moment in time. The integrated samples were collected for 24 hours and mixed, representing a mean sample of the collected period.

For the grab sampling, 100 mL of water were manually collected at 30 cm below the water surface at a pre-established schedule, covering wet and dry weather, in HDPE (high density polyethylene) flasks. The flasks were triple rinsed with the water from the site before sampling. At La Villette, grab samples were collected at points A and B, twice a month from October 2020 to March 2023 (n=85 in each point). At point C, four grab samples were collected: one in 22/06/2021 and three in 16/12/2021. At the river Marne (SMV), grab samples were collected during summer 2022, three times a week in average (n=189). Six of the SMV sampling points formed a main group, where sampling collection was 3 times a week. A total of 4 cruises were conducted in all 18 SMV points on the same day.

The 24-hour samples were collected by an automatic sampler, 30 cm below the surface, at an hourly time step (1 L each), during wet and dry weather. All the samples were stored at 5 °C and manually mixed on a 25 L container before being submitted to the lab analysis. 24-hour samples were collected at La Villette point C (n=10), the river Marne point VGA (n=18), the stormwater outlet points SO-N (n=9) and SO-S (n=8) and the WWTP outlet WWTP-O (n=5).

Summarizing all sampling sites, a total of 414 samples were collected (Supplementary Table S3). All samples were covered in aluminium foil for sunlight protection and stored at 4 °C until being submitted to analysis in the laboratory within 24 h. Only the turbidity of samples from the studied SO (points SO-N and SO-S) were higher than 10 NTU. In order to avoid interference on fluorescence and absorbance measurements (Bedell *et al.*, 2022), the samples were filtered with a 0.45 µm glass microfiber filter (Whatman® Cat No 1825 047). Due to high concentration of FDOM, samples from points SO-N, SO-S and WWTP-O were diluted with ultra-pure water with dilution factor of 1/5.

2.2.1 Faecal indicator bacteria measurement

FIB are represented by *E. coli* and IE. Sample analysis included *E. coli* and IE measurement of the 24-hour integrated samples from Villette C (n=10), Marne VGA (n=18), SO-N (n=9), SO-S (n=8) and WWTP-O (n=5), within 8 h.

From the 50 integrated samples, *E. coli* and IE were measured according to ISO-Norms ISO 9308-3 (1998) and ISO 7899-1 (1998), respectively. The results are expressed as the most probable number per 100 mL (MPN.100mL^{-1}).

2.3 Fluorescence measurements and fluorescence indices

Fluorescence measurements of all samples were performed on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies) with a 10 mm quartz cell, at the laboratory temperature (around 20 °C). A simultaneous scan of excitation wavelength (ex) from 200 to 450 nm with 5-nm intervals and emission wavelength (em) from 250 to 550 nm with 2-nm intervals was performed, with a scan rate of 1200 nm.min^{-1}.

The absorbance was obtained with a Lambda 35 spectrophotometer (PerkinElmer). The samples were analysed in a 10 cm quartz cell and scanned from 200 to 750 nm with 1-nm intervals, at a scan speed of 240 nm.min^{-1}. In both fluorescence and absorbance analysis, an ultra-pure water sample blank was analysed in the beginning of each series of measurements. The cell was tripled-rinsed with ultra-pure water and with the water sample between measurements.

The processing of the Excitation-Emission Matrix (EEM) spectra was conducted according to Murphy *et al.* (2013), using the drEEM toolbox for MATLAB® R2022b (MathWorks®, USA). The absorbance was used for inner filter effect correction (Parker and Barnes, 1957). The values were normalized to Raman Units (RU) according to Lawaetz and Stedmon (2009). The sample EEMs were subtracted by the corresponding blank EEM, removing Rayleigh and Raman scatters. The fluorescence intensities were linearly interpolated to a resolution of 1 nm in excitation and emission wavelengths to calculate and extract peaks and indices. Subsequently, they were smoothed with the Savitzky-Golay method, with a span of 21 data points and polynomial of degree 2.

From the corrected EEMs, the intensities of the fluorophores of interest (Coble, 1996) were obtained. Peak T is the intensity at wavelength excitation (ex) 275 nm and wavelength emission (em) 340 nm. Peak M is the maximum intensity at ex 312 nm and in the interval em 380-420 nm. Peak C is the maximum intensity in the intervals ex 320-360 nm and em 420-460 nm. Peak A is the maximum at ex 260 nm and em 400-460 nm.

In addition, several spectral indices were computed and used to describe DOM composition. The biological index (BIX) was computed at an excitation wavelength of 310 nm (Equation 1), according to Huguet *et al.* (2009).

$$BIX_{ex\,310} = \frac{I_{em\,380}}{I_{em\,430}} \quad (1)$$

Where $I_{em\,380}$ and $I_{em\,430}$ are the fluorescence intensities at 380 nm and 430 nm emission wavelength, respectively.

The humification index (HIX) was calculated at an excitation wavelength of 254 nm (Equation 2), according to Ohno (2002).

$$HIX_{Ohno\,ex\,254} = \frac{\sum(Iem\ 435 \rightarrow 480)}{\sum(Iem\ 300 \rightarrow 345) + \sum(Iem\ 435 \rightarrow 480)} \quad (2)$$

Where $\sum(Iem\ 435 \rightarrow 480)$ is the sum of the fluorescence intensity between 435 nm and 480 nm of emission wavelength, and $\sum(Iem\ 300 \rightarrow 345)$ is the sum of the fluorescence intensity between 300 nm and 345 nm emission wavelength. This index was compared with the HIX expression given by Zsolnay *et al.* (1999) (Equation 3). For brevity, when not specified, HIX will correspond to HIX_{Ohno}.

$$HIX_{Zsolnay\,ex\,254} = \frac{\sum(Iem\ 435 \rightarrow 480)}{\sum(Iem\ 300 \rightarrow 345)} \quad (3)$$

The fluorescence index (FI) was calculated as the ratio between the fluorescence intensity at excitation wavelength of 370 nm (Equation 4), according to Fellman *et al.* (2010).

$$FI_{ex\,370} = \frac{I_{em\,470}}{I_{em\,520}} \quad (4)$$

Where $I_{em\,470}$ and $I_{em\,520}$ are the fluorescence intensities at 470 nm and 520 nm emission wavelength, respectively.

The locations of the fluorophores of interest (peaks T, M, C and A) and of the terms used to obtain the fluorescence indices BIX, HIX and FI on a fluorescence EEM are indicated in Supplementary Figure S3.

2.4 PARAFAC model

PARAFAC (PARAllel FACtor analysis) is a statistic tool applicable to data arranged in three (or more) arrays. The method is based on a generalization of Principal Component Analysis (PCA) to higher order arrays (Murphy *et al.*, 2013). It can be used to decompose a EEM dataset into independent and distinct components that can be overlapped and combined to form the original dataset. The data signal is decomposed into a set of trilinear terms and a residual array, as in Equation 5.

$$x_{ijk} = \sum_{f=1}^{F} a_{if}\, b_{jf}\, c_{kf} + e_{ijk}, \text{ with } i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, K;\ f = 1, \ldots, F, \quad (5)$$

where x_{ijk} is the fluorescence intensity of the i^{th} sample at the j^{th} emission mode and at the k^{th} excitation mode; a_{if} is named "score" and is directly proportional to the concentration of the f^{th} analyte of sample *i*; b_{jf} is a scaled estimate of the emission spectrum of the f^{th} analyte at emission wavelength j; c_{kf} is linearly proportional to the specific absorption coefficient of the f^{th} analyte at excitation wavelength *k*; and e_{ijk} is the residual noise that represents the variability not accounted for by the model (Murphy *et al.*, 2013).

The PARAFAC model was run and validated using drEEM toolbox in MATLAB® R2022b (MathWorks®, USA), following the recommendations of Murphy *et al.* (2013) and Stedmon and Bro (2008). A final step of the dataset preprocessing is needed to run PARAFAC. The addition signal treatment to smooth the EEM dataset is detailed in Supplementary Material A-S1.

The PARAFAC model was developed with a total dataset of 414 samples. 31 samples presented high residual values and were identified as outliers: 27 from La Villette basin, three from the Marne VGA and one from SO-N. The total rejection rate of 7.5% is similar for all study sites. After their removal, the final dataset contained 383 samples (n=158 La Villette, n=204 the Marne, n=8 SO-N, n=8 SO-S and n=5 WWTP-O). Since PARAFAC modelling was used to identify the characteristic components of the dataset, the outliers must be removed.

PARAFAC models with F=4 to 8 components (see Equation 5) were fitted to find the best number of components. The 7-component PARAFAC model was validated through a split-half analysis, a validation with multiple split-half tests to confirm if the model is identically produced in independent subsample groups of the final dataset.

The validation method was an alternating '$S_4C_6T_3$'. This means that four split groups (S) were created. Each sample was alternately assigned to one of the four splits. The splits were assembled in six different combinations (C) to produce three split-half comparison tests (T).

The obtained components were compared to the data in the OpenFluor database (http://openfluor.org) (Murphy *et al.*, 2014). Each component was related to one type of DOM (humic-like, protein-like) and associated with components from the open database for qualitative comparisons.

The fluorescence intensity at maximum loadings of emission (b_{if}) and excitation (c_{if}) modes (Supplementary Material A-S2) was used to quantify components in a sample, denoted C_{if}. For a given sample *i* and a component *f*, C_{if} is given by Equation 6.

$$C_{if} = \max_{jk}(a_{if}\, b_{jf}\, c_{kf}) \quad (6)$$

The relative percentage of each component *f* was calculated by dividing the correspondent Cf value by the sum of Cf of each sample *i* (Equation 7). Cf represents its correspondent PARAFAC component *f*. Hereafter, the component and its fluorescence intensity will both be denoted Cf.

$$\%C_{if} = \frac{Cif}{\sum_{f=1}^{F} Cif}, \text{ with f} = 1, \ldots, F \quad (7)$$

According to Murphy *et al.* (2013), PARAFAC modelling assumes that the fluorescence intensity increases approximately linearly with the concentrations of their corresponding component. Therefore, the relative percentage could be associated with the relative abundance of each component. However, this assumption must be carefully considered.

2.5 Data analysis

The fluorescence data was analysed in five main steps. First, the fluorescence indices of all samples were compared. Second, the PARAFAC components of all samples were compared and associated with fluorophores from Coble (1996). Third, fluorescence indices from the river Marne samples

collected at the same day were used to investigate the variation of DOM composition with land cover in river Marne. Fourth, fluorescence indices from La Villette basin samples were compared to assess the temporal variation of DOM composition, comparing samples from dry and wet weathers. Fifth, the relationship between FIB and protein-like PARAFAC components was obtained through a linear regression model.

Differences between the studied sample groups were determined by a Kruskal-Wallis test, a non-parametric one-way ANOVA test. This multiple comparison test among more than two groups determinates which group differs from the others by using average ranks of the medians of the groups (Helsel *et al.*, 2020). For La Villette samples, Kruskal-Wallis test was also performed to investigate the variability of fluorescence indices between dry (n=85) and wet (n=98) weather samples.

Principal components analysis (PCA) was applied to reduce the dimensionality of fluorescence data (PARAFAC components and indices) and visualize the correlated variables among water types and among sampling sites.

3 Results

This section will successively set forth the results concerning: (1) the sampling weather conditions in La Villette and the river Marne; (2) the fluorescence indices of all sampling sites; (3) the PARAFAC model results; (4) the DOM composition among water types; (5) the impact of land-use in the Marne FDOM; (6) the time variation of FDOM in La Villette; (7) the relationship between PARAFAC components and FIB.

3.1 Sampling weather conditions in La Villette and in the river Marne

In La Villette basin, samples were collected over 29 months (October 2020 to March 2023). From the totality of the samples (n=183), 54% (n=98) were collected in wet weather, according to the definition of wet weather given in paragraph 2.2, and 46% (n=85) in dry weather.

In the river Marne, a total of 34 sampling campaigns were conducted in four months (June to September 2022): 18 in VGA (4 in dry and 14 in wet weather) and 19 in SMV points (11 in dry and 5 in wet weather). The campaigns were predominantly conducted during dry weather (78% of the totality of samples).

The rainfall events associated to wet weather samples have different distribution of intensities in both watercourses, due to the difference in the length of the sampling periods. In La Villette, 24% of the wet weather samples were collected after rainfall between 8 and 10 mm in 72 h. The observed maximum intensity was of 48 mm in 72 h (Supplementary Figure S4-A). In the river Marne, 27% of the wet weather samples were collected after rainfall between 18 and 20 mm in 72 h. The observed maximum intensity was of 26 mm in 72 h (Supplementary Figure S4-B).

3.2 Fluorescence indices for DOM characterization

3.2.1 Comparison of HIX values

Huguet *et al.* (2009) associated DOM characteristics with a range of values calculated according to Zsolnay *et al.* (1999). To find the correspondent range values for Ohno's expression (2002), which provides values in the range [0;1], better adapted to comparison between different water bodies, HIX was calculated with both expressions for all samples (Equation 8). The curve linking both

expressions (Supplementary Figure S5) can be used for the conversion between HIX_{Ohno} and $HIX_{Zsolnay}$. This helps to compare the values found in different waterbodies from the literature.

The HIX ranges from Huguet *et al.* (2009) and their DOM characteristics were then associated with new HIX ranges from Ohno (2002). For comparative purposes, DOM characteristics for HIX and BIX ranges defined by Huguet *et al.* (2009) are grouped in Table 1 (shown in section 4.1.1).

$$HIX_{Ohno} = \frac{HIX_{Zsolnay}}{HIX_{Zsolnay} + 1} \quad (8)$$

3.2.2 DOM characteristics and origin

The characteristics and origin of DOM can be estimated with the indices HIX and BIX (Figure 2, Supplementary Table S4).

In La Villette (n=183), HIX covers a range of [0.55-0.94], with an average of 0.86±0.08. BIX covers a range of [0.64-2.19], with an average of 0.90±0.26. In the river Marne (n=207), HIX covers a range of [0.71-0.90], with an average of 0.85±0.03. BIX covers a range of [0.73-1.62], with average of 0.87±0.09. All the values were very similar along the stretch. They showed less variation than in La Villette.

In WWTP-O effluent (n=5), a low variation of indices was observed. HIX was close to 0.80 and BIX was higher or equal to 1. This indicates a predominant biological origin of DOM. In SO-N (n=9) and SO-S (n=8), the ranges of HIX and BIX values were similar. Higher variations of HIX and BIX than in La Villette and the Marne were observed. HIX lower values in dry weather suggested lower humic character. BIX was lower than 1.0 at both sampling points.

According to Huguet *et al*. (2009), the combination of HIX and BIX values can be used to synthetise the DOM characteristics of the studied sites (mean with one-standard-deviation confidence ellipses are set forth in Figure 3). The river Marne samples showed weak humic character and an important biological activity. La Villette samples showed weak to strong humic character and low to high biological activity. WWTP-O effluents showed very low humic character and high biological origin. The HIX and BIX values of both SO samples are close. The humic character and the biological origin were relatively low. The samples of La Villette, the Marne and WWTP-O matched the classification. Differently, the SO samples were located outside of the expected categories, suggesting that the classification could be extended to different urban waters categories.

3.3 PARAFAC model results

3.3.1 PARAFAC function

The first result was to define a function which could be applied, by using the identified PARAFAC components, to obtain the corresponding PARAFAC values for future samples.

The creation of a robust PARAFAC model requires a large number of EEMs with great diversity of DOM sources of diverse chemical quality. However, once the DOM fluorescence signal is decomposed, new EEMs from the same or similar study site, should not interfere in the obtained

PARAFAC model. Therefore, these new EEMs should fit well in the existing model (Fellman *et al.*, 2009; Murphy *et al.*, 2011).

The defined function provides the new PARAFAC scores, and therefore Cf, on future samples from the same sampling sites. With the emission (b_{jf}) and excitation (c_{kf}) modes obtained from the original PARAFAC model, it was possible to obtain the score a_{if} and compute Cf values of new EEMs (Fellman *et al.*, 2009). To this end, the values of a_{if} to obtain the minimum squared error ε_i in the sample *i* are computed by Equation 9.

$$\varepsilon_i = \sum_{j=1}^{J}\sum_{k=1}^{K}\left(x_{ijk} - \sum_{f=1}^{F} a_{if}\, b_{jf}\, c_{kf}\right)^2 \quad (9)$$

3.3.2 PARAFAC components

A model with 7 components was validated by PARAFAC analysis (C1-C7, Supplementary Figure S6). Each component was associated with similar components from OpenFluor Database (Supplementary Table S5). The similarity score (Tucker congruence coefficient) of all the components was >0.95 for emission and excitation spectra loading, with global score >0.90.

Components C1 to C5 were identified as humic-like components. C1, C2, C3 and C5 were identified as terrestrial humic-like components. The maximum loadings of C1, C2 and C5 were associated to peak A. C2 displayed a greater emission wavelength (red-shifted). The maximum loading of C3 was similar to peak C.

C4 was associated to humic-like substances produced *in-situ* by phytoplankton and microbial activity. The maximum loading of C4 was in lower emission wavelengths (blue-shifted) than peak C and in greater excitation wavelengths than peak M.

Components C6 and C7 were identified as protein-like, close to tryptophan-like fluorescence (peak T). The maximum loading of component C7 has greater excitation wavelength than peak T.

The components C1-C7 were represented by Cf, the fluorescence intensity at the maximum loading of excitation and emission modes for each component f. The components were grouped in terrestrial humic-like (C1, C2, C3 and C5), “low” humic-like (C4), with lower molecular weight, and protein-like (C6 and C7).

The fluorescence intensities (Supplementary Figure S7) and the relative percentage of PARAFAC components from all the sampling points in the five water bodies was estimated. The mean relative percentage of each component *f* (%Cf) was calculated for the five water bodies (Figure 4, Supplementary Table S6). The relative percentage of each component contributes to the signatures of SO and WWTP-O and of the watercourses La Villette and the Marne.

In La Villette, terrestrial humic-like composition was predominant on average (66.0%). The predominant terrestrial humic-like component and with higher variability was C1 (22.4%±6.6%), followed by C5 (15.0%±6.5%). Protein-like components represent 18.5%. The predominant protein-

like component was C7 (11.7%±9.6%). It is more abundant and more variable than C6 (6.8%±4.0%). The highest variations of C1 and C7 occurred mostly during wet weather (Supplementary Figure S8).

In the river Marne, terrestrial humic-like composition was predominant (65.1%). The predominant humic-like component was C5 (20.8%±1.6%). C1 has the highest variability (17.8%±2.7%). The protein-like components represent 20.0%. The predominant protein-like component was C6 (11.8%±2.3%). The higher variability was associated to C7 (8.2%±4.4%) (Supplementary Figure S8).

In WWTP-O (Supplementary Figure S10), terrestrial humic-like components were the most abundant (52.3%). Protein-like components represented 25.6%. The predominant humic-like components were C5 (23.2%±4.2%), terrestrial humic-like, and C4 (21.9%±2.0%), "low" humic-like. The predominant protein-like component was C6 (16.5%±2.1%). The variability of the components was very low, as evidenced by their standard deviations (ranging from 0.5 to 4.2%) and narrow %Cf ranges ([0.0-26.5%]).

In both SO-N and SO-S (Supplementary Figure S10), terrestrial humic-like components were the most abundant (62.5% and 60.1%, respectively). Protein-like components represented 25.7% and 25.8%, respectively. The predominant terrestrial humic-like component in SO-N was C1 (24.9%±17.8%), and in SO-S, C3 (22.5%±1.6%). The dominant protein-like component was C6 in both SO-N (23.4%±11.0%) and SO-S (23.6%±8.0%). These values are the highest measured in all samples. The components with the highest variability were C1 and C6. Higher values of %C1 and lower %C6 were obtained in wet weather, suggesting that they could be potentially used as signature of local precipitation.

In La Villette and the Marne, the average of relative percentage of all terrestrial humic-like components was around 66%, of "low" humic-like component was 15%, and of protein-like components, 19%. In SO, terrestrial humic-like components corresponded to 61%, "low" humic-like components to 13%, and protein-like components to 26%. In WWTP-O, terrestrial humic-like components represented 52% of components abundance, "low" humic-like component, 22%, and protein-like component, 26% (Figure 5). In SO and WWTP-O, protein-like components are more abundant than in La Villette and the river Marne. In WWTP-O "low" humic-like component was more abundant than in SO.

3.4 DOM composition among water types

The distribution of the relative abundance of PARAFAC components among the different water types was compared through Kruskal-Wallis test. The average ranks of the median of each component for each water type was computed. The statistical similarity of each component among the water types is presented in Supplementary Figure S11.

Overall, the contribution of all PARAFAC components were very similar in SO and WWTP-O. The main difference between these groups were the low humic-like component C4 and the protein-like component C7. Both %C4 and %C7 were higher in WWTP-O and significantly different from the SO. In WWTP-O, %C4 was significantly higher from the other water types and %C7 was similar to the watercourses.

The highest %C6 values were measured in SO and WWTP-O. In contrast, %C6 in the river Marne was greater than in La Villette. This may suggest that the outlets which discharge in the river Marne could be the main source of C6 in the river. Higher variability in the composition of protein-like components could contribute to the detection of change in microbial DOM. In the watercourses, the

protein-like component with higher standard deviation and wider range was C7. In SO and WWTP-O, it was C6 (Table S6). This suggests that %C7 could be an indicator of microbial DOM change in La Villette and the river Marne.

The median HIX of La Villette (0.89) was greater than in the Marne (0.86). The median BIX of La Villette (0.80) was lower than in the Marne (0.86). La Villette and the Marne HIX medians were significantly different (Supplementary Figure S11), indicating a potentially higher humification level in La Villette. However, HIX and BIX median values between both watercourses were close enough to be part of the same category, indicating a DOM composition with an important humic character and an intermediate biological activity. Among the samples from SO and WWTP-O, HIX was very similar. WWTP-O samples presented the highest mean BIX values, being significantly similar from the river Marne samples.

The differences between the variances in La Villette and the Marne was likely due to the different characteristics of the sampling periods of each site. In La Villette, samples were collected for more than two years (October 2020 to March 2023), covering all seasons, and including 38 rainfall events of intensities from [2.6-48.2 mm] in 72 h. In the Marne, samples were collected only during Summer 2022, in scheduled dates regardless the weather conditions.

A PCA analysis was performed to analyse the patterns of PARAFAC components and fluorescence indices HIX and BIX among the different water types. The two first axis explained 76.4% of the total variance (Figure 6). In the plan defined by axis PC1 and PC2, the watercourses (La Villette and the Marne river) were well separated from SO and WWTP-O.

The first axis, PC1, explained 49.2% of total variance. PC1 was defined positively by C4, C5, and C6. SO and WWTP-O are distributed along PC1, close to C1 to C6. The second axis, PC2, explained 27.2% of total variation. It was defined positively by C7 and BIX, and negatively by HIX. La Villette data spread along PC2, mainly driven by C7 and BIX. The dispersion of La Villette samples was higher than the Marne samples. This could be due to the longer studied period in La Villette, covering more than two years (29 months) instead of one summer in the Marne.

3.5 Impact of land-use in the river Marne

In the river Marne, the influence of land-use on fluorescence components and their variation along the river stretch, upstream to downstream, were investigated. Samples collected at the same day were compared. A total of 15 sampling campaigns were conducted in the main 6 SMV points (SMV1, SMV6B, SMV7C, SMV8B, SMV10 and SMV14B).

Two WWTP outlets are located in the studied stretch. The most upstream, WWTP-O, is located 7.8 km upstream SMV1. The second, WWTP-PMA, is 1.8 km upstream SMV10. The distance between the sampling points and their nearest WWTP outlet at upstream was considered (DW): the distance between WWTP-O and SMV1, SMV6B, SMV7C, SMV8B; and the distance between WWTP-PMA and SMV10, SMV14B (Supplementary Table S7).

The mean values of the PARAFAC component proportion and the fluorescence indices were very close among all 6 SMV points, from upstream to downstream (Figure 7-A). Among the terrestrial humic-like components, C1 and C5 have the highest mean values, respectively of 18% and 21%.

Regarding the statistical similarity of the protein-like components, some differences appeared in C6 (Figure 7-B). At SMV6B, %C6 presented the highest median (11.7%) and the highest 3rd quartile

(13.4%). Between SMV1 and SMV6B, the 12km-stretch contains 26 stormwater network outlets, including SO-N. The high %C6 level at SMV6B could be due to large outlet loading. %C6 was also slightly higher at SMV10 than SMV8B. Between SMV8B and SMV10 is located the outlet of WWTP-PMA, suggesting an impact of the WWTP on the protein-like DOM composition.

A Principal Component Analysis (PCA) was computed including the seven PARAFAC components, the distance from the closest upstream WWTP outlet (DW) and the land cover in the 1km² upstream each sampling point, grouped in "urban green area", "natural green area", "impervious area" and "water area" (Figure 8). The first two axis, PC1 and PC2, explained 53.8% of total variance of the dataset. PC1 explained 34.5% of the total variance. It was predominantly defined by the positive loadings of the C1 to C6, and the negative loadings of DW. PC2 explained 19.3% of total variance. It was positively defined by BIX, C7 and the impervious area land-use, and negatively defined by urban green area and HIX.

The values were mainly distributed along PC1. The components C1 to C6 were closely distributed and orthogonally to C7. C1 to C6 were inversely related to the distance from a WWTP outlet upstream, suggesting that the lower the distance, the higher the loadings of these components.

A slight influence of the land use was observed. Impervious areas appeared to be more related to BIX and C7; they may contribute to higher microbial loading. Urban green areas appeared to be more related to HIX; they may have higher DOM humification level.

3.6 Rainfall impact on DOM composition in La Villette

The impact of rainfall on fluorescent DOM in La Villette was explored. Samples were collected during dry (85 samples) and wet weather (98 samples), over 2.5 years (October 2020 to March 2023). A Kruskal-Wallis test was performed for assessing the differences between the samples.

The relative contribution of the PARAFAC components, HIX and BIX were compared to identify a potential signature of local contamination due to stormwater network or WWTP outlets or runoff during rainfall events. No significant differences were observed between dry and wet weather for %C3, %C5, %C7 and BIX (Figure 9).

Significant differences were identified in %C1, %C2, %C4, %C6 and HIX. In wet weather, the humic-like components %C1, %C2 and HIX were lower than in dry weather. The values of %C4, a component related to microbial activity (Galletti *et al.*, 2019), and %C6, a protein-like component, were significantly higher. The increase of the components C6 and C4 suggests an increase of microbial loadings during rainy events.

3.7 Relationship of FIB and PARAFAC components

The enumeration of FIB, *E. coli* and IE in MPN.100mL^{-1}, was performed in samples from all water types. The sampling points were La Villette point C (n=10), the Marne VGA (n=18), WWTP-O (n=5), SO-N (n=9) and SO-S (n=8). The same sampling protocol was applied (sub-section 2.2).

The $\log_{10}$-transformed values (in MPN.100mL^{-1}) are presented in Figure 10. The order of magnitude of the FIB varied among the sampling sites. For *E. coli*, in La Villette, it varied by a factor between 2 and 3, in the Marne VGA, between 2 and 4, in WWTP-O, between 4 and 5, in SO-N, between 5 and 6, and in SO-S, between 5 and 7. The order of magnitude of IE, in La Villette varied by a factor

between 0 and 2, in the Marne VGA, between 0 and 3, in WWTP-O, between 3 and 4, in SO-N, between 4 and 5, in SO-S, between 4 and 6.

Interestingly, in the stormwater network outlets SO-N and SO-S, during dry weather, higher *E. coli* values were observed in comparison to wet weather. This unexpected result suggests a constant contamination on the network by sewage infiltration or misconnection of the sewage network. During rain events, the microbiological contaminants are probably diluted by drainage water.

Among all samples, *E. coli* and IE had a strong relationship (R^2=0.88, p-value<0.001) (Supplementary Figure S12). This result reinforced the use of FIB as indicators of faecal contamination.

The correlation between log10-transformed FIB and PARAFAC components was explored using all FIB data (n=50). A good relationship with the protein-like components, C6 or C7, was expected. A significant relationship with $\log_{10}$(C6) with both FIB, *E. coli* (R^2=0.72, Figure 11-A, Equation 10) and IE (R^2=0.61, Figure 11-B), was obtained (p<0.001).

$$log_{10}(E.coli) = 5.46 + 2.81\ log_{10}(C6) \quad (10)$$

Where *E. coli* is in MPN.100mL^{-1} and C6 in RU. According to this relationship, considering the *E. coli* threshold for sufficient bathing water quality (900 MPN.100mL^{-1}), an indicative warning level could be based on C6<0.13 RU.

The relationships between C7 and both FIB were not significant (p>0.5). Therefore, they were not considered.

In 2023, samples were collected in the river Marne at an additional point, point IMP, close to SMV8B. The samples (n=16) were collected from 25/05-03/07/2023, following the same sampling protocol as in point VGA (24-hour integrated samples). C6 values were obtained from the PARAFAC function (Equation 9 from 3.3.1). The relationship between *E. coli* and C6 was applied for *E. coli* estimation (Figure 12-A).

The median of *E. coli* in the IMP samples (2023) was 1500 MPN.100mL^{-1}. The estimated *E. coli* median from Equation 10 was 1378 MPN.100mL^{-1}, very close to the observed value. The distribution of observed and estimated *E. coli* was very similar (Figure 12-B). This result contributes to the validation of Equation 10.

The observed *E. coli* values can be used to determine the overpassing of the bathing threshold (900 MPN.100mL^{-1}). From 25/05 to 03/07/2023, 11 out 16 occurrences (68%) matched the observations, either higher or lower than the threshold. Of the five incorrect estimations, three overestimations and two underestimations were found.

4 Discussion

Hereafter, the results concerning the fluorescence indices and the PARAFAC components in the studied sites are discussed and compared to other results in similar waterbodies. The land cover in the river Marne and the variation of fluorescence components in La Villette are also discussed, as well as the relationship between *E. coli* and protein-like components.

4.1 FDOM patterns in different urban water types

4.1.1 Fluorescence indices HIX, BIX and FI

HIX computed according to Zsolnay *et al.* (1999) or to Ohno (2002) has different scales. The proposed conversion between these scales allows us to compare the values in a unified way (Table 1). BIX is associated to the biological activity from first stages (around 2 days) of organic matter decomposition. BIX and HIX values are related to the DOM characteristics: they are complementary and expected to be inversely proportional to each other. Samples with high bacterial activity should have high BIX and present low humic character, associated to low HIX, and vice-versa (Huguet *et al.*, 2009).

The samples of La Villette and the river Marne matched well the typology proposed by Huguet *et al.* (2009). However, the stormwater network outlets (SO-N and SO-S) do not fit within this typology (see Figure 3). Therefore, such type of urban water probably holds a different FDOM signature, which could be used to extend the classification of DOM characteristics.

Previous HIX and BIX data from the Seine and the Marne were obtained in September 2012 by Nguyen (2014). The mean HIX values from La Villette (0.86) and the Marne (0.85) are similar to 2012 values found in the river Marne (0.88) in the same stretch as ours. In the Seine, downstream Paris, close values (0.87) from La Villette were found. The mean BIX values from La Villette (0.90) and the Marne (0.87) were greater than in the Seine downstream Paris (0.75) and in the Marne (0.67) in September 2012.

HIX values in La Villette and the Marne were compared to 9 rivers and 9 streams, in Berlin area, including two heavily polluted rivers due to WWTP effluents (Romero González-Quijano *et al.*, 2022). HIX in La Villette and in the Marne were similar to the Berlin streams (0.86) and rivers (0.85).

FI was introduced by McKnight *et al.* (2001) for the source assessment of humic substances. Low FI values (~1.2) indicate terrestrial derived DOM and high FI (~1.8), microbial derived DOM (Fellman *et al.*, 2010). In this study, FI was higher than 1.8 in all samples. In La Villette and in the Marne, FI was close, 1.97 and 2.13, respectively. In the SO and WWTP effluent, FI was higher, around 2.4 (Table S4). In Berlin rivers, FI was 1.66, but in the two highly polluted rivers, FI was higher (~1.8) (Romero González-Quijano *et al.*, 2022). In an anthropized watershed in Spain (Marcé *et al.*, 2021), FI was 1.4 in the Ter River but it was higher, 1.6, in the impaired streams and WWTP effluent. Based in high FI values, a dominant microbial contribution on DOM composition cannot be concluded. However, high FI values may indicate a highly urbanized watershed with limited terrestrial DOM components.

4.1.2 PARAFAC components

As the position of the characteristic fluorescence peaks, such as peak T, can be slightly shifted according to the water matrix, PARAFAC model achieves a more accurate location of the fluorescence peaks.

In the obtained 7-component PARAFAC model, the protein-like components were C6, close to peak T, and C7. C6 was similar to the main protein-like fluorophore in WWTP effluents (Cohen *et al.*, 2014). C6 displayed a high correlation with biogeochemical parameters, such as organic carbon, phosphates and faecal bacteria (Tedetti *et al.*, 2012; Cohen *et al.*, 2014).

The comparison of the PARAFAC components among different studies is not straightforward. Similar water types may be represented by different PARAFAC models with different characteristic components (Meng *et al.*, 2013; Lambert *et al.*, 2017; Marcé *et al.*, 2021; Romero González-Quijano *et al.*, 2022). To overcome this limitation, we considered two fractions: protein-like components and humic-like components.

In the Berlin region, Romero González-Quijano *et al.* (2022) obtained a 7-component PARAFAC model for a dataset including rivers, streams, ponds and lakes. They obtained two protein-like components. In the rivers, the average of protein-like components was 16.0%, slightly lower than La Villette (18.5%) and the Marne (20.0%). In the two highly polluted rivers, the protein-like components were 20.5%, similar to the values found in WWTP-O (25.6%). A percentage of protein-like components higher than 20% might indicate a microbiological pollution such as WWTP effluents.

Among our study sites, the tryptophan-like component (C6) was high in WWTP-O effluent. Based on the OpenFluor database, C6 was found very similar to a component found by Cohen *et al.* (2014) in four WWTPs (Israel). C6 was also a dominant component in other WWTP studies (Li *et al.*, 2020; Sciscenko *et al.*, 2022). C6 might be considered as a signature of a microbial activity in a WWTP effluent.

A high percentage of C6 was observed in the stormwater outlets in dry weather (Figure 4). It was higher than in wet weather, contrary to what is expected in a stormwater network. A probable wastewater contamination may occur in the SO during dry weather. This was corroborated by higher *E. coli* (Figure 10-B) in dry than in wet weather samples.

In lakes, tryptophan-like component can also be considered as a potential indicator of eutrophication. In Chinese lakes, a component close to peak T and to our C6 was found (Ren *et al.*, 2021). It was associated to pollutant-degrading activity of microalgae, including cyanobacteria. Therefore, in future studies, it will be interesting to explore the sensitivity of C6 to eutrophication indices.

4.2 Land cover around the river Marne study sites

In the river Marne, according to the PCA results, impervious areas may be associated to BIX and the protein-like component C7. Urban green areas may be associated to HIX, suggesting higher humification level of the FDOM. These results are consistent with previous studies. Urbanized areas have a high microbial organic matter composition, resulting in higher BIX values, and less plant-soil-derived character, with lower HIX (Lambert *et al.*, 2017). The relationship between microbial activity (BIX and C7) and impervious areas may origin from surface runoff, composed by different sources of pollution, and from WWTP effluents (Romero González-Quijano *et al.*, 2022).

4.3 Rainfall impact in fluorescence components in La Villette basin

In general, humic-like fluorescence is strongly controlled by hydrological processes. Protein-like fluorescence is tightly linked with biological processes, depending of water temperature, sunlight intensity, water turbidity, etc. (Parlanti *et al.*, 2000; Fellman *et al.*, 2010; Meng *et al.*, 2013). Higher humic character was found during high-water than in low-water periods in the catchments of the Seine (France) (Nguyen, 2014) and in a river in the USA (Jaffé *et al.*, 2008).

In urban catchments, microbiological contamination peaks occur following rainfall episode (Burnet *et al.*, 2019). Therefore, rainfall may have a significant impact in the protein-like components. In La

Villette, the protein-like components C6 and C7 in wet weather samples increased more than the humic-like components. Therefore, the relative contribution of protein-like components was higher. The wet weather samples presented lower humic character (lower HIX and lower terrestrial humic-like components) and higher microbial loading (higher protein-like components) than the dry weather samples. These results were consistent with the values obtained in an urban river in China (Yuan *et al.*, 2023).

It must be highlighted that La Villette is a highly controlled hydraulic system, with many sluices upstream and one sluice downstream. The variation of discharge is very limited, even during heavy rainfall episodes, reducing their impact on humic-like components. Protein-like components from direct runoff and stormwater networks are dominant. The protein-like components can be used as a fingerprint of rainy events. An improved assessment of rain impact on FDOM in La Villette, would require sampling before, during and after rainy events (for example, every 5 h during four days).

4.4 *E. coli* and protein-like component relationship

The link between *E. coli* and tryptophan-like component has been explored for several years (Cumberland *et al.*, 2012; Baker *et al.*, 2015; Frank *et al.*, 2018; Nowicki *et al.*, 2019; Bedell *et al.*, 2020). The obtained relationships were very variable according to the study sites (Table 2). The water matrix composition changes with the catchment hydrogeological characteristics.

In six rivers in South Africa, Baker *et al.* (2015) obtained a relationship between *E. coli* and peak T (n=136, r=0.75). In two urban rivers, one major river and one sewage treatment effluent in UK, Cumberland *et al.* (2012) obtained a relationship between *E. coli* and peak T (n=46, r=0.87).

A relationship (Equation 10) between C6 and *E. coli* was obtained with our study site samples (n=50, r=0.85). More data of C6 and FIB covering a wider range of FIB values, will allow to have a statistically stronger relationship. Measurement of turbidity and dissolved organic carbon may also provide additional information to improve the relationship between FIB and FDOM.

This relationship was applied to additional samples from the Marne, collected in 2023. The distribution of estimated and observed *E. coli* were close. This result opens the way to the application of Equation 10 to determine, from FDOM measurements, the overpassing of the *E. coli* bathing threshold (900 MPN.100mL^{-1}).

5 Conclusion

The origin, terrestrial or biological, of the main components of the studied urban waters was identified through the fluorescence of dissolved organic matter (FDOM). The continuum of the water types, from WWTP effluent, stormwater outlet, urban river, to a very urbanized basin, cover a wide range of FDOM fingerprints, extending the classification of surface water FDOM.

The validated 7-component PARAFAC model was effective in characterizing the different urban water types. Components C6 and C7, as well as BIX, were identified as fingerprints of microbial contamination. These protein-like components may serve as valuable markers for the assessment of microbial contamination in urban water samples.

In La Villette, the impact of rainfall on water microbiological quality was observed. The FDOM values demonstrated a significant increase in microbial activity following rainy periods. FDOM analysis can provide a sensitive technique for detecting changes in water quality after rainfall events.

In SO and WWTP-O effluent, component C6 exhibited high values. In agreement with previous studies, protein-like fluorescence components, such as C6, could be used as effective indicators for monitoring microbiological water quality. An increase of C6 level would suggest a microbial contamination and deserves further attention for managing water resources.

Urban waterways are highly dynamic systems, with fluctuating microbiological quality depending on the hydrometeorology and sewer accidents. For different uses of water resources, such as bathing, the implementation of an active management has been recommended by the World Health Organization (2021) and the Bathing Water Directive (BWD, 2006/7/EC). A real time monitoring of FDOM could be coupled with a numerical modelling in such a warning system aimed to support the management decision-making.

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

All authors contributed to the article and approved the submitted version.

8 Funding

This work was supported by Ecole des Ponts ParisTech. It is part of the OPUR project and of the Eurostars project (E115737-FORBATH). The first author was granted a doctoral fellowship from Ecole des Ponts ParisTech.

9 Acknowledgments

The authors are grateful to Ville de Paris (Service des Canaux and Service Technique de l'Eau et de l'Assainissement), Eau de Paris, Syndicat Marne Vive, Syndicat Intercommunal d'Assainissement de Marne-la-Vallée (SIAM), CD93 (Direction de l'Eau et de l'Assainissement) and Marin d'Eau Douce for the support and access to the sampling points. We thank CD94 laboratory (Direction des Services de l'Environnement et de l'Assainissement) and Prolog Ingénierie for the support in the field work in the river Marne. We thank the interns for the sampling and laboratory analysis: Vincent Thiercelin, Ana Carolina Pires, Clélia Rancinangue, Romero Lacerda and Anna Luong-Dihn-Giap. We also thank the interns Amanda Tolentino Mendes and Lorena Araújo for the fluorescence data treatment and analysis.

10 References

Baker, A., Cumberland, S. A., Bradley, C., Buckley, C., and Bridgeman, J. (2015). To what extent can portable fluorescence spectroscopy be used in the real-time assessment of microbial water quality? *Science of The Total Environment* 532, 14–19. doi: 10.1016/j.scitotenv.2015.05.114

Bedell, E., Harmon, O., Fankhauser, K., Shivers, Z., and Thomas, E. (2022). A continuous, in-situ, near-time fluorescence sensor coupled with a machine learning model for detection of fecal

720 contamination risk in drinking water: Design, characterization and field validation. *Water*
721 *Research* 220, 118644. doi: 10.1016/j.watres.2022.118644

722 Bedell, E., Sharpe, T., Purvis, T., Brown, J., and Thomas, E. (2020). Demonstration of Tryptophan-
723 Like Fluorescence Sensor Concepts for Fecal Exposure Detection in Drinking Water in
724 Remote and Resource Constrained Settings. *Sustainability* 12, 3768. doi:
725 10.3390/su12093768

726 Boavida, M.-J., and Wetzel, R. G. (1998). Inhibition of phosphatase activity by dissolved humic
727 substances and hydrolytic reactivation by natural ultraviolet light. *Freshwater Biology* 40,
728 285–293. doi: 10.1046/j.1365-2427.1998.00349.x

729 Burnet, J.-B., Sylvestre, É., Jalbert, J., Imbeault, S., Servais, P., Prévost, M., et al. (2019). Tracking
730 the contribution of multiple raw and treated wastewater discharges at an urban drinking water
731 supply using near real-time monitoring of β-d-glucuronidase activity. *Water Research* 164,
732 114869. doi: 10.1016/j.watres.2019.114869

733 Coble, P. G. (1996). Characterization of marine and terrestrial DOM in seawater using excitation-
734 emission matrix spectroscopy. *Marine Chemistry* 51, 325–346. doi: 10.1016/0304-
735 4203(95)00062-3

736 Cohen, E., Levy, G. J., and Borisover, M. (2014). Fluorescent components of organic matter in
737 wastewater: Efficacy and selectivity of the water treatment. *Water Research* 55, 323–334. doi:
738 10.1016/j.watres.2014.02.040

739 Cory, R. M., and McKnight, D. M. (2005). Fluorescence Spectroscopy Reveals Ubiquitous Presence
740 of Oxidized and Reduced Quinones in Dissolved Organic Matter. *Environ. Sci. Technol.* 39,
741 8142–8149. doi: 10.1021/es0506962

742 Cumberland, S., Bridgeman, J., Baker, A., Sterling, M., and Ward, D. (2012). Fluorescence
743 spectroscopy as a tool for determining microbial quality in potable water applications.
744 *Environmental Technology* 33, 687–693. doi: 10.1080/09593330.2011.588401

745 Edberg, S. C., Rice, E. W., Karlin, R. J., and Allen, M. J. (2000). Escherichia coli: the best biological
746 drinking water indicator for public health protection. *Journal of Applied Microbiology* 88,
747 106S-116S. doi: 10.1111/j.1365-2672.2000.tb05338.x

748 EU (2006). Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006
749 concerning the management of bathing water quality and repealing Directive 76/160/EEC.
750 Available at: https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:32006L0007
751 (Accessed November 16, 2020).

752 Fellman, J. B., Hood, E., and Spencer, R. G. M. (2010). Fluorescence spectroscopy opens new
753 windows into dissolved organic matter dynamics in freshwater ecosystems: A review.
754 *Limnology and Oceanography* 55, 2452–2462. doi: 10.4319/lo.2010.55.6.2452

755 Fellman, J. B., Miller, M. P., Cory, R. M., D'Amore, D. V., and White, D. (2009). Characterizing
756 Dissolved Organic Matter Using PARAFAC Modeling of Fluorescence Spectroscopy: A
757 Comparison of Two Models. *Environ. Sci. Technol.* 43, 6228–6234. doi: 10.1021/es900143g

Frank, S., Goeppert, N., and Goldscheider, N. (2018). Fluorescence-based multi-parameter approach to characterize dynamics of organic carbon, faecal bacteria and particles at alpine karst springs. *Science of The Total Environment* 615, 1446–1459. doi: 10.1016/j.scitotenv.2017.09.095

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., and Gilroy, E. J. (2020). Statistical methods in water resources. Reston, VA: U.S. Geological Survey. doi: 10.3133/tm4A3

Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J. M., and Parlanti, E. (2009). Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Organic Geochemistry* 40, 706–719. doi: 10.1016/j.orggeochem.2009.03.002

HydroPortail, E. (2022). Station hydrométrique - F664 0001 04 : La Marne à Gournay-sur-Marne - Pont - Fiche de synthèse - Données hydrologiques de synthèse | SCHAPI - HydroPortail. *HydroPortail*. Available at: https://www.hydro.eaufrance.fr/stationhydro/F664000104/synthese (Accessed November 29, 2022).

Institut Paris Region (2022). Occupation du sol 2021 en 11 postes de légende d'Île-de-France. Available at: https://data-iau-idf.opendata.arcgis.com/datasets/iau-idf::occupation-du-sol-2021-en-11-postes-de-l%C3%A9gende-d%C3%AEle-de-france/about (Accessed March 14, 2023).

ISO (1998a). ISO 7899-1:1998 - Water quality — Detection and enumeration of intestinal enterococci — Part 1: Miniaturized method (Most Probable Number) for surface and waste water. Available at: https://www.iso.org/standard/14852.html (Accessed February 22, 2022).

ISO (1998b). ISO 9308-3:1998 - Water quality — Detection and enumeration of Escherichia coli and coliform bacteria — Part 3: Miniaturized method (Most Probable Number) for the detection and enumeration of E. coli in surface and waste water. *ISO*. Available at: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/02/08/20878.html (Accessed June 15, 2020).

Jaffé, R., McKnight, D., Maie, N., Cory, R., McDowell, W. H., and Campbell, J. L. (2008). Spatial and temporal variations in DOM composition in ecosystems: The importance of long-term monitoring of optical properties. *Journal of Geophysical Research: Biogeosciences* 113. doi: 10.1029/2008JG000683

Kim, J., Kim, Y., Kang, H.-W., Kim, S. H., Rho, T., and Kang, D.-J. (2020). Tracing water mass fractions in the deep western Indian Ocean using fluorescent dissolved organic matter. *Marine Chemistry* 218, 103720. doi: 10.1016/j.marchem.2019.103720

Lambert, T., Bouillon, S., Darchambeau, F., Morana, C., Roland, F. A. E., Descy, J.-P., et al. (2017). Effects of human land use on the terrestrial and aquatic sources of fluvial organic matter in a temperate river basin (The Meuse River, Belgium). *Biogeochemistry* 136, 191–211. doi: 10.1007/s10533-017-0387-9

Lawaetz, A. J., and Stedmon, C. A. (2009). Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water. *Appl Spectrosc* 63, 936–940. doi: 10.1366/000370209788964548

Li, J., Wang, L., Geng, J., Li, S., Yu, Q., Xu, K., et al. (2020). Distribution and removal of fluorescent dissolved organic matter in 15 municipal wastewater treatment plants in China. *Chemosphere* 251, 126375. doi: 10.1016/j.chemosphere.2020.126375

Marcé, R., Verdura, L., and Leung, N. (2021). Dissolved organic matter spectroscopy reveals a hot spot of organic matter changes at the river–reservoir boundary. *Aquat Sci* 83, 67. doi: 10.1007/s00027-021-00823-6

Markager, S., and Vincent, W. F. (2000). Spectral light attenuation and the absorption of UV and blue light in natural waters. *Limnology and Oceanography* 45, 642–650. doi: 10.4319/lo.2000.45.3.0642

McCabe, K. M., Smith, E. M., Lang, S. Q., Osburn, C. L., and Benitez-Nelson, C. R. (2021). Particulate and Dissolved Organic Matter in Stormwater Runoff Influences Oxygen Demand in Urbanized Headwater Catchments. *Environ. Sci. Technol.* 55, 952–961. doi: 10.1021/acs.est.0c04502

McKnight, D. M., Boyer, E. W., Westerhoff, P. K., Doran, P. T., Kulbe, T., and Andersen, D. T. (2001). Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography* 46, 38–48. doi: 10.4319/lo.2001.46.1.0038

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

Mladenov, N., Zheng, Y., Miller, M. P., Nemergut, D. R., Legg, T., Simone, B., et al. (2010). Dissolved Organic Matter Sources and Consequences for Iron and Arsenic Mobilization in Bangladesh Aquifers. *Environ. Sci. Technol.* 44, 123–128. doi: 10.1021/es901472g

Morris, D. P., Zagarese, H., Williamson, C. E., Balseiro, E. G., Hargreaves, B. R., Modenutti, B., et al. (1995). The attenuation of solar UV radiation in lakes and the role of dissolved organic carbon. *Limnology and Oceanography* 40, 1381–1391. doi: 10.4319/lo.1995.40.8.1381

Murphy, K. R., Hambly, A., Singh, S., Henderson, R. K., Baker, A., Stuetz, R., et al. (2011). Organic Matter Fluorescence in Municipal Water Recycling Schemes: Toward a Unified PARAFAC Model. *Environ. Sci. Technol.* 45, 2909–2916. doi: 10.1021/es103015e

Murphy, K. R., Stedmon, C. A., Graeber, D., and Bro, R. (2013). Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Anal. Methods* 5, 6557. doi: 10.1039/c3ay41160e

Murphy, K. R., Stedmon, C. A., Wenig, P., and Bro, R. (2014). OpenFluor– an online spectral library of auto-fluorescence by organic compounds in the environment. *Anal. Methods* 6, 658–661. doi: 10.1039/C3AY41935E

Nguyen, P. T. (2014). Study of the aquatic dissolved organic matter from the Seine River catchment (France) by optical spectroscopy combined to asymmetrical flow field-flow fractionation. Université de Bordeaux. Available at: https://theses.hal.science/tel-01170629 (Accessed March 21, 2023).

Nowicki, S., Lapworth, D. J., Ward, J. S. T., Thomson, P., and Charles, K. (2019). Tryptophan-like fluorescence as a measure of microbial contamination risk in groundwater. *Science of The Total Environment* 646, 782–791. doi: 10.1016/j.scitotenv.2018.07.274

Ohno, T. (2002). Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. *Environ. Sci. Technol.* 36, 742–746. doi: 10.1021/es0155276

Parker, C. A., and Barnes, W. J. (1957). Some experiments with spectrofluorimeters and filter fluorimeters. *Analyst* 82, 606–618. doi: 10.1039/AN9578200606

Parlanti, E., Wörz, K., Geoffroy, L., and 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. doi: 10.1016/S0146-6380(00)00124-8

Polubesova, T., Sherman-Nakache, M., and Chefetz, B. (2007). Binding of Pyrene to Hydrophobic Fractions of Dissolved Organic Matter: Effect of Polyvalent Metal Complexation. *Environ. Sci. Technol.* 41, 5389–5394. doi: 10.1021/es070722r

Portail assainissement collectif (2021). Stations de traitement des eaux usées. *Portail assainissement collectif*. Available at: https://www.assainissement.developpement-durable.gouv.fr/PortailAC/data (Accessed March 14, 2023).

Ren, W., Wu, X., Ge, X., Lin, G., Zhou, M., Long, Z., et al. (2021). Characteristics of dissolved organic matter in lakes with different eutrophic levels in southeastern Hubei Province, China. *J. Ocean. Limnol.* 39, 1256–1276. doi: 10.1007/s00343-020-0102-x

Romero González-Quijano, C., Herrero Ortega, S., Casper, P., Gessner, M., and Singer, G. (2022). Dissolved organic matter signatures in urban surface waters: spatio-temporal patterns and drivers. Biogeochemistry: Limnology. doi: 10.5194/bg-2021-340

Sciscenko, I., Arques, A., Micó, P., Mora, M., and García-Ballesteros, S. (2022). Emerging applications of EEM-PARAFAC for water treatment: a concise review. *Chemical Engineering Journal Advances* 10, 100286. doi: 10.1016/j.ceja.2022.100286

Sorensen, J. P. R., Carr, A. F., Nayebare, J., Diongue, D. M. L., Pouye, A., Roffo, R., et al. (2020). Tryptophan-like and humic-like fluorophores are extracellular in groundwater: implications as real-time faecal indicators. *Sci Rep* 10, 15379. doi: 10.1038/s41598-020-72258-2

Stedmon, C. A., and Bro, R. (2008). Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnology and Oceanography: Methods* 6, 572–579. doi: 10.4319/lom.2008.6.572

Tedetti, M., Longhitano, R., Garcia, N., Guigue, C., Ferretto, N., and Goutx, M. (2012). Fluorescence properties of dissolved organic matter in coastal Mediterranean waters influenced by a municipal sewage effluent (Bay of Marseilles, France). *Environ. Chem.* 9, 438. doi: 10.1071/EN12081

USEPA (1977). Clean Water Act of 1977. Available at: https://www.regulations.gov/document/EPA-HQ-OW-2021-0602-0076 (Accessed November 27, 2023).

USEPA (1986). Ambient Water Quality Criteria for Bacteria - 1986. Washington DC: Office of Water. Available at: https://nepis.epa.gov/ (Accessed November 13, 2023).

Wetzel, R. G. (1992). Gradient-dominated ecosystems: sources and regulatory functions of dissolved organic matter in freshwater ecosystems. *Hydrobiologia* 229, 181–198. doi: 10.1007/BF00007000

Williams, C. J., Yamashita, Y., Wilson, H. F., Jaffé, R., and 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. doi: 10.4319/lo.2010.55.3.1159

Wilson, H. F., and Xenopoulos, M. A. (2009). Effects of agricultural land use on the composition of fluvial dissolved organic matter. *Nature Geosci* 2, 37–41. doi: 10.1038/ngeo391

World Health Organization (2021). "Guidelines on recreational water quality: Volume 1 Coastal and fresh waters," in *Guidelines on recreational water quality: Volume 1 Coastal and fresh waters* Water, Sanitation, Hygiene and Health. (World Health Organization), 164. Available at: https://www.who.int/publications-detail-redirect/9789240031302 (Accessed November 30, 2023).

Xenopoulos, M. A., Barnes, R. T., Boodoo, K. S., Butman, D., Catalán, N., D'Amario, S. C., et al. (2021). How humans alter dissolved organic matter composition in freshwater: relevance for the Earth's biogeochemistry. *Biogeochemistry* 154, 323–348. doi: 10.1007/s10533-021-00753-3

Yuan, K., Wan, Q., Chai, B., Lei, X., Kang, A., Chen, J., et al. (2023). Characterizing the effects of stormwater runoff on dissolved organic matter in an urban river (Jiujiang, Jiangxi province, China) using spectral analysis. *Environ Sci Pollut Res* 30, 50649–50660. doi: 10.1007/s11356-023-25933-6

Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., and Saccomandi, F. (1999). Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere* 38, 45–50. doi: 10.1016/S0045-6535(98)00166-0

12 Data Availability Statement

The dataset of this study can be found in the OpenFluor Database: https://openfluor.lablicate.com/of/measurement/15067.

13 Figures and Tables

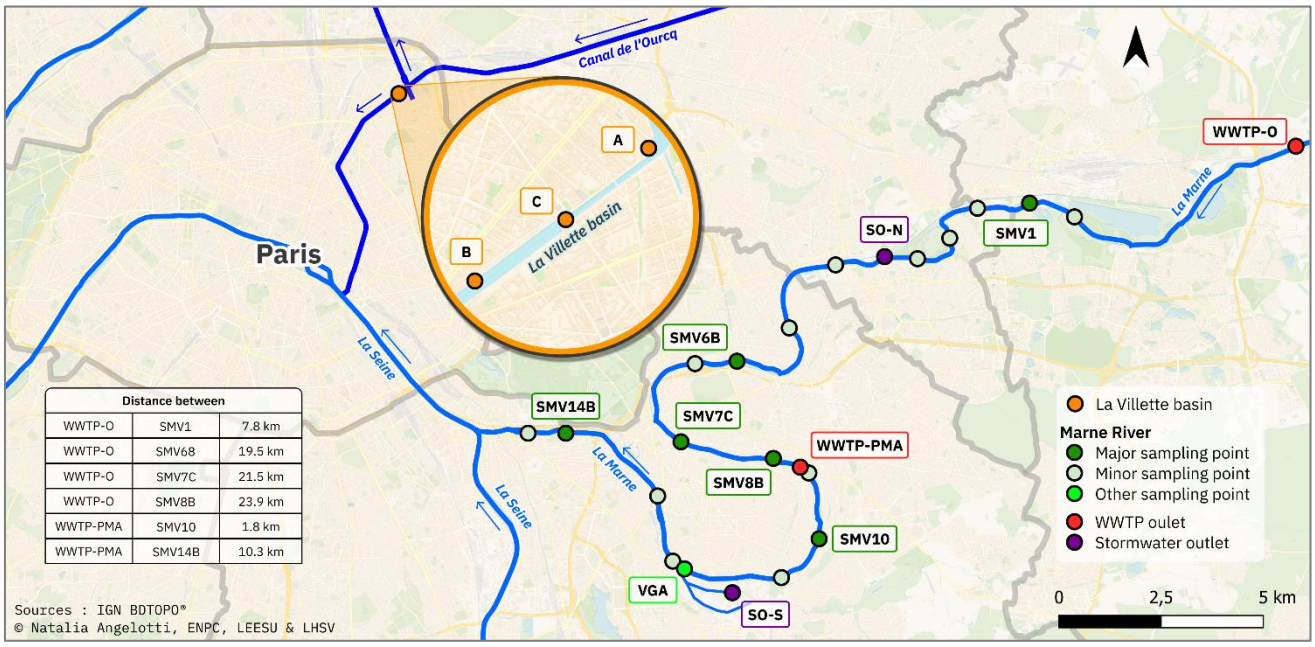


Figure 1. Sampling points. La Villette basin (orange), the river Marne (SMV in dark green; VGA in light green), stormwater outlets (SO-N and SO-S in purple), and WWTP outlet (WWTP-O in red).

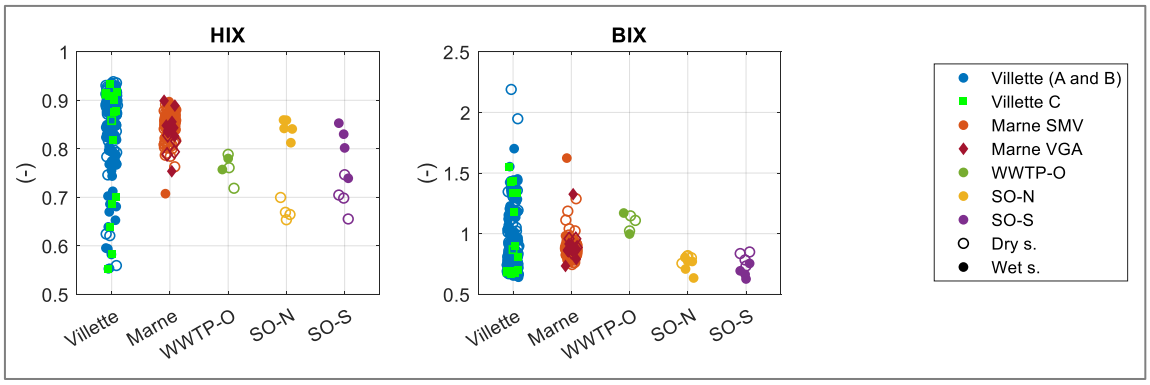
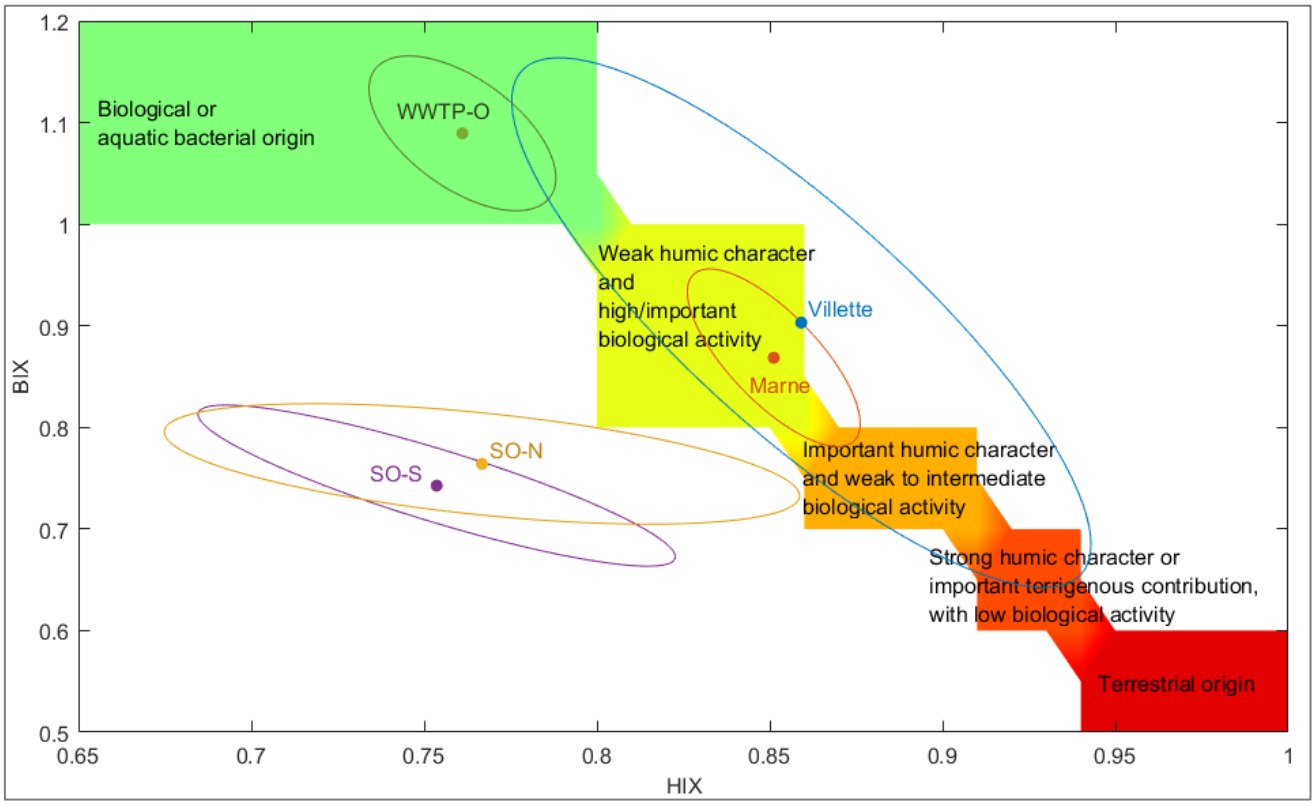
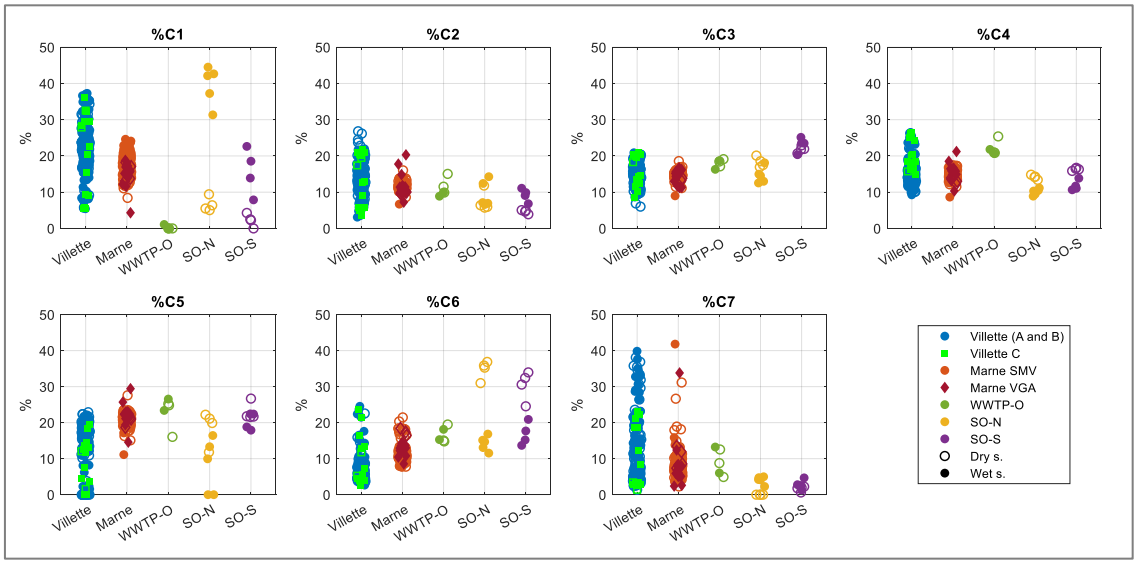


Figure 2. HIX and BIX values in La Villette basin at points A (o), B (o), and C (squares), the Marne SMV (o) and VGA (diamond), SO-N, SO-S and WWTP-O. Dry weather samples are open markers and wet weather samples are filled markers.



915

916 **Figure 3. DOM characteristics associated with HIX and BIX according to Huguet *et al.* (2009);**
917 **mean (filled circles) and one standard deviation (confidence ellipse) of the site samples.**



918

919 **Figure 4. Relative percentage of PARAFAC components (%) in La Villette basin points A**
920 **(blue), B (blue) and C (bright green), in the river Marne points SMV (orange) and VGA (dark**
921 **red), in the outlet of WWTO-O (dark green) and the outlets from stormwater networks SO-N.**

922

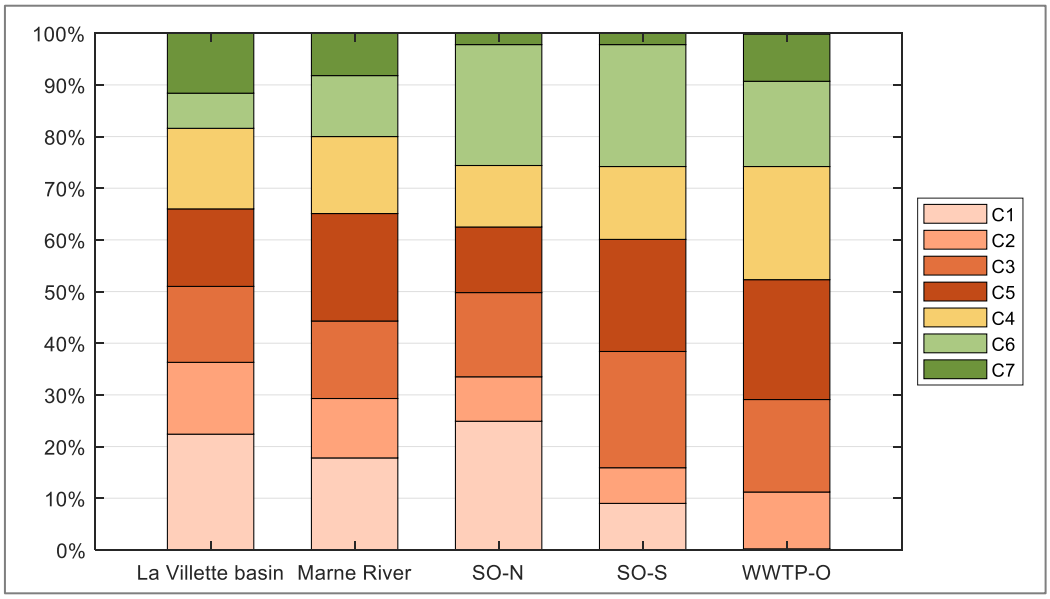


Figure 5. Relative abundance of the PARAFAC components. Components were gathered in three groups: terrestrial humic-like (orange), "low" humic-like (yellow) and protein-like (green).

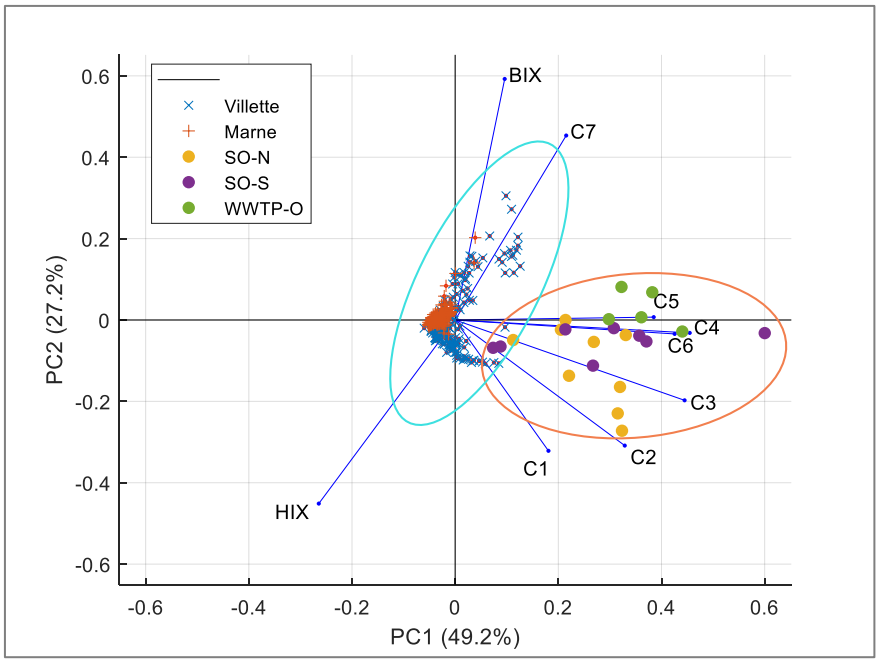
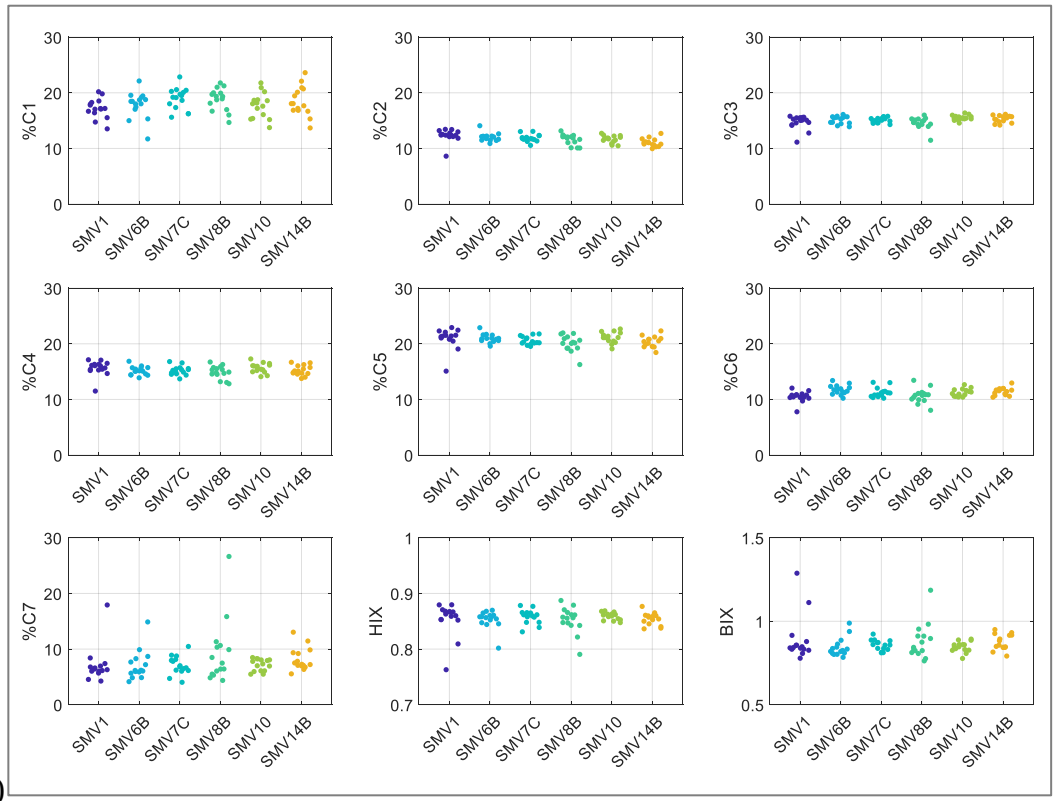
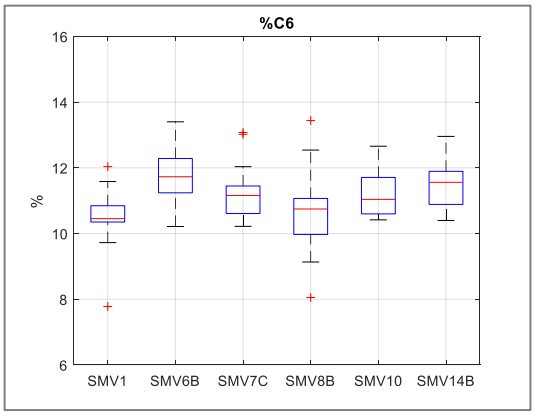


Figure 6. PCA of PARAFAC components, HIX and BIX, with dataset of La Villette basin (blue, n=183), Marne (orange, n=207), SO-N (yellow, n=9), SO-S (purple, n=8), and WWTP-O (green, n=5). Blue ellipse groups watercourses samples (La Villette basin and the river Marne) and orange ellipse the SO and WWTP-O.



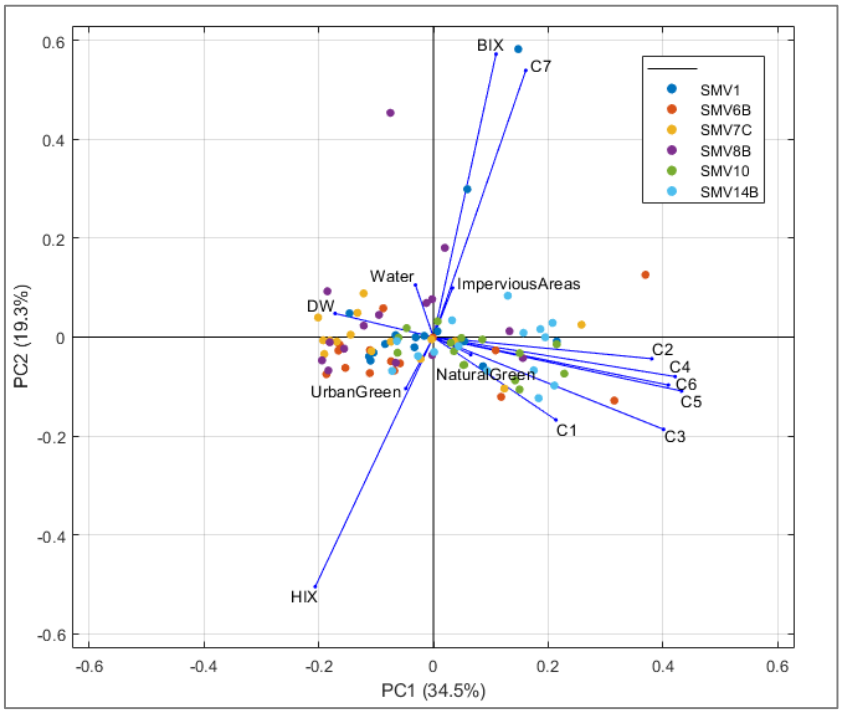
(A)



(B)

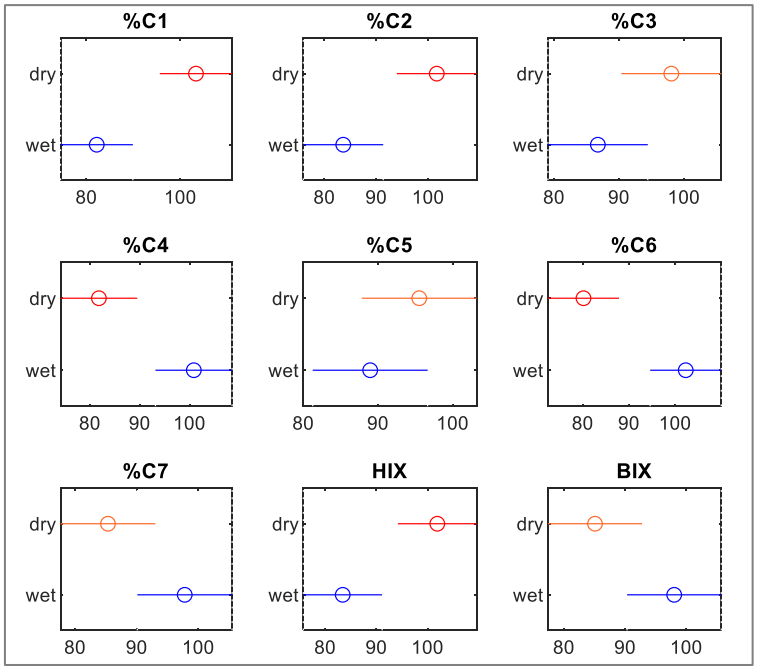
Figure 7. (A) Swarm chart of %C1 to %C7 and (B) Boxplot of %C6 of the main group of 6 SMV sampling points along the river Marne.

939



940 **Figure 8. PCA of the main group of 6 SMV river Marne points with C1-C7, land-use and**
941 **distance from the nearest WWTP outlet upstream (DW).**

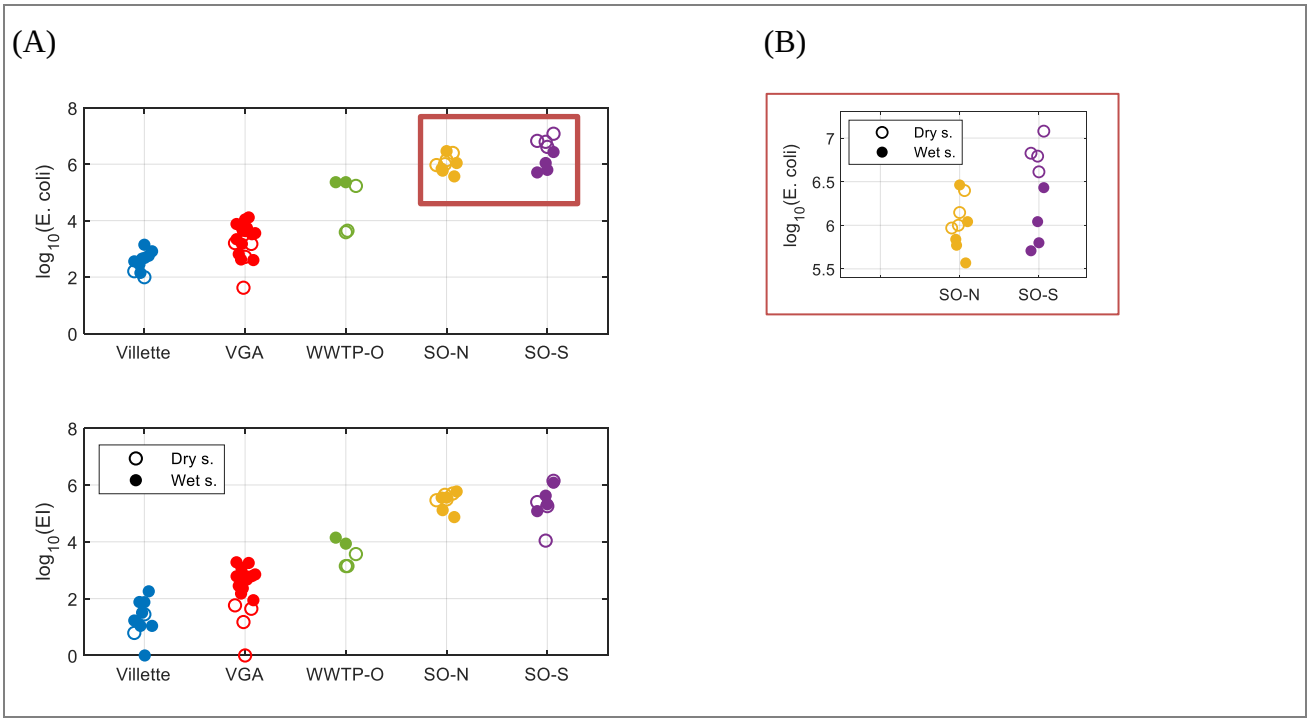
942



943

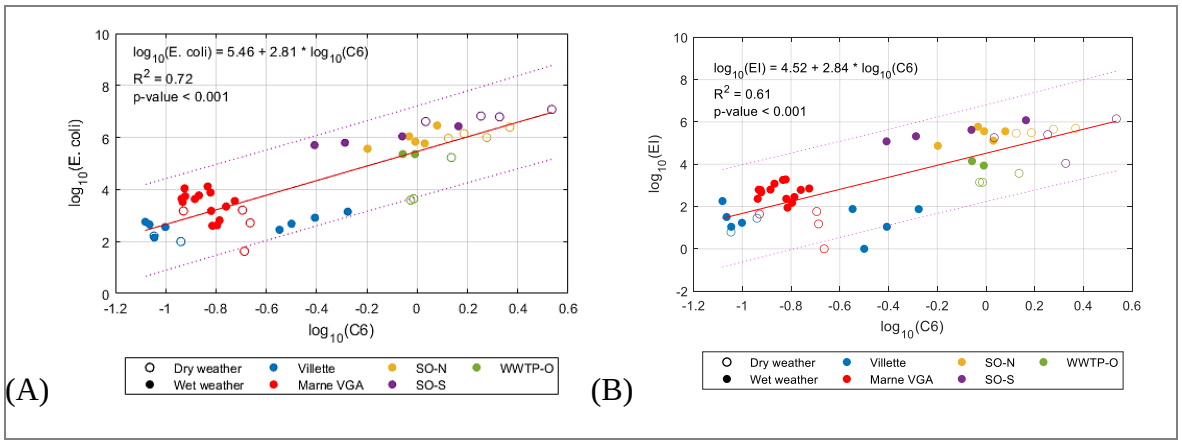
944 **Figure 9. Mean ranks of Kruskal-Wallis test of relative PARAFAC components, HIX and BIX,**
945 **for comparison between dry (n=85, red) and wet (n=98, blue) weather La Villette samples.**

946



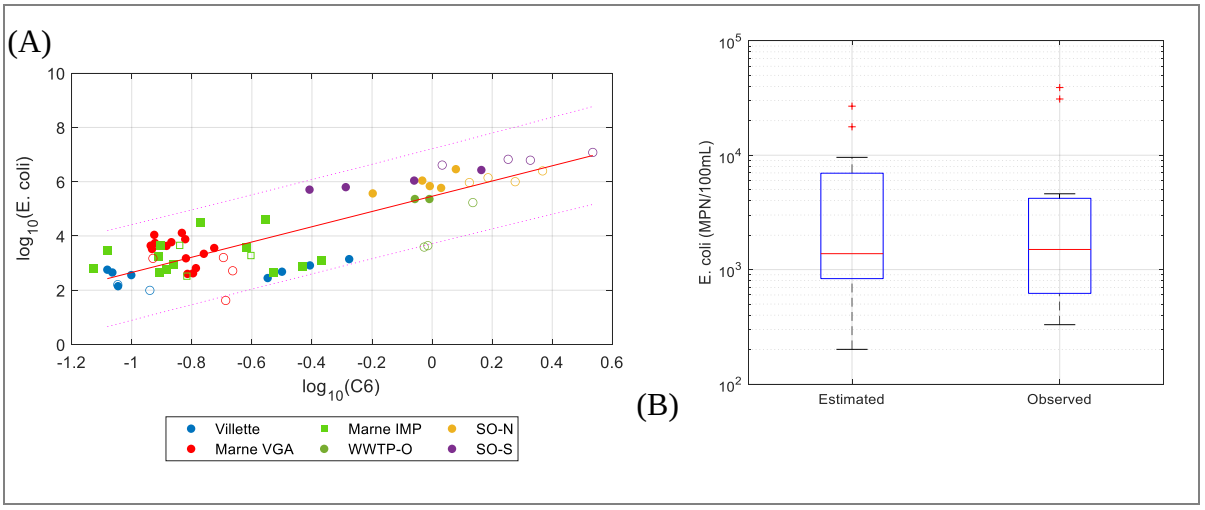
947 **Figure 10. (A) log-10 transformed FIB dataset in MPN.100mL^{-1} of the samples from La Villette**
948 **point C, the Marne VGA, WWTP-O, SO-N and SO-S; (B) zoom-in in SO-N and SO-S *E. coli***
949 **data.**

950



951 **Figure 11. Linear relationship between log$_{10}$(C6) and: (A) log$_{10}$(*E. coli*) (n=50), (B) log$_{10}$(EI)**
952 **(n=50).**

953



954 **Figure 12. (A) *E. coli* and C6 relationship with 2023 additional values superimposed (green**
955 **squares). Wet weather in filled markers and dry weather in empty markers; (B) Estimated**
956 ***E. coli* from C6 values and observed from additional samples (n=16).**

957

958 **Table 1. DOM characteristics associated with HIX values calculated according to Zsolnay et al.**
959 **(1999) and Ohno (2002) expressions. Adapted from Huguet et al. (2009).**

$HIX_{Zsolnay}$	**HIX_{Ohno}**	**DOM characteristics from HIX**	**DOM characteristics from BIX**	**BIX**
10-16	0.91-0.94	Strong humic character/important terrigenous contribution	Low biological activity	0.6-0.7
6-10	0.86-0.91	Important humic character and weak recent autochthonous component	Average / Intermediate biological activity	0.7-0.8
4-6	0.80-0.86	Weak humic character and important recent autochthonous component	High biological activity	0.8-1
<4	<0.80	Biological or aquatic bacterial origin	Biological or aquatic bacterial origin	>1

960 NOTE: In Huguet et al. (2009), "biological activity" is used as equivalent to "autochthonous
961 component".

962

963 **Table 2. Summary of studies with relationships between *E. coli* and tryptophan-like**
964 **fluorescence. The relationship type "binary classification" is established for the overpassing of**
965 **a defined threshold.**

Water type	n	Relationship type	Correlation factor/ accuracy	Reference
River and sewage treatment effluent (West Midlands, UK)	46	Linear	r=0.87	(Cumberland et al., 2012)
River (KwaZulu Natal, South Africa)	136	Linear	r=0.75	Baker et al., (2015)
Karst aquifer springs (Vorarlberg, Austria)	n_1=96 n_2=50 n_3=45	Rank correlation	$r_{Spearman1}$= 0.58 $r_{Spearman2}$=0.08 $r_{Spearman3}$=0.16	(Frank et al., 2018)
Water points (wells and borehole, Kwale County, Kenya)	162	Rank correlation	$r_{Kendall}$=0.59	(Nowicki et al., 2019)
River (Colorado, USA)	298	Binary classification for *E. coli*>10 CFU*/100mL	83% accuracy (95% CI: 78% - 87%)	(Bedell et al., 2022)
River, SO and WWTP-O (Ile-de-France, France)	50	Linear	r=0.85	Present study

966 *CFU: colony forming units, comparable to most probable number (MPN).

967

frontiers

Supplementary Material

Fluorescence spectroscopy for tracking microbiological contamination in urban waterbodies

Natália Angelotti de Ponte Rodrigues*, Rémi Carmigniani, Arthur Guillot - Le Goff, Françoise S. Lucas, Claire Therial, Manel Naloufi, Aurélie Janne, Francesco Piccioni, Mohamed Saad, Philippe Dubois, Brigitte Vinçon-Leite*

*** Correspondence:**

Corresponding Authors: natalia.angelotti-de-ponte-rodrigues@enpc.fr, b.vincon-leite@enpc.fr

1 Supplementary information

A-S1. Additional dataset preprocessing for the use of PARAFAC modelling

The first steps of EEM dataset preprocessing include IFE correction, blank subtraction, and the conversion from AU to RU (section 2.3). To obtain better PARAFAC modelling results, the EEMs of the dataset need to be smoothed.

The reminiscent scatter signals, that were not removed with blank subtraction, were removed by the smoothing function, without interpolation. They were numerically replaced by NaN on the following extents:

- 1st order Rayleigh: 25 nm above, 30 nm below, in the emission axis.
- 1st order Raman: 10 nm above, 15 nm below, in the emission axis.
- 2nd order Rayleigh: 20 nm above, 18 nm below, in the emission axis.
- 2nd order Raman: 20 nm above, 14 nm below, in the emission axis.

The region at excitation wavelengths below 250 nm was removed due to high levels of noise. Since negative concentrations and fluorescence intensities are chemically and physically impossible, a non-negativity constraint was applied to have realistic results.

Prior to PARAFAC modelling, all EEMs of the dataset were normalised to its total signal. Then, the model was able to identify chemical variations of the sample composition independently of the magnitude of their fluorescence intensity.

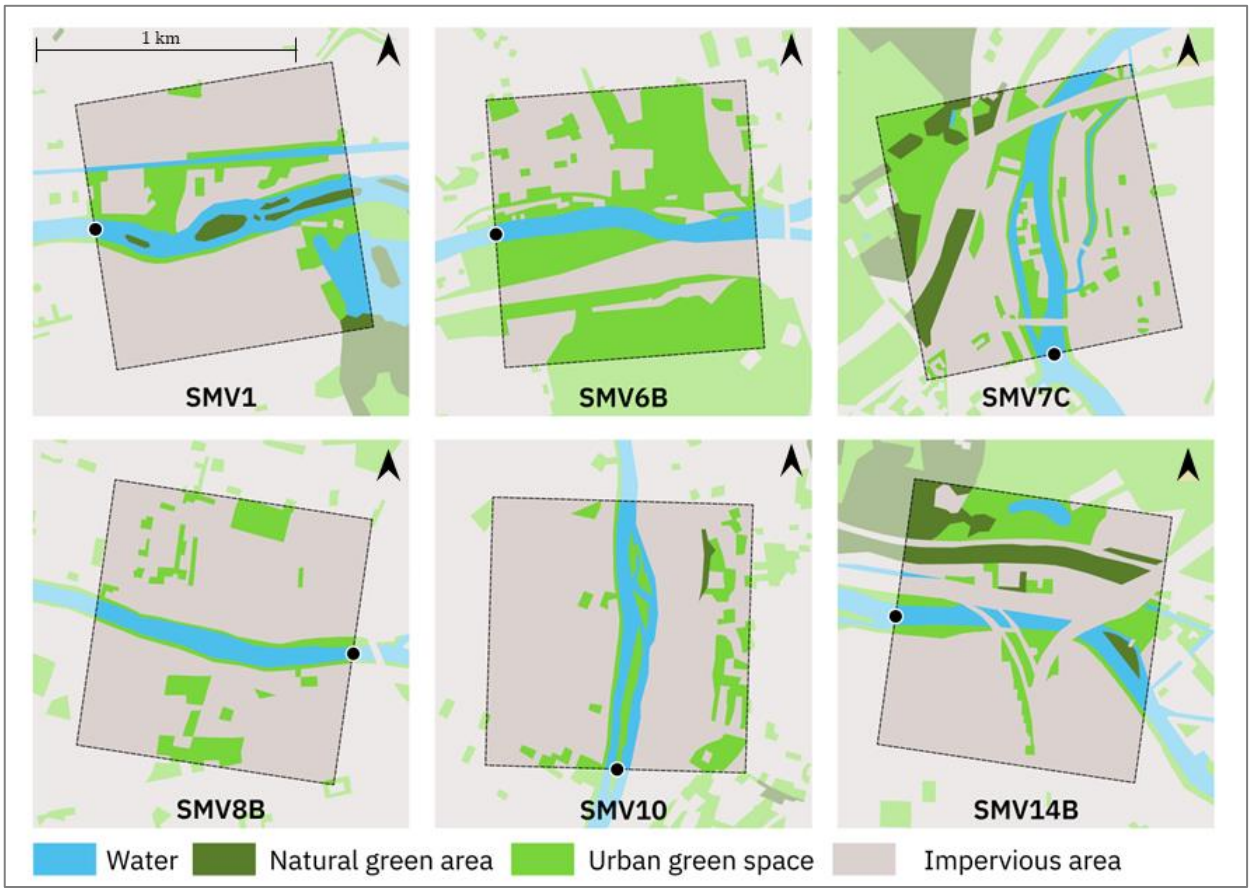
A-S2. Obtention of maximum fluorescence intensity of PARAFAC components

The fluorescence of each component at the same scale as the original EEMs, Fmax, corresponds to the maximum fluorescence calculated at the maximum emission and maximum excitation loadings of each component in each sample (Equation 1).

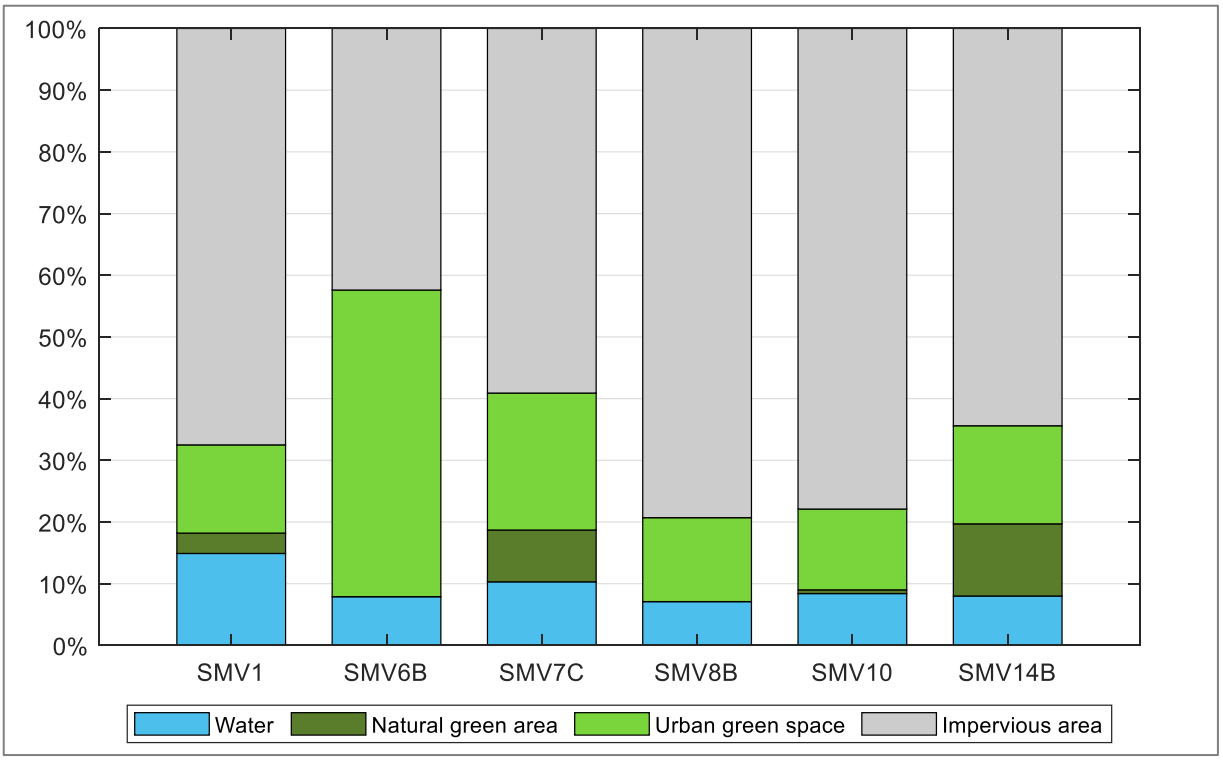
$$F_{ijkf} = a_{if}\, b_{jf}\, c_{kf}, \text{ with i} = 1, \ldots, I; \text{ j} = 1, \ldots, J; \text{ k} = 1, \ldots, K; \text{ f} = 1, \ldots, F \quad (1)$$

Where F_{ijkf} is the computed fluorescence intensity of the fth component in the ith sample, at the jth variable of the emission mode and the kth variable of the excitation mode. The a_{if}, b_{jf} and c_{kf} were defined in Equation 5. Cf, or Fmax$_f$, is F_{ijkf} at maximum loadings of b_{jf} and c_{kf} for each fth component in the ith sample.

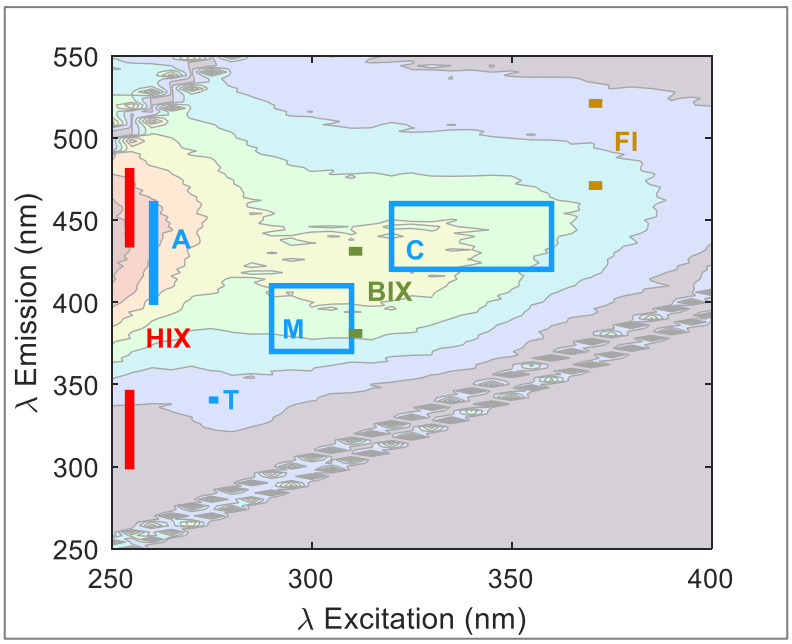
2 Supplementary Figures



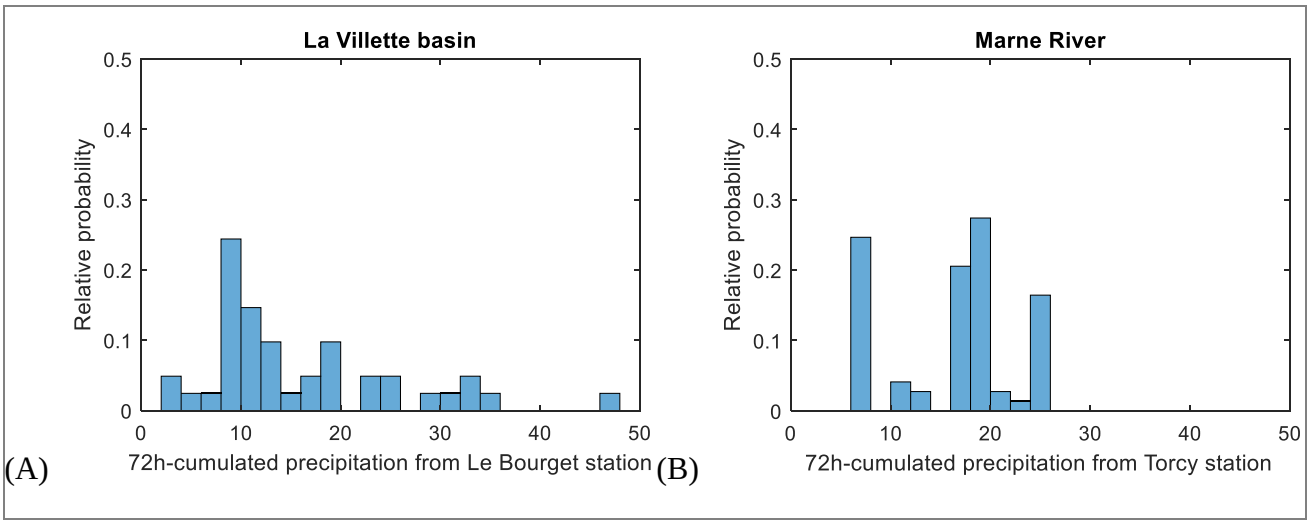
Supplementary Figure S1. Land-use of the river Marne main SMV points (Sources: IGN BDTOPO, IAU Île-de-France MOS).



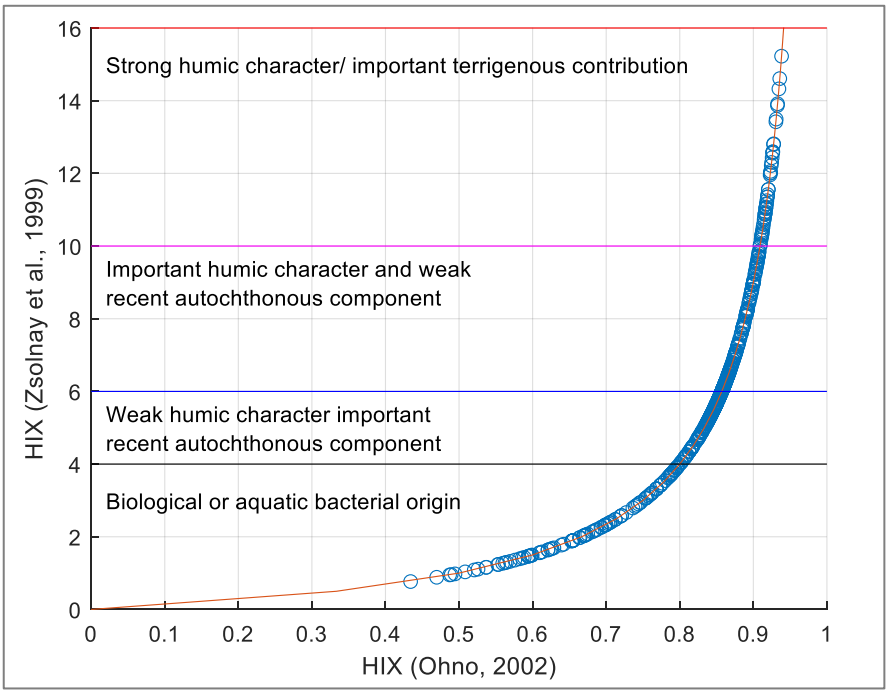
Supplementary Figure S2. Land-use (%) of the squared area of 1km² immediately upstream of the main river Marne sampling sites.



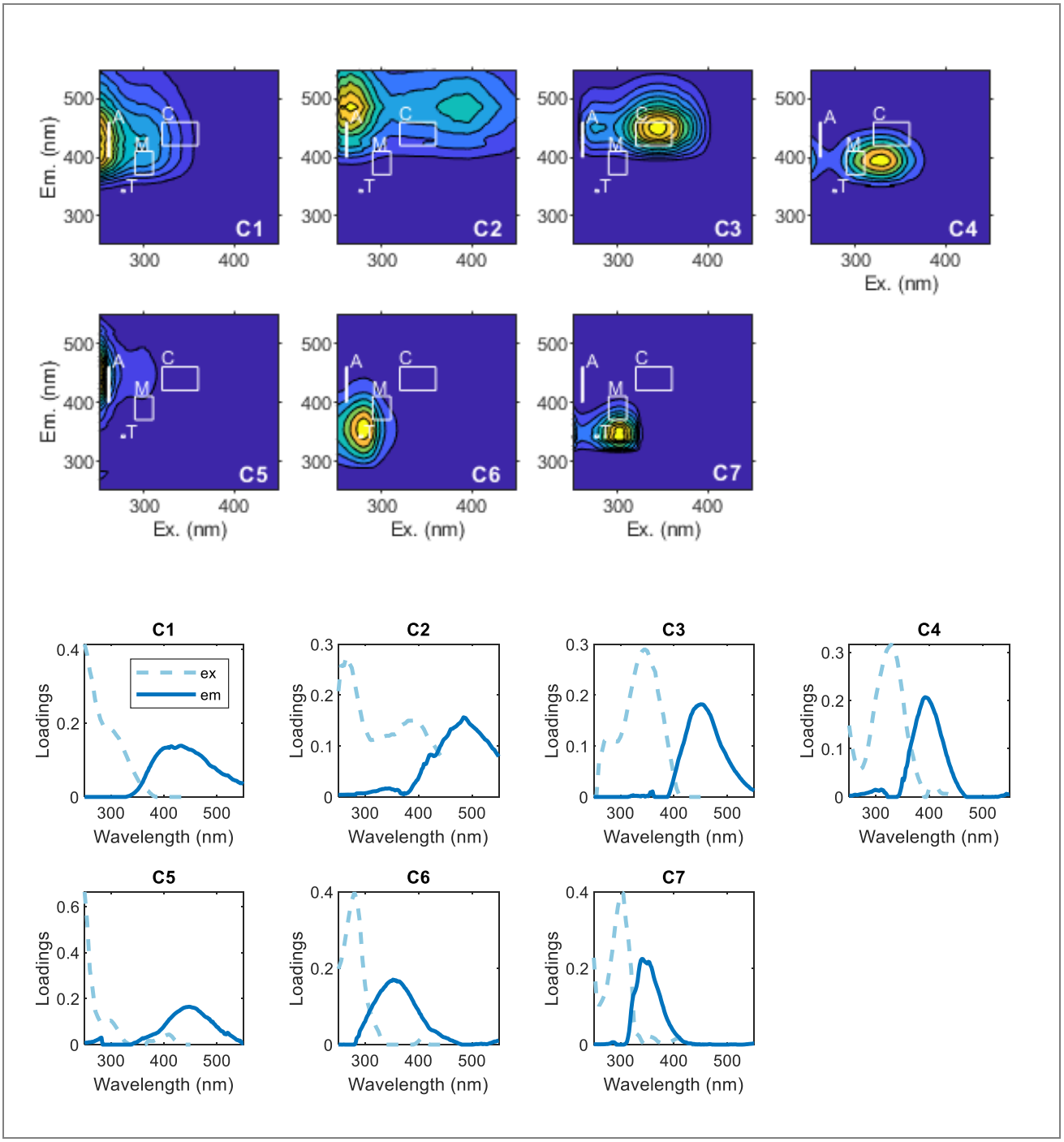
Supplementary Figure S3. Typical Excitation-Emission Matrix (EEM). Location of the fluorescence peaks and fluorescence intensities used for indices computation are depicted in colour: blue for peaks T, M, C and A; red for HIX; green for BIX and brown for FI.



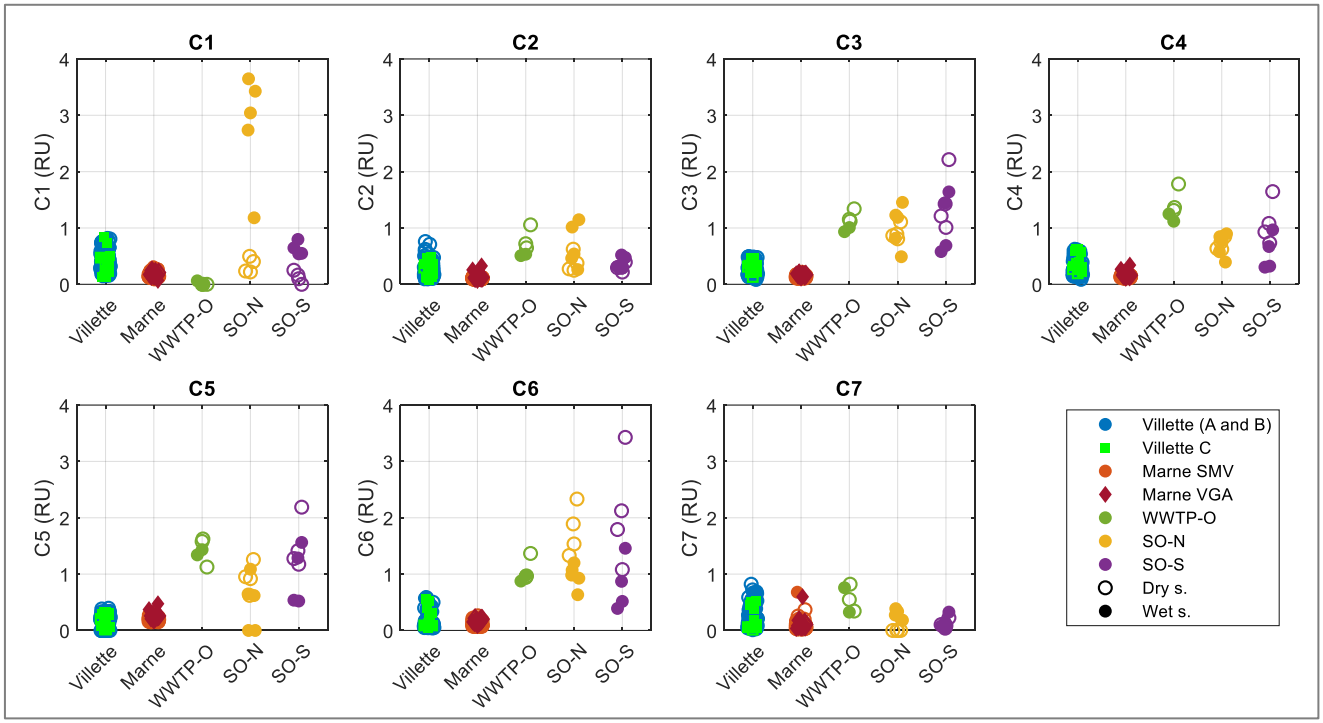
Supplementary Figure S4. Distribution of relative probability of 72h-cumulated precipitation before sampling in La Villette basin (A) and the river Marne (B).



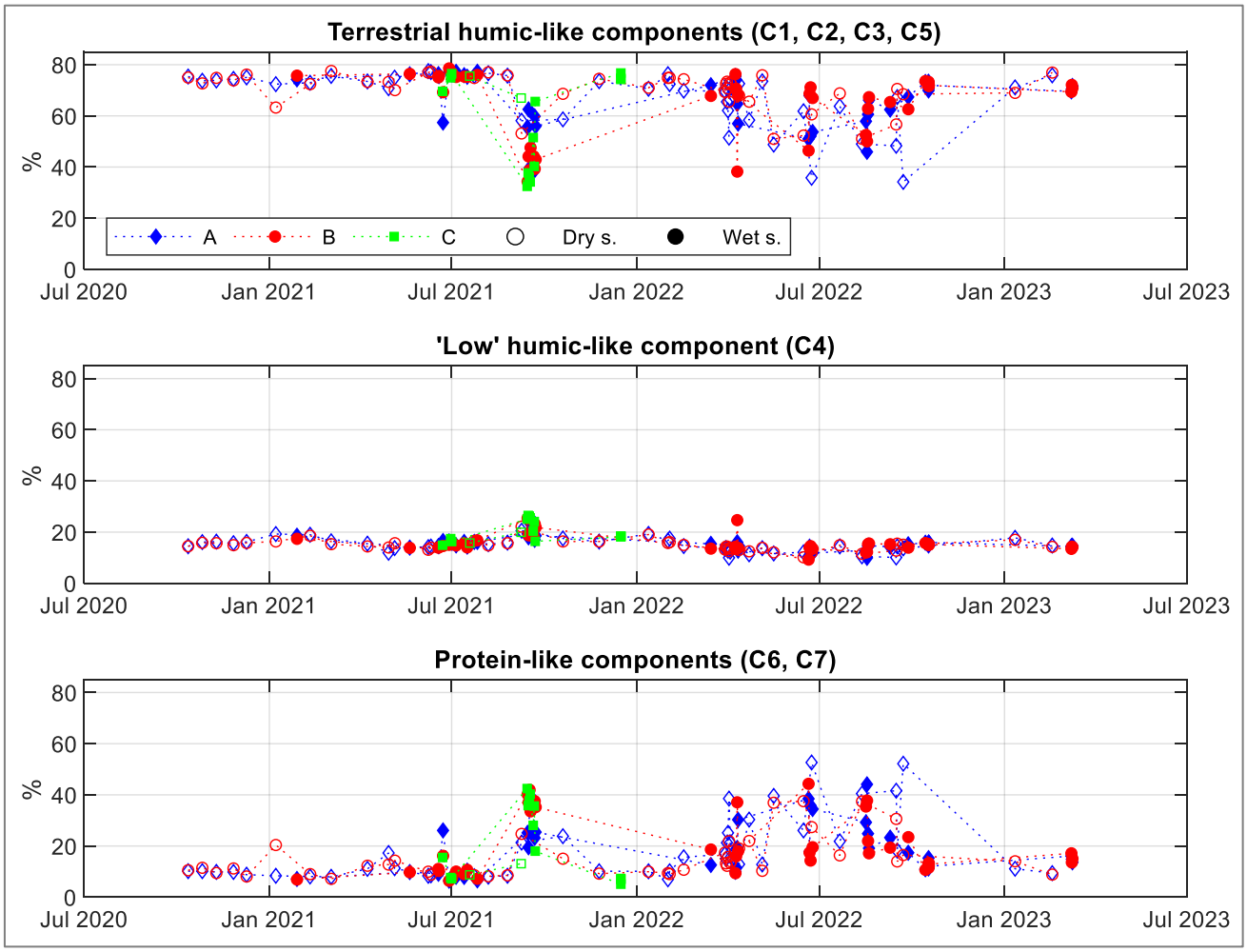
Supplementary Figure S5. HIX values calculated according to Ohno (2002) and Zsolnay *et al.* (1999). The dataset included samples from La Villette, the Marne, SO-N, SO-S and WWTP-O.



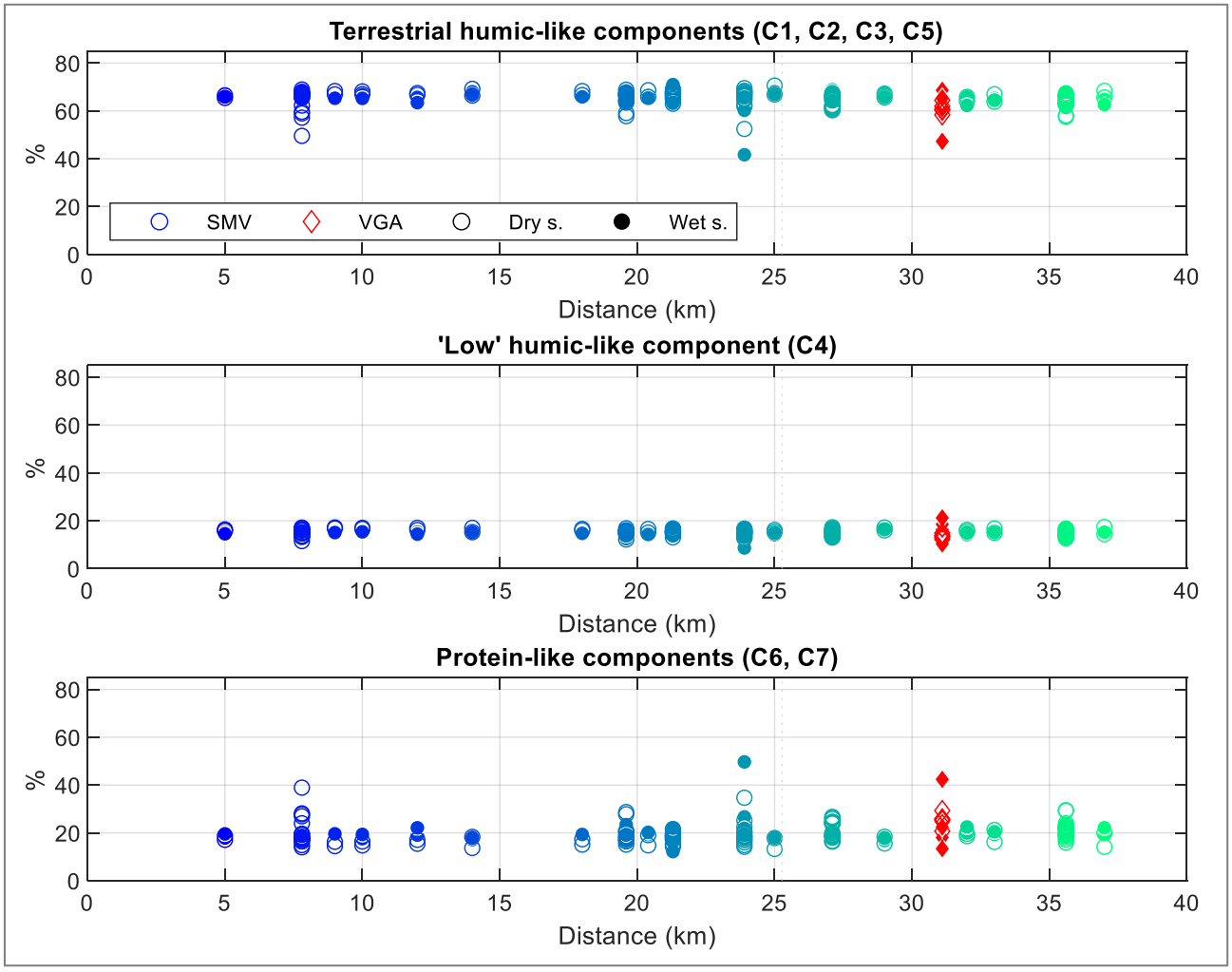
Supplementary Figure S6. EEMs with the loadings of the 7-component PARAFAC model (top) and spectral loadings results (bottom).



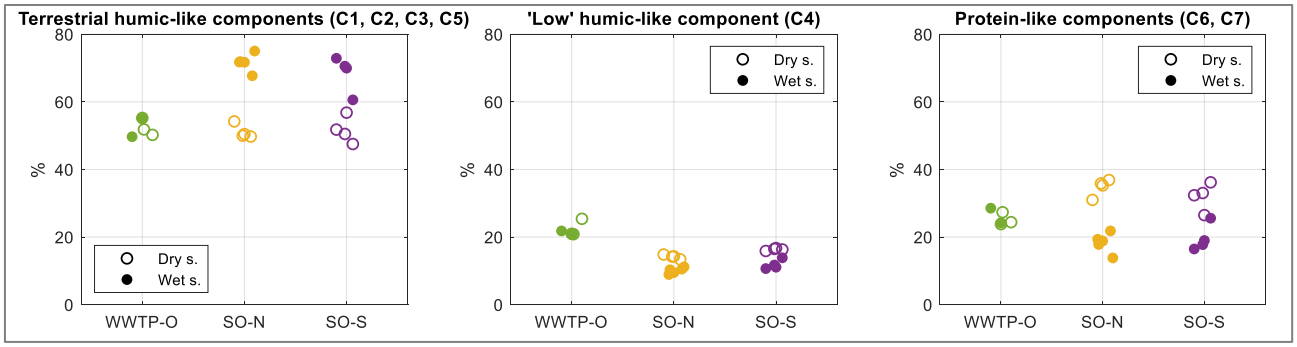
Supplementary Figure S7. Fluorescence intensity of PARAFAC components (in RU) in La Villette basin points A (blue), B (blue) and C (bright green), in the river Marne points SMV (orange) and VGA (dark red), in the outlet of WWTO-O (dark green) and the outlets from stormwater networks SO-N.



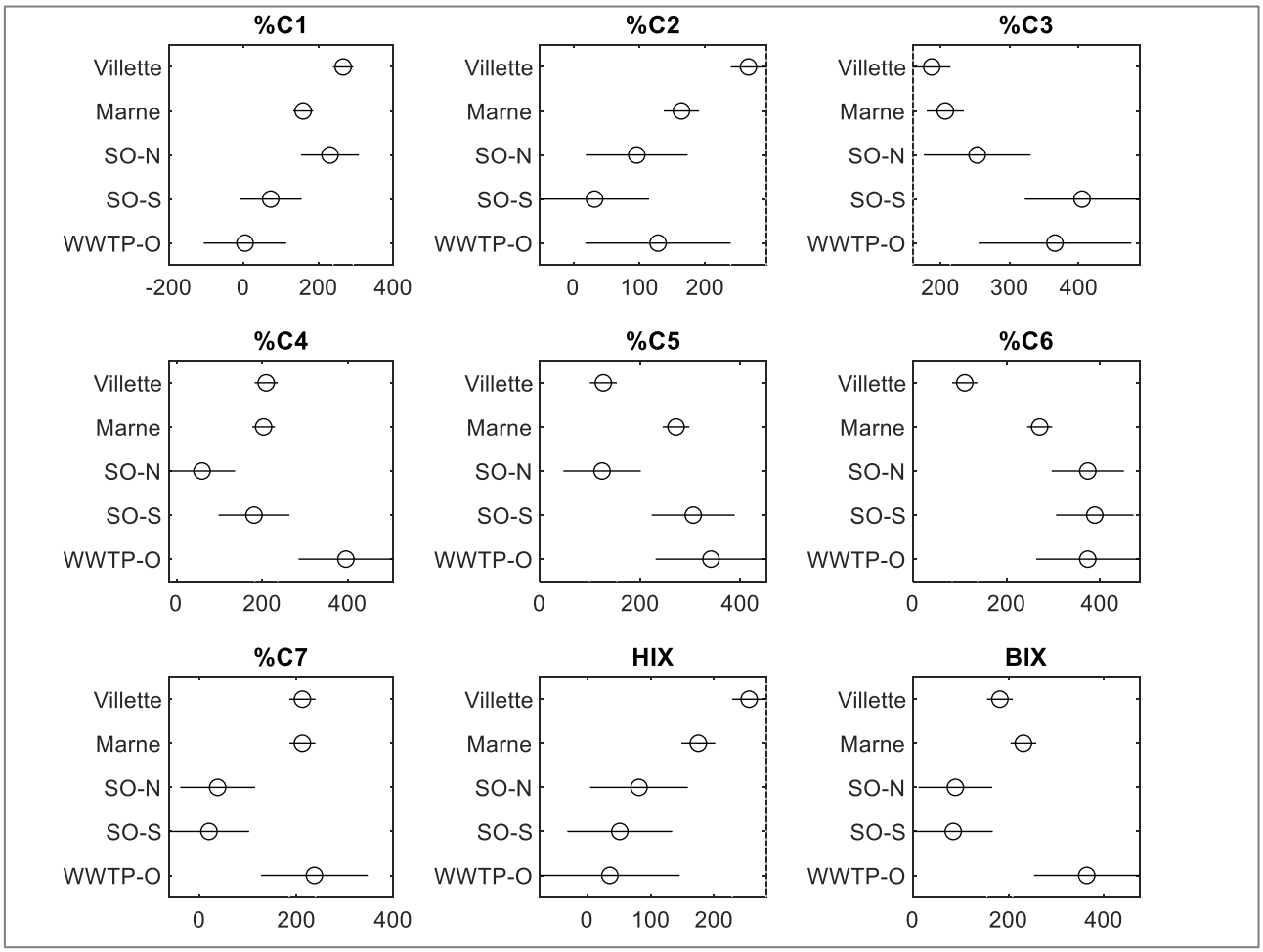
Supplementary Figure S8. Sum of the relative abundance of PARAFAC components (in %), grouped in terrestrial humic-like (top), "low" humic-like (middle) and protein-like (bottom), at La Villette basin sampling points A, B and C in time, for dry and wet weather samples.



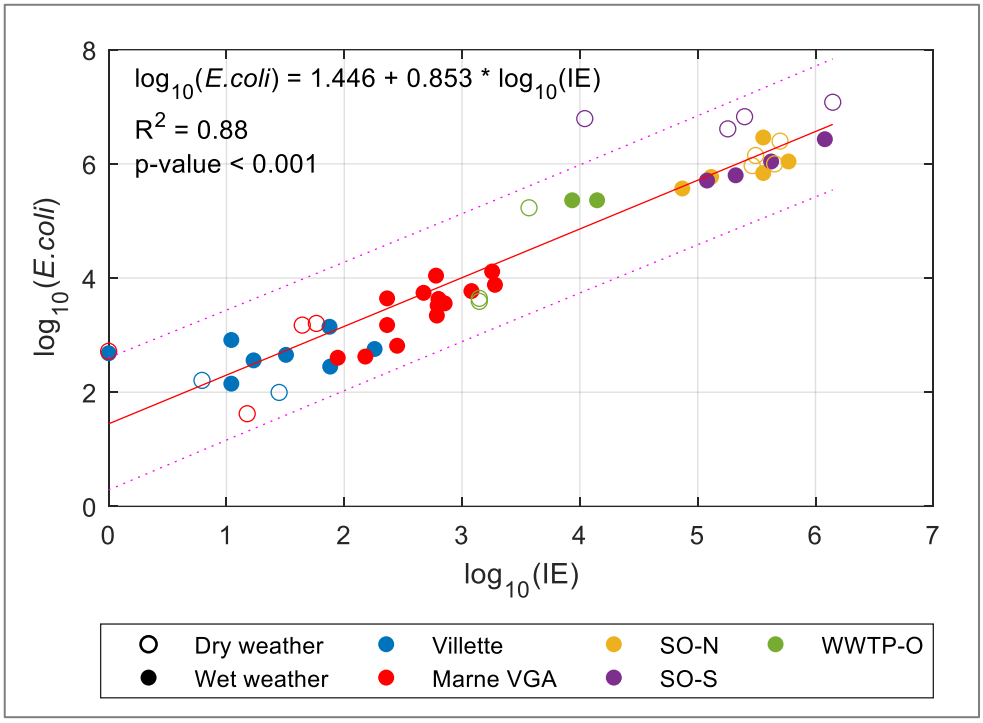
Supplementary Figure S9. Sum of relative abundance of PARAFAC components (in %), grouped in terrestrial humic-like (top), "low" humic-like (middle) and protein-like (bottom), at the river Marne sampling points for dry and wet weather samples, with VGA in red.



Supplementary Figure S10. Sum of relative abundance of PARAFAC components (in %), grouped in terrestrial humic-like (left), "low" humic-like (middle) and protein-like (right), of the outlets from stormwater networks (SO-N and SO-S) and WWTP (WWTP-O).



Supplementary Figure S11. Kruskal-Wallis mean ranks of median relative abundance of PARAFAC components (in %) and fluorescence indices HIX and BIX of all water types.



Supplementary Figure S12. Relationship between $\log_{10}$-transformed *E. coli* and IE (n=50).

3 Supplementary Tables

Supplementary Table S1. Names, coordinates and distance from closest point of interest of sampling sites (other sampling point, stormwater network or WWTP outlet upstream). Latitude and longitude are given in decimal degrees North and East, respectively.

Site ID	Site name and city	Water type	Latitude (N)	Longitude (E)	Distance from the closest point of interest
SO-N	Urban centre of Noisy-le-Grand	Stormwater network outlet	48.852811	2.541739	~2km upstream WWTP-PMA.
SO-S	Sucy-en-Brie	Stormwater network outlet	48.771383	2.508284	Upstream points VGA and SMV11.
WWTP-O	Station of Lagny-sur-Marne Saint-Thibault-das-Vignes	Wastewater treatment plant outlet	48.875812	2.672124	~7km upstream point SMV1
A	Upstream La Villette canal, Paris	Canal	48.891755	2.385479	~20 km downstream the WWTP from Villeparisis
B	La Villette bathing area, Paris	Canal	48.885802	2.373753	1.1 km downstream from point A
C	Crimée bridge, Paris	Canal	48.888371	2.379246	0.6 km downstream from point A
SMV1	Chelles	River	48.865599	2.583061	7.8 km downstream from outlet WWTP-O.
SMV6B	Nogent-sur-Marne	River	48.830919	2.477830	19.6 km downstream WWTP-O. 11.8 km downstream point SMV1.
SMV7C	Joinville-le-Pont	River	48.819127	2.469216	21.3 km downstream WWTP-O. 1.7 km downstream point SMV6B.

Site ID	Site name and city	Water type	Latitude (N)	Longitude (E)	Distance from the closest point of interest
SMV8B	Champigny-sur-Marne	River	48.810578	2.502057	23.9 km downstream WWTP-O. 2.7 km downstream point SMV7C.
SMV10	Saint-Maur-des-Fossés	River	48.792178	2.522323	1.8 km downstream discharge outlet of WWTP-PMA. 3.2 km downstream point SMV8B.
VGA	VGA Voile Club, Saint-Maur	River	48.788302	2.474838	5.8 km downstream discharge outlet of WWTP-PMA. Immediately upstream SMV12 and 4.0 km downstream SMV10.
SMV14B	Saint-Maurice	River	48.816339	2.441363	10.3 km downstream discharge outlet of WWTP-PMA. 8.5 km downstream point SMV10.

Supplementary Table S2. Land-use (%) and mean population density (people.km^{-2}) of the squared area of 1km² immediately upstream of each sampling site. "Agriculture" and "Forest" (green) were grouped in "natural green areas".

River Marne sampling points	Dist. (km)	Water	Agriculture	Forest	Natural green areas	Urban green space	Impervious area	Mean pop. density (people/km²)
SMV1	7.8	15%	0%	3%	3%	14%	68%	477.9
SMV6B	19.6	8%	0%	0%	0%	50%	42%	854.7
SMV7C	21.3	10%	2%	7%	9%	22%	59%	714.8
SMV8B	23.9	7%	0%	0%	0%	14%	79%	780.0
SMV10	27.2	8%	0%	1%	1%	13%	78%	437.2
SMV14B	35.6	8%	0%	12%	12%	16%	64%	742.7
Average	-	*9%*	*0%*	*4%*	*4%*	*21%*	*65%*	*667.9*

Supplementary Table S3. Summary of sample collection

Point	Sampling period	Sampling type	Total number of samples (dry + wet weather)	Sample preparation	
				Dilution factor	Filtration
Villette A	10/10/2020 to 08/03/2023	Grab samples	85 (40 + 45)	1	No
Villette B	10/10/2020 to 08/03/2023	Grab samples	85 (40 + 45)	1	No
Villette C	30/06/2021 to 22/09/2021	24-hour integrated	10 (2 + 8)	1	No
Villette C	22/06/2021 and 16/12/2021	Grab samples	5 (5 + 0)	1	No
Marne VGA	15/06/2022 to 20/10/2022	24-hour integrated	18 (4 + 14)	1	No
Marne SMV	06/07/2022 to 08/09/2022	Grab samples	189 (125 + 64)	1	No
SO-N	04/05/2022 to 09/09/2022	24-hour integrated	9 (4 + 5)	1/5	Yes
SO-S	19/09/2022 to 07/11/2022	24-hour integrated	8 (4 + 4)	1/5	Yes
WWTP-O	30/06/2021 to 04/10/2021	24-hour integrated	5 (3 + 2)	1/5	No

Supplementary Table S4. Fluorescence indices (mean ± SD and range) of the different sampling sites and the total number of samples (N) per site.

Water type	HIX (-)		BIX (-)		FI (-)	
	Mean ± SD	**Range**	**Mean ± SD**	**Range**	**Mean ± SD**	**Range**
Villette (n=183)	0.86 ± 0.08	[0.55-0.94]	0.90 ± 0.26	[0.64-2.19]	1.97 ± 0.07	[1.76-2.28]
Marne (n=207)	0.85 ± 0.03	[0.71-0.90]	0.87 ± 0.09	[0.73-1.62]	2.13 ± 0.08	[1.90-2.40]
WWTP-O (n=5)	0.76 ± 0.03	[0.72-0.79]	1.09 ± 0.08	[1.00-1.17]	2.41 ± 0.09	[2.29-2.53]
SO-N (n=9)	0.77 ± 0.09	[0.65-0.86]	0.76 ± 0.06	[0.63-0.82]	2.25 ± 0.32	[1.93-2.79]
SO-S (n=8)	0.75 ± 0.07	[0.66-0.85]	0.74 ± 0.08	[0.63-0.85]	2.56 ± 0.15	[2.30-2.79]

Supplementary Table S5. Designation, excitation (Ex) and emission (Em) wavelengths peaks of PARAFAC components, number of references with matching components from OpenFluor Database (checked on 30/06/2023), with minimum similarity score (Tucker congruence) of 0.95 for excitation and emission wavelength loads (Murphy et al., 2014).

PARAFAC component	Ex (nm)	Em (nm)	%Fmax range	Matches in OpenFluor (0.95)	Description and references
C1	250	432	0.0-44.5	79	Terrestrial humic-like, similar to peak A. Frequently found in lentic freshwater (C4) (Marcé et al., 2021); widespread in freshwaters (C1) (Lambert et al., 2017).
C2	265	484	3.1-26.8	89	Terrestrial humic-like, associated with high aromaticity and molecular weight (C1) (Lambert et al., 2016); plant-derived material (C2) (Yan et al., 2020); associated mainly with peak A.
C3	345	450	6.0-25.1	14	Terrestrial humic-like, similar to peak C, widespread in freshwaters, (C2) (Lambert et al., 2017); associated with aromatic molecules of high molecular weight (C3) (Lambert et al., 2016).
C4	330	392	8.6-26.5	5	Humic-like, low molecular weight, close to peaks C and M; potentially related to agricultural catchments (C6) (Graeber et al., 2012); associated to humic-like substances produced *in-situ* by phytoplankton and microbial activity (C4) (Galletti et al., 2019).
C5	250	448	0.0-27.6	9	Terrestrial humic-like, estuarine streams of agricultural catchment (C6) (Asmala et al., 2018); associated with peak A (C3) (Li et al., 2015).
C6	280	352	2.7-36.9	44	Protein-like, similar to peak T. Indicative of recent autochthonous production, commonly related to anthropogenic occupation (C7) (Lambert et al., 2017) and WWTP effluents (C1) (Cohen et al., 2014); aliphatic (non-aromatic) with low molecular weight (C6) (Lambert et al., 2016).
C7	305	340	0.0-41.8	21	Protein-like, close to peak T; might be associated with sediment-derived or autochthonous tyrosine, and may be a potential indicator of eutrophication (C3) (Ren et al., 2021); indicator of DOM bioavailability (C4) (Ryan et al., 2022).

Supplementary Table S6. Relative abundance of PARAFAC components (mean ± SD) of the sampling sites (n=number of samples).

Water type	%C1 (%)	%C2 (%)	%C3 (%)	%C4 (%)	%C5 (%)	%C6 (%)	%C7 (%)
Villette (n = 183)	22.4 ± 6.6	13.9 ± 4.1	14.7 ± 2.9	15.6 ± 3.2	15.0 ± 6.5	6.8 ± 4.0	11.7 ± 9.6
Marne (n = 204)	17.8 ± 2.7	11.5 ± 1.1	15.0 ± 1.0	14.9 ± 1.3	20.8 ± 1.6	11.8 ± 2.3	8.2 ± 4.4
SO-N (n = 9)	24.9 ± 17.8	8.6 ± 3.3	16.3 ± 2.6	11.9 ± 2.3	12.7 ± 8.4	23.4 ± 11.0	2.3 ± 2.3
SO-S (n = 8)	9.0 ± 8.4	6.9 ± 2.7	22.5 ± 1.6	14.1 ± 2.6	21.7 ± 2.6	23.6 ± 8.0	2.2 ± 1.2
WWTP-O (n = 5)	0.2 ± 0.5	11.0 ± 2.4	17.9 ± 1.2	21.9 ± 2.0	23.2 ± 4.2	16.5 ± 2.1	9.1 ± 3.7

Supplementary Table S7. Fluorescence indices (mean ± SD) in the main group of 6 points along the river Marne. Distance from the nearest WWTP outlet upstream of each sampling point (DW).

Sampling point (n=15/ point)	**DW (km)**	**%C1**	**%C2**	**%C3**	**%C4**	**%C5**	**%C6**	**%C7**	**HIX**	**BIX**
SMV1	7.8	17.2 ±1.8	12.3 ±1.1	14.7 ±1.2	15.6 ±1.3	21.0 ±1.9	10.5 ±0.9	8.7 ± 7.0	0.85 ±0.03	0.89 ±0.13
SMV6B	19.6	17.8 ±2.4	12.0 ±0.7	15.1 ±0.7	15.2 ±0.8	21.0 ±0.8	11.7 ±0.8	7.1 ± 2.7	0.85 ±0.02	0.84 ±0.06
SMV7C	21.3	18.9 ±2.0	11.9 ±0.6	15.1 ±0.5	15.2 ±0.8	20.6 ±0.7	11.2 ±0.9	7.1 ± 1.7	0.86 ±0.01	0.86 ±0.03
SMV8B	23.9	18.8 ±2.0	11.6 ±0.9	14.6 ±1.0	14.9 ±1.2	20.1 ±1.5	10.7 ±1.3	9.3 ± 5.7	0.85 ±0.02	0.88 ±0.11
SMV10	1.8	17.7 ±2.3	11.7 ±0.7	15.5 ±0.5	15.6 ±0.9	21.1 ±1.0	11.2 ±0.7	7.1 ± 1.1	0.86 ±0.01	0.85 ±0.03
SMV14B	10.3	18.5 ±2.6	11.0 ±0.7	15.3 ±0.6	15.2 ±0.9	20.3 ±1.0	11.4 ±0.7	8.2 ± 2.0	0.85 ±0.01	0.87 ±0.05
Global	**-**	**18.2 ±2.2**	**11.8 ±0.8**	**15.1 ±0.8**	**15.3 ±1.0**	**20.7 ±1.1**	**11.1 ±0.9**	**7.9 ±3.4**	**0.9 ±0.0**	**0.9 ±0.1**