Automated Monitoring of Organic and Bacterial Drinking Water Constitution via PARAFAC Fluorescence Spectroscopy and Flow Cytometry

Dissertation (monograph)

approved by the Doctoral Degree Committee of

Hamburg University of Technology

in pursuit of the academic degree of

Doktor-Ingenieur (Dr.-Ing.)

written by

Jonas Schuster

from

Langen (Hessen)

2025

Reviewers

1. Reviewer: Prof. Dr. Mathias Ernst

2. Reviewer: Prof. Dr. Manuela Antonelli

Chairman of the Examination Committee

Prof. Dr. Kay Smarsly

Date of the oral examination

February 12, 2025

DOI

https://doi.org/10.15480/882.14887

ORCID ID

https://orcid.org/0000-0001-5077-5312

Acknowledgements

First, I would like to express my gratitude to Mathias for providing the framework for this research, for his trust in me, and for the valuable professional discussions we had. My sincere thanks also go to the entire team at the Institute of Water Resources and Supply as well as the DVGW Research Centre. A special thanks is owed to Dorota and Thorsten for the exchange, their support in the laboratory, and for creating such a pleasant atmosphere in the lab spaces.

An indescribably great thank you goes to all the students I had the privilege of supervising during my doctoral research. I thoroughly enjoyed working with you, learned a lot from each of you, and sincerely hope you had a rewarding experience. This extensive work would not have been possible without you. My heartfelt thanks go to Eleni, Hanna, Helana, Johanna, Jon, Juri, Lars, Lena, Mehrad, Ole, Parham, Philip, Priyanka, Sam, Shabnam, and Xia.

For the invaluable interdisciplinary and international experiences, I was fortunate to gain during my work, I would like to extend my special thanks to our project partners from Technion in Haifa and TU Ilmenau. The regular discussions with Leonid and Hao were a refreshing change from the daily lab routine at TUHH. I would also like to thank my colleagues at Hamburg Wasser for their collaboration. In this context, a special thanks goes to Mike for providing the pilot plant and for his hands-on support during its modification.

Further thanks go to Prof. Antonelli for reviewing my dissertation and to Prof. Smarsly for taking on the role of examination chair.

Beyond professional support, many people have shaped me personally, and I sincerely thank them, as this work would not have been possible without them.

Starting with my fellow PhD colleagues at the institute, I can confidently say that our mutual support and camaraderie were truly exceptional. I am immensely proud of what we have achieved together and the ways we have helped each other along the way. We were, are, and always will be a great team. A special shout-out goes to Charlotte, the two Jakobs, and Tomi, who were always there for me and made my time in Harburg all the more enjoyable. I would also like to thank my two soil-colleagues, Lukas and Claas, for our regular, fruitful discussions.

Next, my heartfelt appreciation goes to all the wonderful, important friends and people I have met along my journey through Langen, Karlsruhe, Munich, Gothenburg, and Hamburg. Each and every one of you knows that good friendships and social interactions mean more to me than work and high-impact publications. I feel incredibly fortunate to know so many amazing people with whom I share countless individual memories and experiences. You all played a crucial role in helping me navigate these last four years and successfully complete this dissertation.

Last but by no means least, a huge thank you to my dear family, who have always supported me and had my back in everything I do.

Abstract

Ensuring high drinking water quality is of major public health concern, protecting consumers from harmful consequences. One relevant concern about drinking water quality is the content of dissolved organic matter (DOM) which can be related to source water quality and treatment needs. Additionally, the bacterial constitution is crucial to monitor evolving changes in time and respond appropriately. However, DOM characterizing methods are a compromise between analytical effort, time and data quality while bacterial methods such as cultivation are labor-intensive, time-consuming and difficult to adapt to online measurements. Therefore, developing an automated monitoring system that quickly and consistently analyzes water quality data with high detail is vital.

This work examines the application of parallel factor analysis (PARAFAC) fluorescence spectroscopy to characterize DOM content and flow cytometry to monitor the presence of bacteria in drinking water. The specific research objectives of this work were mainly to further develop, combine and automate, and test both methods in real-world applications.

Using PARAFAC fluorescence spectroscopy, DOM compositions of investigated drinking waters, originating from groundwater in Northern Germany, were described through a respective fluorescence score of six compounds (C1–C6). Additionally, the total cell count (TCC) and the proportion of high nucleic acid cells (%HNA) were determined via flow cytometry describing the bacterial constitution. Scenarios of water quality changes were simulated in the laboratory by spiking one specific drinking water with increasing amounts of water samples of varying water quality. Water quality could be described by generating parameter-specific baselines defining their thresholds. Strongly deviating waters, e.g., wastewater effluent and rainwater, were identified in lower volume proportions than less strongly deviating waters. Regarding sensitivity, C1–C3 and TCC were the most performant for detecting water quality changes, e.g., due to a contamination event.

Automation of both methods required hardware and software extensions and development. Sampling, data analysis, evaluation, and visualization were the automation objects addressed in this work. In the context of near real-time drinking water analysis, the time required from sampling to data visualization was reduced to less than 15 minutes.

In pilot plant trials utilizing a model drinking water distribution system, combined methods were tested to detect water quality changes in a flowing system. Tests were performed to simulate events of water quality change. The combined system provided characteristic fingerprints of flowing water and detected sudden changes. 300 m behind the injection point, continuous induction of wastewater effluent and rainwater could be detected under appropriate conditions. In a large-scale measurement campaign, a three-component PARAFAC model (c1–c3) was proven to characterize groundwater DOM individually and can be applied to estimate humic substances concentrations in groundwater.

In conclusion, it can be stated that PARAFAC fluorescence spectroscopy in conjunction with flow cytometry represents a rapid and powerful system for the comprehensive characterization of DOM and the presence of bacteria in both drinking water and groundwater. The continuous and automatic monitoring of the emphasized parameters enables to recognize even minor deviations in water characteristics

Zusammenfassung

Die Sicherstellung einer hohen Trinkwasserqualität ist essentiell, um Verbraucher vor gesundheitsschädlichen Folgen zu schützen. Wesentliche Aspekte der Trinkwasserqualität sind der Gehalt an gelösten organischen Substanzen (DOM), der mit der Rohwasserqualität und entsprechender Trinkwasseraufbereitung in Zusammenhang steht, und die bakterielle Konstitution des Trinkwassers. DOM Charakterisierungsmethoden sind jedoch ein Kompromiss zwischen analytischem Aufwand, Zeit und Datenqualität, während bakterielle Methoden wie die Kultivierung arbeitsintensiv, zeitaufwendig und schwer als Online-Messungen zu integrieren sind. Die Entwicklung eines automatisierten Monitoring-Systems, das Wasserqualitätsdaten schnell und konsistent mit hoher Detailgenauigkeit analysiert, ist daher erstrebenswert.

Diese Arbeit untersucht die Anwendung der Fluoreszenzspektroskopie mit paralleler Faktorenanalyse (PARAFAC) zur Charakterisierung des DOM-Gehalts und die Durchflusszytometrie zur Überwachung der Bakterienzahl im Trinkwasser. Die Forschungsziele dieser Arbeit bestanden darin, beide Methoden weiterzuentwickeln, zu kombinieren und zu automatisieren sowie in realen Anwendungen zu testen.

Durch die PARAFAC-Fluoreszenzspektroskopie wurde der DOM-Gehalt der untersuchten Trinkwässer, die aus Grundwasserquellen in Norddeutschland stammten, anhand von sechs Substanzen (C1–C6) mit korrespondierendem Fluoreszenz-Score beschrieben. Mittels Durchflusszytometrie wurden die Gesamtzellzahl (TCC) und der Anteil an Zellen mit hohem Nukleinsäure-Gehalt (%HNA) bestimmt. Im Labor wurden Trinkwasserproben mit zunehmender Konzentration variierender Wässer versetzt um gezielt Wasserqualitätsänderungen zu simulieren. Parameterspezifische Baselines definierten dabei Schwellenwerte der Trinkwasserqualität. Stark abweichende Wässer, z.B. Abwasser und Regenwasser, wurden in geringeren Volumenanteilen erkannt als weniger stark abweichende Wässer. C1–C3 und TCC zeigten die höchste Leistungsfähigkeit bei der Erkennung von Wasserqualitätsänderungen.

Die Automatisierung beider Methoden erforderte Hardware- und Softwareerweiterungen und -entwicklungen. Dabei lag der Fokus auf Probenahme, Datenanalyse, Auswertung und Visualisierung beider Methoden. Im Kontext der nahezu Echtzeit-Analyse von Trinkwasser wurde die Zeit von der Probenahme bis zur Datenvisualisierung auf weniger als 15 Minuten verkürzt.

Um Wasserqualitätsänderungen in einem Fließsystem zu erkennen wurden beide Methoden mittels Pilotversuchen in einem Modell-Leitungssystem evaluiert. Dabei wurden gezielt Ereignisse von Wasserqualitätsänderungen simuliert. Das Monitoring-System lieferte charakteristische Fingerabdrücke des fließenden Wassers und erkannte unmittelbare Veränderungen. Zum Beispiel konnte unter gegebenen Bedingungen der Pilotanlage die Einleitung von Abwasser und Regenwasser auch 300 Meter nach der Einleitungsquelle detektiert werden. In einer groß angelegten Messkampagne lag der Fokus auf der Charakterisierung und Quantifizierung von Huminstoffen im Grundwasser. Ein PARAFAC Modell bestehend aus drei Komponenten (c1–c3) erwies sich dafür als geeignet.

Zusammenfassend liefert die PARAFAC-Fluoreszenzspektroskopie in Verbindung mit der Durchflusszytometrie ein schnelles und leistungsfähiges System zur umfassenden Charakterisierung von DOM und Bakterien sowohl im Trinkwasser als auch im Grundwasser. Die kontinuierliche und automatische Überwachung der hervorgehobenen Parameter ermöglicht es, selbst geringfügige Abweichungen in der Wasserqualität zu detektieren.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

AOC	assimilable organic carbon
ATP	adenosine triphosphate
bdl	below detectable level
BDOC	biodegradable carbon
CDOM	chromophoric/colored dissolved organic matter
cfu	colony-forming unit
DMSO	dimethyl sulfoxide
DBP	disinfection byproduct
DI	deionized
DOC	dissolved organic carbon
DOM	dissolved organic matter
drEEM	decomposition routines for emission-excitation matrices
dw	drinking water
dw_{wds}	drinking water from various location within the network
dw_{wwe}	drinking water from waterworks exit
E. coli	*Escherichia coli*
EEM	excitation-emission matrix
FA	fulvic acid
FDOM	fluorescent dissolved organic matter
FSC	forward scatter
gw	groundwater
HA	humic acid
HNA	high nucleic acid
HS	humic substances
HSNOM	Lake Hohlohsee surface water
ICC	intact cell count
IFE	inner-filtering effect
IHSS	International Humic Substances Society
IQR	interquartile range
LB	Luria-Bertani
LC	liquid chromatography
lmw	low molecular-weight

LNA	low nucleic acid
n/a	not available
NOM	natural organic matter
OCD	organic carbon detector
PARAFAC	parallel factor analysis
PDOC	photochemically degradable carbon
POC	particulate organic carbon
Q1	first quartile
Q3	third quartile
Re	Reynolds number
rw	rainwater
SAC	spectral absorbance coefficient
SDG	Sustainable development goal
SG	SYBR Green I
SMP	soluble microbial product
SRFA	Suwannee River fulvic acid
SRHA	Suwannee River humic acid
SRNOM	Suwannee River natural organic matter
SSC	sideward scatter
SUVA	specific ultraviolet absorbance
TCC	total cell count
TOC	total organic carbon
TrinkwV	Trinkwasserverordnung
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
UVD	ultraviolet detector
WDS	water distribution system
WHO	World Health Organization
wwe	wastewater effluent

1. Introduction

Ensuring the access to safe high-quality drinking water is one of the global sustainable development goals (SDGs) that requires both appropriate treatment of raw water sources and secured distribution to end consumers. In addition to the scarcity of source water, its water quality is affected by a number of factors, e.g., anthropogenic impact from environmental pollution and population growth or climate change (Anderson et al., 2023; DVGW, 2022; IPCC, 2023).

Although drinking water quality in Northern Germany is high due to proper source management and production from deep, well protected groundwater, it is presumed that climate change will significantly increase the seasonal uncertainties of water supply to all regions of the global North (Leveque et al., 2021). Droughts and heavy rainfall are just two examples of how extreme weather events pressure the quality of surface water and of surface water impaired groundwater (Barbieri et al., 2023). Due to this, direct impact on the drinking water quality may occur depending on the water treatment as well as the water distribution system (WDS). Nevertheless, not only the effects of climate change need to be considered, but other factors must also be taken into account. One such factor is the ageing water supply infrastructure, which can lead to water loss and act as an external influence on drinking water quality (Folkman, 2018).

It is often the case that significant alterations to drinking water quality, which may affect public health, are attributable to bacterial contamination of the WDS (WHO, 2014). However, conventional plating methods for determining the bacterial constitution of drinking water are generally time-consuming with an analysis time up to approximately two days (DIN EN ISO 6222, 1999; DIN EN ISO 9308-1, 2017). Furthermore, the size of an urban WDS, such as Hamburg with approximately 5500 km of pipe infrastructure requires an enormous labor input for regular surveillance of the water quality all over the WDS. Hence a continuous and automated system that monitors drinking water parameters with regard to microbial water quality is desirable.

Fluorescence spectroscopy and flow cytometry are two known methods proven for their relatively rapid (< 15 min) and consistent analysis of water characteristics as well as for their robustness when used in the field (W. Chen & Yu, 2021; Safford & Bischel, 2019). Thereby, fluorescence spectroscopy with coupled parallel factor analysis (PARAFAC) characterizes and quantifies the dissolved organic matter (DOM) reflecting the total organic carbon (TOC) content in water, while flow cytometry counts and characterizes present bacteria in water. However, both technologies are not established as general drinking water quality surveillance techniques.

To address this issue, both methods were evaluated and further developed in the present study to enhance a detailed characterization of drinking water. Specifically, the focus was on sophisticated analysis of multivariate data and automation of both methods. Through the generation of near real-time fingerprints of drinking water using PARAFAC fluorescence spectroscopy and flow cytometry, the possibility of detecting even minor deviations in drinking water quality,

in terms of DOM and present bacteria, was investigated. This work's specific aims can be grouped into the following major research areas:

- **Further development** of PARAFAC fluorescence spectroscopy and flow cytometry. How powerful are PARAFAC fluorescence spectroscopy and flow cytometry, applied solely and in a combined approach, for rapid detection of DOM and bacterial changes in drinking water? What are the sensitivities of monitoring parameters regarding drinking water quality changes and contamination events?
- **Automatization** of PARAFAC fluorescence spectroscopy and flow cytometry. What could an automated software and hardware setup look like? Developing a general approach to handling multivariate water quality data.
- **Transfer** to real-world scenarios. How could the developed software and hardware setup be integrated into a real-life drinking water system? In what ways could PARAFAC fluorescence spectroscopy be used to characterize and quantify groundwater DOM?

2. Theoretical Background

2.1. Drinking Water Quality

The minimum requirement for drinking water is that it be pure and safe to drink. This means the water should contain neither harmful substances nor pathogens. In Germany, the responsibility for producing and supplying this flawless water lies with the water supply company (BMG, 2021). In 2019, approximately 69 % of Germany's drinking water was generated from groundwater sources whereas the remaining amount comes from surface water, riverbank filtrate, and artificially recharged groundwater (BMG, 2021). In Hamburg, nearly 100 % of drinking water sources originate from groundwater (Statistisches Bundesamt, 2019). Groundwater is generally of high quality, as various substances such as particles, dissolved organic and inorganic substances, bacteria and trace substances are naturally removed as the water percolates through the soil layers (Keesstra et al., 2012).

Groundwater is often treated using aeration combined with rapid sand filtration (Favere, 2022). Aeration brings the water in contact with air which facilitates the precipitation of containing iron, manganese, and organic compounds, while the subsequent sand filter removes these substances as well as containing microorganisms (Otieno et al., 2012). The application of disinfection methods, such as chlorine disinfection (Cheswick et al., 2020), ultraviolet (UV) radiation (Guo et al., 2022) or ozonation (Kämmler, 2023) can enhance sophisticated microbial treatment. However, disadvantages include the explicit formation of disinfection-by-products (DNPs) during chlorination (Ding et al., 2019), which is why it can be dispensed with in Germany if groundwater quality is sufficient (BMG, 2021).

In Germany, the Drinking Water Ordinance ("Trinkwasserverordnung", TrinkwV) specifies the quality measurements to be carried out and the respective limit values of drinking water (TrinkwV, 2023). Water suppliers are responsible for ensuring that the water is of sufficient high-quality, from treatment to the water distribution system (WDS), to the house connection. The frequency of drinking water quality tests varies according to the size of a WDS (TrinkwV, 2023). For larger systems like Hamburg, water quality must be checked once a day at least. Even the designation of high-quality drinking water means a large number of microorganisms, organic and inorganic substances in water, which are not harmful for the end consumer per se.

To meet drinking water's general microbiological requirements, it must not contain *Escherichia coli* (*E. coli*) or intestinal enterococci. In addition, no *Pseudomonas aeruginosa* may be detected in the water for release into sealed containers. It is also necessary to determine the following indicator parameters. Within the WDS, the colony count at 22°C and 36°C must not show abnormal deviations. Furthermore, there are strict limit values for various locations in the WDS, e.g., sealed containers and tapping points, between 20 and 1000 colony-forming units (cfu) per mL. Additionally, there are special indicator parameters for drinking water installations (*Legionella spec.*: 100 per 100 mL) and for the occurrence of certain microbial hazards (somatic

coliphages: 50 plaque-forming units per 100 mL of raw water). Concerning further organic carbon related parameters, the “Trinkwasserverordnung” states two general indicator parameters for drinking water. The total organic carbon content must not show any unusual deviations while the coloring, determined by the spectral absorption coefficient at 436 nm, must be below 0.5 m^{-1} all over the WDS. (TrinkwV, 2023)

Although approximately 99 % of the microbiological indicator-parameter tests showed negative responds in the years 2017 to 2019, continuous monitoring is essential since even the little occurrence of microorganisms requires rapid examination of the causes to prevent harmful impacts on end consumers (BMG, 2021). In this context, continuous monitoring of bacterial and organic characteristics of drinking water would allow for the detection of sudden changes, which could negatively impact the consumer's health. According to (Van der Kooij & Van der Wielen, 2013), the reasons for sudden water quality changes can be an indication of failing filtration and/or treatment which allows more bacteria and nutrients to enter the WDS or of failing disinfection or loss of disinfectant residual in the WDS, respectively. Moreover, it can be caused by external contamination such as a pipe break or other events (Moreira & Bondelind, 2017). Scenarios that favor this abrupt change in water quality within the WDS, now and in the future, could be the following:

- Climate change – droughts, floods, and heavy rainfalls
- Agricultural runoffs
- WDS infrastructure
- Human failure or cyberattacks

Climate change and its consequences have a significant effect on ground- and thus drinking water quality due to the increasing appearance of droughts, floods and heavy rainfalls (Taylor et al., 2013). The long-term effects of droughts on ground water quality by reducing aquifer rechargeability and increasing anthropogenic pressures are discussed in more detail by Barbieri et al. (2023) and Schreiner-McGraw and Ajami (2021). Regarding flooding events, it was shown that prolonged floodwater can induce several microbiological contaminants in ground water aquifers which directly affect drinking water quality, even far from the actual event (Masciopinto et al., 2019). Heavy rains can stress the entire treatment and distribution system and furthermore can result to an increase in the normal groundwater level. This can restrict the soil filtration capacity and thus the natural treatment of water (Tornevi, 2015). Moreover, Tornevi (2015) reported that heavy rain falls often lead to water tanks overflows, which could release pathogens. In addition, contaminants are generally mobilized through an impervious surface and reach undesirable locations. General climate warming may lead to an increase in water temperature, including in the WDS, which certainly affects microbial communities and, e.g., their growing behavior (Yuan et al., 2021). Chemicals exposed by agricultural runoff can also negatively impact groundwater quality (Hallberg, 1987). In this context, the increased input of

phosphorous-based fertilizers is mentioned in particular, as phosphorous is considered a limiting nutrient for bacteria in drinking water (Vrouwenvelder et al., 2010).

Ageing WDS infrastructure and associated pipe bursts not only lead to water loss, such as up to 10 % of the total pumping volume in some of the areas investigated in the USA and Canada (Folkman, 2018). Also the pressure loss that occurs in the leakage area can lead to groundwater, rainwater, surface water or even sewage from surrounding pipes penetrating the WDS (LeChevallier et al., 2003). In urban areas in particular, advanced technologies have significantly improved the control of water systems. However, remote control devices also introduce new risks such as human failures and cyberattacks. According to the US Department of Homeland Security 25 water system related cyber-attacks were reported in 2015 (Clark et al., 2017). A general overview of publicized cyberattacks in the water sector between 2000 and 2020 is compiled by (Tuptuk et al., 2021).

With all these challenges in mind, in order to detect evolving drinking water quality changes immediately and react appropriately, detailed and strict monitoring of organic matter and bacteria is required during the treatment process and within the WDS.

2.2. Organic Matter in Water

Organic matter in water can be subdivided into different groups which are introduced in the following part. An overview of all organic matter groups in water that are relevant to this work is provided in Figure 2-1.

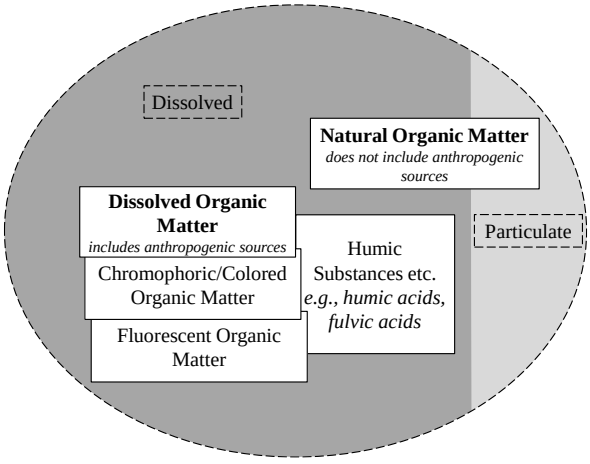

Figure 2-1: Overview of organic matter groups in water with relevance to this work.

Natural organic matter (NOM) refers to a complex mixture of organic compounds, such as plant and microbial residues, found in soil, water, and sediments, contributing to the dynamic cycling of carbon in ecosystems and can be present either in dissolved or particulate form (Sillanpää & Park, 2023). NOM is derived in water through the decomposition of terrestrial plants or as a by-product of bacteria, algae, and aquatic plants either in allochthonous or autochthonous processes (Filella, 2009).

In the literature, the term dissolved organic matter (DOM) defines the fraction of organic matter that passes through a filter, typically between 0.2 and 0.7 µm (Hansell & Carlson, 2014). However, contrary to NOM the term DOM includes anthropogenic sources of organic matter. There are two subfractions of DOM that play a crucial role in water analytics. Chromophoric (Tietjen et al., 2005) or colored (Gao & Zepp, 1998) dissolved organic matter (CDOM) describes organic matter capable of absorbing light. Fluorescent dissolved organic matter (FDOM) represents organic matter with intrinsic fluorescent properties (Carstea et al., 2020).

The complex mixed structure of DOM consists of varying chemical composition, molecular size, and molecular weight with different combinations of functional groups such as esters, phenols, quinine, carboxylic, hydroxyl, amino, and nitroso, which are all usually negatively charged at neutral pH (Gjessing, 1976). Furthermore, DOM may show regional and seasonal variations, e.g., depending on pH values and biological processes in a specific region and season (Leenheer & Croué, 2003). Since a detailed determination of particular substances is challenging due to DOM's complexity, there is a general characterization into the following hydrophobic and hydrophilic component groups which are present in natural waters: humic substances, polysaccharides, proteins, biopolymers, and low-molecular dissolved substances like organic acids and neutrals (alcohols, aldehydes, amino acids etc.) (Filella, 2009; S. A. Huber et al., 2011). According to Thurman (1985) humic substances are the dominant fraction of DOM in natural waters with approximately 50 % of the total organic matter content. Humic substances are typically high molecular weight molecules which cannot be described in a standardized chemical structure. However, Thurman (1985) and Sillanpää and Park (2023) describe the elemental composition amount (w%) as: C > O >> H > N > S with the following proportions of functional groups (w%): carboxyl > phenol and alcohol > carbonyl > methoxy. Humic substances can further be distinguished into three types of hydrophobic acids (Sillanpää & Park, 2023):

- Fulvic acids (FAs), which are soluble in alkali and acid and make up the majority of humic substances in natural waters (Thurman et al., 1982).
- Humic acids (HAs), which are soluble in alkali, but insoluble in acid (pH < 2) and are more aromatic than FAs, which tend to result in its more hydrophobic characteristic (Islam et al., 2020; Schnitzer et al., 1991). HAs are generally larger and have a higher molar weight as well as contain more carbon than FA (S. A. Huber et al., 2011; Thurman et al., 1982).
- Humins, which are insoluble in alkali and acid.

2.2.1. Analytical Methods to Characterize Organic Matter in Water

This chapter provides an introduction to established analytical methods for measuring and characterizing organic matter in water, in particular DOM, CDOM, and biodegradable fractions of

DOM. A more detailed introduction regarding the characterization of FDOM in water is provided in chapters 2.2.2 and 2.2.3.

In general, the content of organic matter in water can be described through the amount of total organic carbon (TOC) which is a measure of all organic carbon present in water. This usually reveals a result of carbon concentration in mg L^{-1}. TOC is furthermore subdivided into dissolved organic carbon (DOC) and particulate organic carbon (POC). Thereby, DOC describes organic carbon components which pass a filter size of 0.45 µm, while remaining components are referred to as POC (Danielsson, 1982). Typical DOC ranges for different waters are presented in Figure 2-2.

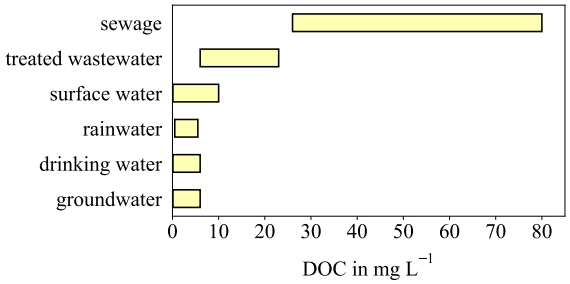

Figure 2-2: Approximate range of dissolved organic carbon (DOC) in different waters. Adapted and revised from Escalas et al. (2003), Grohmann et al. (2011), Kämmler (2023), Kördel et al. (1997), Li (2017), Torres et al.(2014) and Volk et al. (2002).

Regarding a more sophisticated characterization of DOM in water, size exclusion chromatography, more precisely liquid chromatography (LC) in combination with an organic carbon detector (OCD) is known as a method to separate the different DOM fractions depending on their different sizes and chemical functions (S. A. Huber et al., 2011). The LC-OCD analysis creates characteristic peaks over the analysis duration. The following DOM fractions can be distinguished and quantified based on the peak's amplitude and retention time: biopolymers, humic substances (HA, and FA), building blocks, low molecular-weight organic acids, and neutral compounds. Figure 2-3 shows an example of an LC-OCD diagram with typical peaks expected when analyzing natural waters. An estimate of the concentration of each fraction can be calculated according to the areas under the respective peaks.

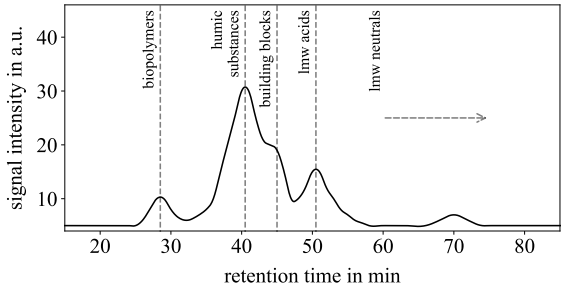

Figure 2-3: Example of an LC-OCD diagram indicating the intensity (concentrations) of the respective fractions according to S. A. Huber et al. (2011): biopolymers, humic substances, building blocks, low molecular-weight acids, and neutrals.

Regarding chromophoric or colored compounds in water (CDOM), their ability to absorb light in the ultraviolet (UV) and visible range is used to determine the CDOM concentration semi-quantitatively. The relation between absorption and CDOM concentration can be basically described via Beer-Lambert law (Equation 2-1), where A is the absorbance, ε is the molar absorbance coefficient in $mol^{-1}cm^{-1}$, c is the molar concentration in mol, and l is the optical path length in cm (Lakowicz, 2006).

$$A = \varepsilon c l \qquad \text{Equation 2-1}$$

A way to express the absorbance of a water sample is through determining the spectral absorbance coefficient (SAC) which corresponds to the attenuation of light of a certain wavelength over a distance of one meter. According to Sillanpää and Park (2023), any wavelength between 220 and 280 nm can be used for CDOM measurement, although different wavelengths are discussed to characterize different chromophores. For example, measuring the absorbance at 254 nm describes the content of aromatic groups and can therefore be used as a surrogate DOC measurement (Korshin et al., 2009). However, absorbance behavior in general is not dependent on the CDOM concentration only but also on the type of DOM (Weishaar et al., 2003) or other interfering factors such as high $NO3^-$ concentrations (Leenheer & Croué, 2003).

A parameter describing the biodegradable part of DOM in water is the determination of assimilable organic carbon (AOC). AOC characterizes the part of DOM which is responsible for bacterial growth in water (Huck, 1990). Van der Kooij (1990) described the AOC as accounting for very low amounts of total DOM in water. The initial approach to assessing AOC in water involves the use of an inoculum which can consist of either a blend of pure bacterial strains cultivated under controlled laboratory conditions or a mixture of environmental bacteria demonstrating broad nutritional adaptability (Van der Kooij, 1992). The related bacterial growth in water samples is tracked through colony counts. The mean growth observed during incubation is translated into AOC units in $\mu g\ L^{-1}$ as acetate-carbon. This conversion is accomplished by employing bacterial growth yields obtained from calibration curves, which are generated using standard concentrations of organic compounds like acetate or oxalate (Escobar & Randall, 2001). Hammes and Egli (2005) invented a comparatively new method of determining AOC. Without an inoculum, flow cytometry is applied to observe cell growth, which allows detailed monitoring of cell growth kinetics over time. According to Hammes et al. (2006), natural cell growth is calculated by using a conversion factor of 1×10^7 per µg of AOC. Low AOC concentrations have been shown to limit bacterial growth in water. Literature recommends to consider AOC values between 10 and 100 $\mu g\ L^{-1}$ as biostable waters (Favere, 2022; LeChevallier et al., 1996; Van der Kooij, 1992).

Apart from the introduced techniques, various sophisticated DOM characterization methods are available which can be summarized under the following approaches: biological tests, isolation and concentration, fractionation, spectroscopic methods, and chromatographic methods (W. Chen & Yu, 2021; Leenheer & Croué, 2003; Matilainen et al., 2011; Sillanpää & Park, 2023).

2.2.2. Fluorescent Dissolved Organic Matter in Water

The following sections describe the properties and challenges of intrinsic fluorescence, which is used to characterize FDOM in water.

Principle of Fluorescence

Fluorescence is one part of an optical phenomenon called luminescence. It is generally the emission of light from any substance from electronically excited states within nanoseconds (Lakowicz, 2006). Fluorescence can be explained by the Jablonski diagram shown in Figure 2-4.

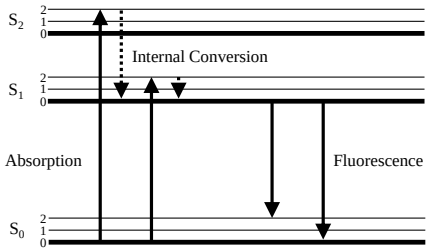

Figure 2-4: A form of the Jablonski diagram showing the fluorescence mechanism. Adapted and revised from Lakowicz (2006).

In the ground state (S_0) both valence electrons of the conjugated or delocalized π-electron system are in singlet state which means they have an opposite spin. During absorption, the incident light excites one of these electrons, without changing the spin, to a higher vibrational level (S_1, S_2, ..., S_n) energetically equal or higher than S_1. S_1 is defined as the first excited singlet state. Afterwards the electron usually rapidly relaxes to the lowest vibrational level S_1. This step is called internal conversion. During fluorescence, the system emits a photon and restores S_0 which is either reached at a minimal higher ground state level ($S_{0,1}$, $S_{0,2}$) or at the original ground state level ($S_{0,0}$). The system reaches equilibrium and is stable from that point on. There are various processes that compete with fluorescence, like intersystem crossing, quenching and photochemical decomposition. However, only quenching has a significant impact on fluorescence. Quenching means the radiationless transition of the excited electron to S_0 due to collisions with other molecules, e.g., dissolved oxygen. (Lakowicz, 2006; Wagner, 2014)

As shown in the Jablonski diagram, the absorption energy is higher than the energy emitted by fluorescence. Conversely, the wavelength of the excited light (λ_{ex}) is thus always higher than the wavelength of the emitted light (λ_{em}). This phenomenon was first discovered by Stokes and is generally called the “Stokes shift” (Stokes, 1852).

Regarding the fluorescence measurement of a single compound or a mixture of compounds, the sample is analyzed over a spectrum of UV and visible light. For this purpose, the sample is excited step by step at different excitation wavelengths. At the same time, the intensity of the

corresponding emitted light is detected. Typically, the analysis result is visualized in wavelength specific excitation-emission spectra or excitation-emission matrix (EEM) as shown in Figure 2-5.

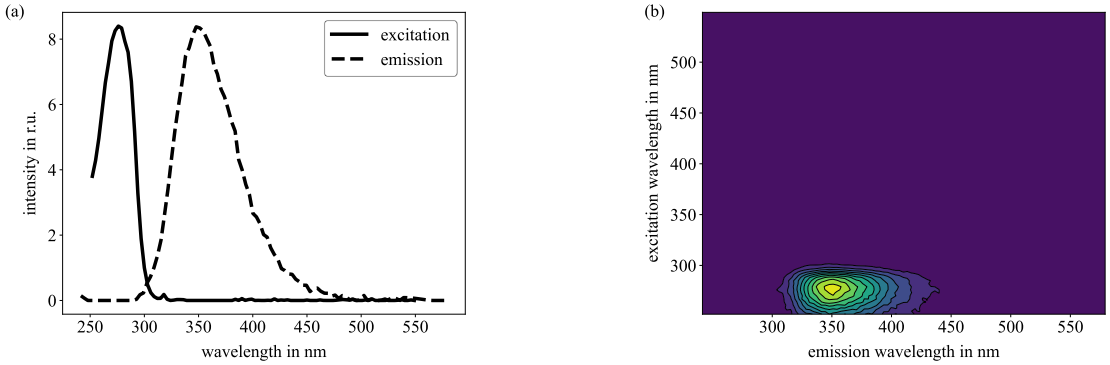

Figure 2-5: Fluorescence spectroscopy analysis of a single compound (here: tryptophan). (a) shows the excitation spectrum at λ_{em} = 350 nm, and the emission spectrum at λ_{ex} = 280 nm. (b) shows the excitation-emission matrix (EEM) over the entire analyzed UV and visible light range.

Fluorescence of Water

Fluorescence only occurs for aromatic compounds or compounds with conjugated double bonds (sp^2 hybridized carbon) while saturated bonds do not show fluorescence due to the required high energy demand and the localization of the electron bonds (Wagner, 2014). Aromatic, phenolic, and carbonyl-containing components are identified as the primary contributors to FDOM in water (McKay, 2020; Weishaar et al., 2003), while compounds like amino acids (e.g., tryptophan and tyrosine), polyphenols, and other organic acids (e.g., those containing functional groups like NH, NH2, and C=O) play a minor role (Baker, 2001; Maie et al., 2007; Wünsch et al., 2015).

Regarding the amount of FDOM on total DOM in natural water samples, there is uncertainty due to the individual composition of each sample. Leresche et al. (2022) described a modelling approach to investigate the mass ratio of CDOM to DOM of Suwannee River fulvic acid (SRFA), which is a commonly used DOM standard. They showed that approximately 50 % of carbon atoms are attached to a chromophore. This can be compared to the numbers from the International Humic Substances Society (IHSS) website (IHSS, 2024). As a comparison method to characterize DOM of river water samples, Stubbins et al. (2014) applied ultrahigh resolution mass spectrometry (FTICR-MS). They observed that up to approximately 60 % of nontarget DOM showed fluorescent properties (FDOM).

The characterization and quantification of FDOM in water through fluorescence spectroscopy and generated EEM was introduced by Coble et al. (1990). In this context, the generally applicable peak nomenclature was introduced and updated consecutively (Coble, 1996; Coble et al., 2014). Figure 2-6 shows a typical EEM of a drinking water sample including peak region letters nomenclature according to Carstea et al. (2020).

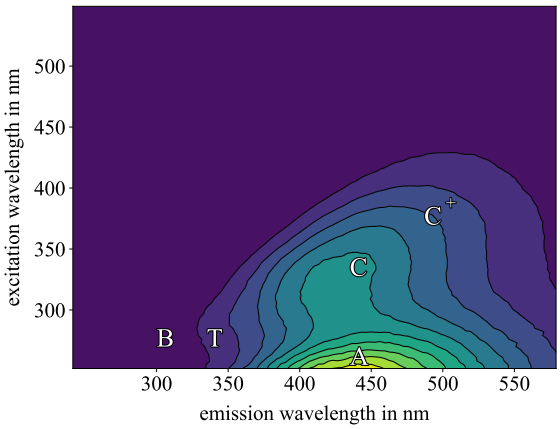

Figure 2-6: Typical drinking water excitation emission matrix (EEM) including peak positions from peak nomenclature.

EEMs are typically divided into five subregions: B, T, A, C, and C^+. Peak B ($\lambda_{ex}/\lambda_{em}$ = 230/305 and 275/305 nm) refers to autochthonous tyrosine and tyrosine-like compounds but also free and or combined amino acids with only one aromatic ring (Barsotti et al., 2016; Coble et al., 1998; Conmy et al., 2014; Parlanti et al., 2000; Stedmon & Markager, 2005). Peak T ($\lambda_{ex}/\lambda_{em}$ = 230/340 and 275/340 nm) is associated with autochthonous tryptophan-like substances with at least one aromatic ring like indoles, amino acids, etc. (Barsotti et al., 2016; Coble et al., 1998; Parlanti et al., 2000; Stedmon & Markager, 2005). Peak A ($\lambda_{ex}/\lambda_{em}$ = < 280/420–520 nm), C ($\lambda_{ex}/\lambda_{em}$ = 300–250/400–450 nm), and C^+ ($\lambda_{ex}/\lambda_{em}$ = 350–400 /475–525 nm) refer to the group of autochthonous and allochthonous humic-like substances with at least two aromatic rings (Barsotti et al., 2016; Coble et al., 2014).

Natural Effects Influencing the Fluorescence of Water

Environmental factors influencing fluorescence are pH, the solvent, metal ions, particles, bubbles, DOM concentration range, and temperature. Those effects can be detected by a deviation in fluorescence intensity or a shift in the emission spectrum. Carstea et al. (2020) and Conmy et al. (2014) reviewed all factors and their impacts.

A lower pH can cause decreasing fluorescence intensities since hydrogen ions can quench specific fluorophores (Lakowicz, 2006). However, Groeneveld et al. (2022) showed that the effect of pH on fluorescence is minor for pH range between 5.5 and 7.5. Metal ions cause quenching and thus reduce fluorescence intensity, while at the same time they form complexes with humic substances which in turn can lead to changes in fluorescence (Du et al., 2019; Wagner, 2014). According to Henderson et al. (2009), a temperature increase of 1°C can entail decreasing fluorescence intensity of ≈ 1 % within the range of 10–45°C, although there are several studies introducing temperature compensation factors (De Oliveira et al., 2018; Khamis et al., 2015; McKay et al., 2018). Non-organic particles can significantly affect fluorescence because of scattering and absorption effects which might result in attenuation and redirection of the light

(Hulst, 1981). There are studies presenting the correction of scattering effects, especially depending on turbidity thresholds (De Oliveira et al., 2018; Khamis et al., 2017).

The inner filtering effect (IFE) is the most significant effect affecting particle-free water fluorescence. It describes the attenuation of the fluorescence intensity due to light absorbance by organic matter before reaching the fluorescence detector and can be observed at rather high FDOM concentrations (Ohno, 2002). Primary IFE refers to the absorption of light before the fluorescence process (incident light), while secondary IFE describes the absorption of light after the fluorescence process (emission light) (Mobed et al., 1996). An established approach to compensating IFEs is an empirical correction based on the absorbance profile of the same sample (Lakowicz, 2006; Parker & Barnes, 1957). Rayleigh scattering is a form of elastic scattering in which incident light is scattered by small molecules ($< \lambda_{ex}$) contained in the water. The scattered light has either the same (1^{st} order) or twice (2^{nd} order) the wavelength of the incident light. Inelastic Raman scattering is caused by vibrational and rotational transitions in molecular atomic bonds. The scattered light appears red-shifted relative to the incident light due to energy transmission (1^{st} order) and two times the 1^{st} order scattered light wavelength (2^{nd} order). (Senesi, 1990). An example of the position of the Rayleigh- and Raman scattering bands is provided in appendix (Figure I-1).

2.2.3. Parallel Factor Analysis (PARAFAC)

Theoretical Background

Parallel factor analysis (PARAFAC) can be expressed as an unsupervised multi-way method working with an iterative algorithm developed by Harshman (1970), who based this idea on an earlier approach (Cattell, 1944). According to Bro (1997), PARAFAC can generally be described as an extension to the bilinear principal component analysis (PCA) model into a trilinear model with the mathematical expression of PARAFAC given in Equation 2-2. Since this was initially a general multi-way method to analyze three- or higher-order arrays, Bro (1997) examined the opportunity of applying PARAFAC modeling to fluorescence spectroscopy data sets in particular. Regarding fluorescence spectroscopic data, the three-dimensional character is achieved by the excitation-emission wavelengths and detected intensity of the emitted light. Hence, in the last decades PARAFAC found applications in diverse research fields such as describing crystallization and precipitation processes in biopharmaceutical modalities (Wegner & Hubbuch, 2022), quantifying metal complexes as catalyst materials (Ebrahimi et al., 2008), or the general application in food science (Lenhardt Acković et al., 2018). In the same way, it gained increasing application within the water sector for qualifying and quantifying FDOM explicitly (Fernández-Pascual et al., 2023; Ishii & Boyer, 2012; Sciscenko et al., 2022; Stedmon et al., 2003).

Murphy et al. (2013) published a tutorial fluorescence spectroscopic PARAFAC analysis based on MATLAB toolboxes which became the base for many subsequent research, especially

FDOM analysis in water. The tutorial works with two free MATLAB-implemented toolboxes called the N-way toolbox and the drEEM (decomposition routines for emission-excitation matrices) toolbox. While the N-way toolbox provides the PARAFAC engine and was initially developed by Andersson and Bro (2000), the drEEM toolbox, introduced by Murphy et al. (2013), supports the application, visualization, and interpretation of PARAFAC analysis. The drEEM toolbox is based on previously developed toolboxes DOMFluor (Stedmon & Bro, 2008) and FDOMcorr (Murphy et al., 2010, 2011), hence combining and extending their capabilities significantly. Since then, the drEEM toolbox has been updated several times to its current state (version 0.6.5). Noteworthy, using the drEEM toolbox in MATLAB requires a programming background, which is why studies were published introducing a user-friendly graphical MATLAB interface (Mazivila et al., 2018; Micó et al., 2019). Although applying PARAFAC in MATLAB environment using the mentioned toolboxes, other PARAFAC software are accessible, such as recently developed open source solutions staRdom package based on R (Pucher et al., 2019) or commercially available 'Solo' Software (Eigenvector Research, Inc., USA).

PARAFAC analysis is based on the mathematical expression of Equation 2-2. In this equation x_{ijk} represents the fluorescence data of a single sample i at an emission wavelength j and excitation wavelength k, respectively. f corresponds to one PARAFAC component. a_{if} represents the score of one component f and one sample i. b_{jf} stands for the value of an emission and c for the excitation wavelength (j and k) of the particular component f. e_{ijk} represents the residual which cannot be explained by the model.

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \qquad \text{(Equation 2-2)}$$

$$with\ i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K$$

An example of applying PARAFAC to a real water data set using Equation 2-2 is given in Figure 2-7. With a wide set of samples, each sample's EEM can be divided into, here, three different components. Therefore, the number of components and their position depend on the corresponding data set. PARAFAC modelling requires three key assumptions of the data set (Murphy et al., 2013):

- Variability: Each component must have individual intensity and excitation and emission spectra.
- Trilinearity: Neither do emission spectra vary over the excitation wavelengths nor do excitation spectra vary over the emission wavelengths. Fluorescence intensity increases linearly with concentration.
- Additivity: The total signal (EEM) results from the linear superposition of a fixed number of components.

The requirements of trilinearity and additivity can be directly related to Beer-Lambert law which implies that the extinction of light is directly proportional to the concentration of absorbing matter (Equation 2-1) (Lakowicz, 2006). However, the investigation of real water samples often leads to violations of the above-mentioned requirements.

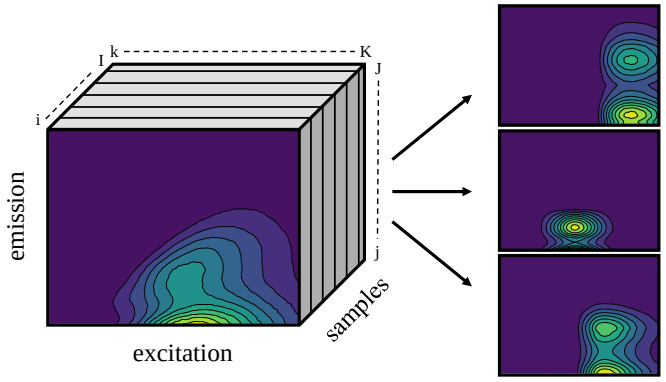

Figure 2-7: PARAFAC decomposition of an EEM data set into three fluorescent components. Adapted and revised from Murphy et al. (2013).

Applying the PARAFAC Model

The PARAFAC procedure is divided into five steps as shown in Figure 2-8. The procedure is generally valid for all types of analyses using the drEEM toolbox. Unless otherwise noted, PARAFAC modelling steps are described according to Murphy et al. (2013).

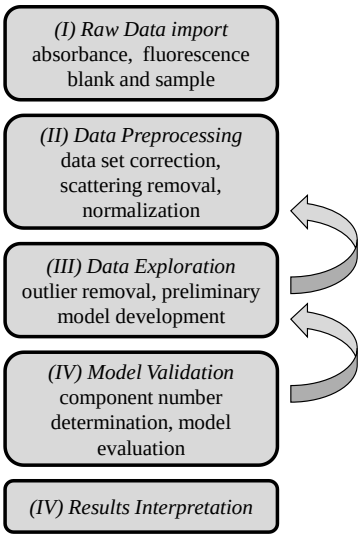

Figure 2-8: Schematic sequence of the PARAFAC working steps using the drEEM toolbox. Adapted and revised from Murphy et al. (2013).

(I) Raw Data Import

The raw data of one water sample consist of three files generated during one measurement: absorption of the sample, excitation-emission intensities of the sample, and excitation-emission intensities of a blank. In general, a blank represents the reading of an ultrapure water sample. A requirement for the appropriate PARAFAC procedure is the size of the data set (number of

samples). Sciscenko et al. (2022) recommended a data set containing more than 50 samples, although there are approaches to reduce this number significantly (Wünsch & Murphy, 2021).

(II) Data Preprocessing

Fluorometers can contain systematic biases in the dataset that must be corrected. This can be caused by imperfections in, e.g., the optical components or their alignment or variations in monochromator light transmission. Correction matrices can be generated and multiplied with the initial signal matrices to correct the dataset, either by integrated fluorometer programs or by manual post-treatment (e.g., drEEM toolbox).

As previously described, IFEs affect the linearity between fluorescence intensity and fluorophores concentration, especially in samples containing high DOM concentrations. A still-current method of correcting the IFE by utilizing the absorbance spectrum of a sample was proposed by Kubista et al. (1994). Based on the absorbance, a correction matrix is calculated for all pairs of excitation and emission wavelengths. However, the absorbance-based correction only works when the absorbance is below 2.0 in a 1 cm cell (Holland et al., 1977). This means that samples with higher absorbance should be diluted before being analyzed.

Another part of preprocessing is the elimination of non-trilinear data due to Rayleigh and Raman scattering. The typical treatment of these four scatter bands is either by replacing them with missing data (Bro, 1997) or with interpolated data from both sides of the bands (Bahram et al., 2006). The decision generally depends on the appearance of the investigated fluorophores. It is recommended that these data be replaced with missing data if the fluorophore does not appear in the region of e.g., 1^{st} order Rayleigh scattering or 2^{nd} order Rayleigh scattering(Andersen & Bro, 2003). Since the remaining scattering regions usually affect fluorophore signals directly, interpolation is suggested (Murphy et al., 2013).

The third step of normalizing the dataset is optional for PARAFAC modelling, but recommended for datasets having large concentration gradients. Normalization in relation to total signal intensity adjusts the weighting between higher and lower concentrated components. The decision about the need for normalizing a dataset can be made using the tutorial of Wünsch (2024).

(III) Data Exploration

PARAFAC analysis encompasses the careful selection of an optimal data set for modeling purposes. It also includes the initial estimation of the appropriate number of PARAFAC components for the model. At this step, outlier identification is crucial. Outliers refer to samples or wavelengths that exhibit significant deviations from the overall data set, potentially stemming from errors in sampling and analysis. It is imperative to individually verify all outliers to ascertain the reasons behind their deviations. Detection of outliers involves assessing samples/wavelengths based on their error residuals and departure from their average distribution (leverage).

Leverage is a number between zero and one. A leverage of zero for one sample/wavelength means that this sample/wavelength is not very different from others. In contrast a leverage of one means a very large deviation from the remaining data. Regarding further modelling, the leverage for all samples/wavelengths must be in a similar region. If not, the samples/wavelengths should be removed to avoid cohesive modelling errors. The determination of components and the identification of outliers (including leverage verification) are intertwined processes, constituting an iterative approach and is usually carried out by increasing the number of components from low to a supposedly correct number (Stedmon & Bro, 2008). The selection of too many components not only means that noise is increasingly modelled, but also that the true number of components is modelled by more correlated components (Bro, 1997).

(IV) Model Validation

After choosing a suitable number of components, four steps must be reviewed for successful model validation: randomness of the residuals, visualization of the spectral loadings, core consistency, and split-half analysis.

Residuals are signals remaining from the original EEM when removing all previous defined components. Should any samples show residual signals in the form of peak regions, this can be a sign of an inappropriate choice of number of components. This means the model is not yet sufficiently adapted (Stedmon & Bro, 2008). In the case that residual signals rather looking like background noise without any obvious peak regions, it can be concluded that the fit is sufficiently accurate. The next step can be performed.

Taking into account the "Stokes shift", the optical spectra of excitation and emission of each component must be examined in order to determine the physical reasonability (Lakowicz, 2006). In addition to the shift in the emission spectrum towards longer wavelengths, the spectra of the respective single component should fulfil the following characteristics, defined by Lakowicz (2006):

- There should be a minimal overlap between excitation and emission spectra (< 50 nm).
- The emission spectrum can exhibit only one specific peak, whereas the excitation spectrum can show multiple peaks. In water, FDOM usually has fewer than three peaks.
- In case of multiple excitation peaks, absorption generally occurs between these peaks.
- Excitation and emission spectral behavior should not show atypical and abrupt changes.

Bro and Kiers (2003) introduced the mathematical approach to core consistency diagnostic, which is applied during PARAFAC modelling for evaluating the appropriateness of the model. The core consistency is determined for each model with a different number of components, starting with a high number of components and then descending. The aim is to find the model with the number of components for which the core consistency approaches 100 % before it drops to the next model with a lower number of components. For real applications such as the

determination of FDOM components in water, this method is not always expedient. Models with low core consistency can also be appropriate, as components with low intensity (Bro & Kiers, 2003) or strongly covarying components (Murphy et al., 2008), for example, might affect core consistency. Part of this evaluation step is the finding of the so-called global minimum to ensure that there is no random solution of a PARAFAC model (Wünsch, 2024). For this purpose, a model with a certain number of components is calculated 100 times in the most ideal case and compared with each other. As long as the deviations from the individual calculations are small enough, the next step can be taken.

Split-half analysis is one of the most meaningful validation steps. The data set is split into at least two halves and each half are modelled independently. The spectral properties of the components derived from each half are compared and checked for variability. The model is considered robust if the components' properties of the two halves are identical. Successful split-half validation can fail if the data set is too small and contains very different samples (Stedmon & Bro, 2008).

(V) Results Interpretation

Ideally, each PARAFAC component in the validated model represents an independent fluorophore or possibly fluorophore groups with very similar spectral properties. Finally, a fluorescence intensity for each component of each sample (F_{max}) is given. In the case of an artificial sample with known fluorophores, it is possible to draw conclusions from F_{max} to concentrations by adding known amounts of the substance. However, real water samples usually contain unknown DOM substances for which this conversion is not trivial. In this case, it is recommended to work with F_{max} in relation to the sum of all components $\sum F_{max}$ rather than with concentrations ($F_{max}/\sum F_{max}$). A helpful tool for categorizing the developed model and its components is *OpenFluor* introduced by Murphy et al. (2014b). On the open-source website, everyone can share their models and data about the investigated water samples. This facilitates the interpretation and contributes to a better understanding of the developed model. This is done by comparing individual PARAFAC components to various published data.

2.3. Bacteria in Water

2.3.1. General Characteristics of Cells

A cell is an independent compound separated from other cells by a membrane and/or a cell wall. From a chemical point of view, cells are macromolecules that contain four components: nucleic acids, lipids, proteins, and polysaccharides. All cells are characterized by an individual metabolism that enables them to absorb nutrients from the environment, convert them and excrete waste products. Nutrients provide the basic substances cells require to reproduce. The reproducing or increasing of cell numbers is achieved through the cell division and is generally defined as microbial growth. (Madigan et al., 2015)

During microbial growth, the cell grows to twice its original size before dividing. Under optimal laboratory conditions (temperature, nutrient concentration, etc.), this process, known as exponential growth, is repeated continuously. This phase can be described by Equation 2-3 and Equation 2-4 in which the instantaneous growth x (biomass, protein etc.) at the time t is described as a function of the growth rate μ. (Cypionka, 2010)

$$\frac{dx}{dt} = \mu x \qquad \text{(Equation 2-3)}$$

$$x_t = x_0 e^{\mu(t-t_0)} \qquad \text{(Equation 2-4)}$$

Under optimal conditions, exponential growth is reached after an accumulation phase (lag phase), in which cells are already metabolically active but have not divided yet. In the case of a growth limiting factor like a depletion of an essential nutrient, cell growth stops and the cells reach the stationary phase. It is called the decline phase when cells die due to, e.g., persistent lack of nutrients or a decline in temperature. (Madigan et al., 2015)

Cells are fundamentally differentiated between two groups: prokaryotic and eukaryotic cells. Prokaryotes (bacteria) describe cells without a distinct nucleus, including *Eubacteria* and *Archaea*. The general cell size of prokaryotes varies from 0.1–0.2 to 50 µm whereby common bacteria such as *Escherichia coli* (*E. coli*) have a size between 1 and 3 µm. Eukaryotes, whose cells have membrane-bound nuclei, are further classified into single-celled eukaryotes (protozoa), animals and plants (algae or fungi). Their size is relatively large compared to prokaryotes (2 to 200 µm). The small size of prokaryotes with the resulting larger ratio between surface area and volume generally favors faster metabolic and growth rates. Accordingly, prokaryotes can strongly influence their environment in a relatively short time. (Reineke & Schlömann, 2023)

2.3.2. Bacteria in Aquatic Systems

Natural waters and engineered waters such as drinking water are not sterile but contain a large number of microorganisms (Ho et al., 2020). According to Fenchel and Jørgensen (1977) and Grohmann et al. (2011), bacteria in particular play the dominant role of microorganisms in aquatic systems. A distinction between different waters can be made based on the total bacteria amount. An established method for indicating the concentration of bacteria in water is the determination of total cell count (TCC). TCC ranges for different waters are displayed in Figure 2-9. Typical bacteria concentrations in different waters are in the range between 10^3 and 10^5 cells mL^{-1} for groundwater and drinking water, 10^4 and 10^7 cells mL^{-1} for surface water and 10^6 and 10^9 cells mL^{-1} for treated wastewater effluent (Berney et al., 2008; Ho et al., 2020; Kötzsch et al., 2012; Kötzsch & Sinreich, 2014; Manti et al., 2008). The TCC of rainwater prior to contact with the surface can be estimated at 10^2 to 10^5 cells mL^{-1} (Guillemette et al., 2023). However, a correlation between the number of pathogens present in different waters and the TCC was not identified in general (Van Nevel et al., 2017b).

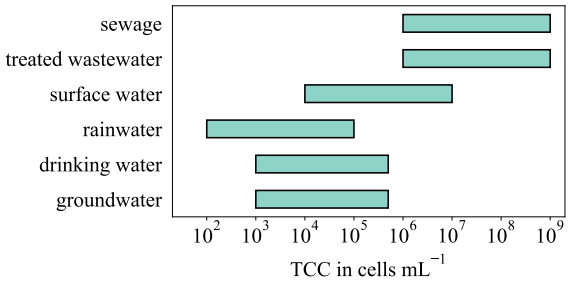

Figure 2-9: Typical ranges of total cell count (TCC) for different waters. Adapted and revised from Berney et al. (2008), Guillemette et al. (2023), Ho et al. (2020), Kötzsch et al. (2012); Kötzsch and Sinreich (2014) and Manti et al. (2008).

In particular, drinking water requires particular attention to pathogens since it is directly consumed by humans (WHO, 2022). There are two ways pathogens affect the WDS quality. The external entering into WDSs as external contaminants (e.g., human and animal) via insufficient treatment, leaking pipe systems etc. (chapter 2.1) or via bacterial growth in the system itself (WHO, 2014). Apart from the harmful impact of pathogens, all types of bacteria contribute to the water quality either by affecting odor and taste or leading to technical problems (Prest et al., 2016; Simpson, 2008). There are ways to remove (e.g., via membrane) or destruct (e.g., disinfection) bacteria. Nevertheless, there are usually remaining bacteria within the WDS, especially if it is not disinfected, usually occurs in Germany. Most of these bacteria can generally be considered harmless (Berney et al., 2008; Cabral, 2010; WHO, 2014).

Bacteria in WDSs generally appear dissolved in the bulk phase (free bacteria), or as a biofilm bound to the pipe walls, or attached to suspended solids or loose deposits, but all in dynamic exchange (Favere, 2022). Apart from the fact that free bacteria have a direct impact on consumers, attached bacteria, especially in biofilms, can have negative impacts such as an increase in flow resistance, material destruction or product quality. However, biofilm is also applied to the drinking water treatment, e.g., sand filtration (Maiyo et al., 2023). Considering bacterial growth within a WDS, bacteria in the biofilm require a lower concentration of nutrients (AOC) in the water stream than free bacteria, as nutrients can be accumulated in the biofilm (Grohmann et al., 2011).

Conventional plating methods such as heterotrophic plate counts (HPC) are commonly used to determine the total number of viable aerobic heterotrophic bacteria in water (Allen et al., 2004). However, there is a discrepancy between cultivable cell count and TCC, since only a few bacteria present in drinking water are reported as cultivable by HPC (Hoefel et al., 2003). According to Ferris et al. (2021) approximately 90 % of the total bacteria population is not culturable. Moreover, culture-based methods are limited due to their high labor and time requirements and are therefore not suitable for short-term water quality changes (Buysschaert et al., 2018; Ramírez-Castillo et al., 2015; Van Nevel et al., 2017b). This leads to the disadvantage of cultivation methods as continuous surveillance methods for drinking water. Other non-cultivation methods, e.g., polymerase chain reaction (PCR) and adenosine triphosphate (ATP) are available and described in detail by Botes et al. (2013) and Berney et al. (2008), respectively.

2.3.3. Flow Cytometry

Flow cytometry has been used in clinical hematology laboratories since 1990's as a non-cultivation method for single cell counting (Brown & Wittwer, 2000). The method distinguishes itself from other bacterial analysis by being fast (< 15 min) and sensitive (detection limit between 10^2 and 10^3 cells mL^{-1}). This method was also used as early as 2000 for assessing water quality in the workplace (Hammes & Egli, 2010; Safford & Bischel, 2019; Y. Wang et al., 2010). Recently, flow cytometry has become increasingly important as an online monitoring tool for drinking water surveillance (Favere et al., 2021; Gabrielli et al., 2021; Gunter et al., 2023; López-Gálvez et al., 2023).

Flow cytometry is a method for analyzing single events (particles, bacteria etc.) based on their scattering or fluorescent properties. Its general working principle is shown in Figure 2-10. The sensitivity limit for event sizes is usually 300 to 500 nm, but can be reduced to 100 to 150 nm under certain circumstances (dedicated devices, and trigger strategies) (Arkesteijn et al., 2020; Chandler et al., 2011; Van der Pol et al., 2014). Through hydrodynamic focusing, events in a fluid are aligned into a single stream and passed one after the other through at least one laser. Due to light scattering or fluorescence, the event affects the incident monochromatic light. The interaction between the event and the light is detected by 180° forward scatter (FSC), 90° sideward scatter (SSC) and fluorescence (FL) detectors. (Macey, 2007)

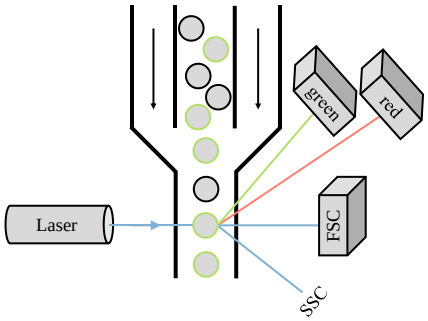

Figure 2-10: General working principle of flow cytometry.

The scatter detected via the FSC provides an indication of the surface, size or volume of the event (Bouvier et al., 2001), while the scatter obtained via the SSC is related to the complexity of the event, e.g., the granularity (Lewis et al., 2004). Various fluorescence detectors can be applied to detect light in a very narrow spectrum. Fluorescence can be caused either by fluid containing fluorophores or by adding fluorescent stain that is selectively adsorbive. Fluorescent dyes are commonly applied to differentiate between bacteria and abiotic events. Although there are several types of fluorescent dyes with similar properties, SYBR Green I (SG) is one of the most established nucleic acid dyes for aquatic analysis (Safford & Bischel, 2019). SG interacts with double-stranded DNA, more precisely with the nucleic acid of DNA, of both intact and damaged cells and shows a maximum excitation wavelength of 497 nm, whereas the fluorescence emitted light occurs at 520 nm (Dragan et al., 2012). Excitation and emission spectra of

SG are attached to the appendix (Figure I-2). Another commonly used fluorescent dye is propidium iodide. This is a charged molecule and therefore cannot penetrate intact cell membranes, allowing it to effectively discriminate between viable and dead cells (Safford & Bischel, 2019). Both SG and propidium iodide can be combined to obtain a summary of TCC and the relation between dead and viable cells in an aquatic sample.

Aquatic samples usually follow a staining procedure that is widely used and deviates only minimally. A crucial aspect of both fluorescent dyes' performance is the staining and incubation process. Ho et al. (2020) described the optimal conditions for water quality assessment as follows. Fluorescent dyes must be added to the sample and incubated at 37°C for 13 minutes. Tris(hydroxymethyl)aminoethane (Tris) was shown to be suitable as an appropriate solvent for dyes. Additional procedure recommendations are provided by Kötzsch et al. (2012), Nescerecka et al. (2016), Van Nevel et al. (2013) and Zunabovic-Pichler et al. (2018). After incubation the actual measurement starts by analyzing a specific volume of the sample.

2.3.4. Flow Cytometric Data Analysis

Flow cytometric data is challenged by the correct processing and interpretation of the generated multivariate data (Hammes & Egli, 2010). The term 'fingerprint' is often used to describe flow cytometry data sets. A fingerprint is an image or dataset that is unique for one samples, either from raw or further processed (Van Nevel, 2014). A way of further processing raw data is to visualize it via 1-D histogram plots or 2-D scatter or density plots (Prest et al., 2013). 2-D plots allow the combination of two different detectors' data. A common way is to choose two fluorescence detectors (red and green) for this comparison (e.g., chapter 3.2.2, Figure 3-1) (Prest et al., 2013). In 2-D density plots, each dot refers to a counted event. In the red-green plot, the dot's position contains information about its red and green fluorescence intensity.

Anyhow, the previously described fluorescent dye staining procedure leads to a shift in events containing DNA. Thus, bacteria and non-bacteria can be discriminated by the position of the occurring dot. Therefore, a widely used approach is to apply a polygon TCC gate, which defines all dots inside as bacteria (Ho et al., 2020). However, defining the boundaries of the TCC gate is subjective. Ho et al. (2020) provides a detailed description of an approach to standardizing the TCC gate process with Evian mineral water, which is briefly summarized here.

To remove background noises (bottom left signals) from the TCC gate, a sterile filtered (0.1 µm) sample is measured. All remaining signals belong to background noises whereby the vertical limit can be defined. Apart from the bottom left signals, water samples often show signals in the rather red fluorescence regions. These signals do not belong to bacteria-like materials and are classified as impurities like inorganic particles, humic substances, dust etc. (Ho et al., 2020; Van Nevel, 2014). Since fluorescent dyes, such as SG partly emit light in the visible red-light area (Figure I-2), not only a simple distribution over the horizontal axis, but also a

distribution over the vertical axis can be observed. According to Ho et al. (2020), an unfiltered Evian mineral water sample can be applied to describe the gradient of the TCC gate diagonal.

TCC also differentiates between bacteria with low and high nucleic content (LNA and HNA). (Y. Wang et al., 2009) approach this differentiation by counting cells as LNA cells if they pass a 0.45 µm filter. Another approach of Ho et al. (2020) is the application of Evian mineral water which usually contains two clear bacteria clusters (LNA and HNA), to define the vertical limit in between those clusters. In the context of this work the term 'fingerprint' is defined as the red-green fluorescence plot, including information about TCC, as well as HNA and LNA amounts.

2.4. State of the Art in Drinking Water Quality Monitoring

2.4.1. Monitoring via General Water Quality Parameters

There are parameters already established in online drinking water quality monitoring, such as pH, turbidity, conductivity, free chlorine, dissolved oxygen, and TOC (Banna et al., 2014). These parameters are relatively simple to monitor and integrate into a digital surveillance system, since they usually generate one value, which in most cases gives direct feedback about the sufficiency or insufficiency of water quality. According to "Trinkwasserverordnung" there are even strict limit values for conductivity (< 2790 µS cm^{-1} at 25°C), pH (6.5–9.5), and turbidity (nephelometric turbidity unit < 1.0) (TrinkwV, 2023). However, pH, conductivity, dissolved oxygen, and TOC are cumulative parameters, which means that a potential deviating value would not provide characteristic information (Frimmel & Abbt-Braun, 2011).

Concerning biological relevant drinking water quality parameters, such as *E. coli* and coliforms, Silva et al. (2022) emphasized novel approaches of an alternative monitoring which can be partly applied as automatic online methods. Tok et al. (2019) developed an online system that uses the absorption and fluorescence of water together with digitally processed images to determine the presence of bacteria. They detected *E. coli* and coliforms within approximately 16 hours down to a detection limit of 1 cfu per 100 mL. Zhang et al. (2018) developed an online approach using an automatic electrical bacterial growth sensor based on a multichannel capacitively coupled contactless conductivity detector. By qualitative and quantitative comparison of bacterial growth behavior over incubation time, viable bacteria concentrations can be estimated. Despite their automatic approach, both methods still require (incubation) time, and further testing in real environments. However, laboratory-based plate cultivation methods as described in chapter 2.3.2 are still the standard for analyzing bacteria and especially pathogens in drinking water (TrinkwV, 2023; WHO, 2022).

In the case of multivariate data generated by fluorescence spectroscopy and flow cytometry which are not considered with a fixed limit value in "Trinkwasserverordnung", the question arises as to the point at which drinking water quality can be considered inadequate. Nevertheless, there are several approaches to applying fluorescence spectroscopy and flow cytometry to drinking water quality monitoring. These approaches are described in the following chapter.

2.4.2. Monitoring via (PARAFAC) Fluorescence Spectroscopy

Fluorescence spectroscopic generated EEMs enables data evaluation either by peak-picking or PARAFAC analysis (chapters 2.2.2 and 2.2.3). For both, the results are available as fluorescence intensities, either of one peak region or of one particular fluorescent compound, which is difficult to convert into FDOM concentrations in case of investigating an unknown aquatic system. There are studies dealing with qualitative evaluation, but also quantitative evaluation in both online and offline operations.

Qualitative online investigations were performed by Carstea et al. (2018), Khamis et al. (2015) and Shutova et al. (2016). Carstea et al. (2018) installed three fluorometers to monitor the T and C peak regions of effluent organic matter at a wastewater treatment plant over one month. They identified daily variations in the DOM's composition which were related to minor changes in effluent quality and process performance issues. Khamis et al. (2015) focused on the T peak region when online monitoring of groundwater over 28 days to investigate the dependency between temperature, turbidity and fluorescence signals. Accordingly, they showed sufficient monitoring of groundwater fluorescence without any parameter compensation. Shutova et al. (2016) monitored the fluorescence of peak C and T over one month for the influent (groundwater) and effluent of a drinking water treatment plant. Peak C was observed to strongly correlate with DOC of both waters, whereas both peaks showed correspondence to general water quality fluctuations and operational conditions.

A deep interest gained in the monitoring of tryptophan-like fluorescence (T-peak) and its relation regarding microbial activities, both online and offline (Bedell et al., 2020; Dapkus et al., 2023; Nowicki et al., 2019; Sorensen et al., 2018; Ward et al., 2020, 2021). Bedell et al. (2020) focused on developing and testing a novel offline fluorescence sensor device measuring T-peak fluorescence to detect *E. coli* contamination in drinking water. They highlighted the correlation between T-peak intensity and *E. coli* in artificial waters (deionized water + *E. coli* strain) with a sensitivity limit of 4 cfu mL^{-1}. Dapkus et al. (2023) compared online monitored T-peak intensities with manual sampling from a karst basin, including collected information about air temperature, turbidity and antecedent precipitation. A strong correlation between *E. coli* and T-peak intensity was observed for *E. coli* concentrations > 1000 cfu per 100 mL. Nowicki et al. (2019) investigated offline the relation between *E. coli* and T-peak fluorescence of a groundwater stream and applied WHO risk indicator (No, low, medium, high, and very high risk) levels for both parameters (WHO, 2017). They recommended not using T-peak as an *E. coli* proxy on an individual sample basis, but as an additional valuable parameter regarding groundwater risk assessment. Ward et al. (2020) and Ward et al. (2021) presented data from a large-scale investigation of an offline groundwater and drinking water campaign. T-peak intensities were compared with *E. coli* presence, utilizing WHO risk classes (WHO, 2017). They highlighted that T-peak intensities may be an indicator for *E. coli*, although not as an individual parameter of bacterial contamination. They also emphasized the influence of C-peak on T-peak

which is affected significantly by varying DOC concentrations and concluded by T-peak intensity monitoring as a high-level screening tool for potential bacterial risks. Sorensen et al. (2018) investigated offline T-peak data from a wide set of studies investigating surface water, groundwater, drinking water applying the same WHO risk categories. Although a strong correlation was observed for higher contaminated waters, it was unsuccessful to monitor very low contaminated waters (< 10 cfu per 100 mL).

A selection of studies using the PARAFAC approach for monitoring DOM behavior in different waters are highlighted in the following section. These studies can be considered most recently published or studies with a high level of impact. Sciscenko et al. (2022) provided a detailed review. All of these studies applied offline PARAFAC modelling to test either the water treatment efficiency of DOM removal or the water quality within the WDS.

An offline campaign was conducted by Gabrielli et al. (2023), who investigated groundwater, treated drinking water and water from the WDS throughout one year. The waters were characterized through PARAFAC modelling and evaluated the effectiveness of the respective component removal depending on the type of water treatment (sand filter, granular activated carbon filter). The investigations confirmed results from Heibati (2019) and Heibati et al. (2017), who identified effects of deviating DOM in household taps which are eventually caused by plastic applications or biofilm in the tap. A study by Wells et al. (2022) examined the offline analysis of drinking water reuse in water treatment plants. They showed interactions between different components, especially within protein-like environments. This resulted with the recommendation of separating treatment processes according to the removal of the desired compound. A worldwide offline campaign of characterizing surface waters and drinking waters via PARAFAC modelling was conducted by Philibert et al. (2022). In combination with LC-OCD analysis, they emphasized the monitoring of component ratios of drinking water before and after treatment to recognize crucial changes in water quality. This confirmed the work of Shutova et al. (2014), who also highlighted the significant correlation between monitoring component ratios and treatability, especially coagulation-sedimentation, and coagulation membrane filtration. Regarding drinking water treatment via biofiltration, Moona et al. (2021) investigated the relation between empty bed contact time and DOM removal. The authors showed appropriate monitoring of DOM using PARAFAC modelling. They highlighted the increasing removal of even protein-like compounds which are usually difficult to remove apart from biofilters.

Besides the mentioned studies focusing mainly on DOM removal during water treatment, there are two publications aiming at different objectives. Fernández-Pascual et al. (2023) reviewed the formation potential of disinfection byproducts (DBPs) with regard to PARAFAC components. They showed a significant relationship between carbonaceous DBPs formation and the presence of ubiquitous humic- or fulvic-like compounds. Moreover, they mentioned cases showing correlations between protein-like components and nitrogenous DBPs. An approach to applying PARAFAC modelling to detect contamination events was introduced by Stedmon et

al. (2011). During laboratory experiments, they found one humic-like compound and one protein-like compound that could detect induced wastewater contamination. They showed the potential of using PARAFAC fluorescence spectroscopy not only to characterize DOM removal within treatment plants, but also as an appropriate tool to assess contamination events.

2.4.3. Monitoring via Flow Cytometry

Flow cytometry as and offline and online method is commonly used to characterize bacterial communities in various aquatic environments. The following section highlights articles related to the present work. A detailed review on flow cytometry applications is provided by Safford and Bischel (2019) and Schönher et al. (2021).

Van Nevel et al. (2017a) emphasized flow cytometry as a method of bacterial stability surveillance after pipe maintenance. They evaluated offline flow cytometry and fecal indicator plating. By appropriate pipe flushing TCC of maintenance pipes reached the same TCC as reference water, and the bacterial stability of the pipes was confirmed hours later by plate counting methods. Flow cytometry instead of plate counting saved time and water. Schleich et al. (2019) characterized (TCC, HNA, LNA, and dead/alive) bacterial communities within a WDS over one year. This revealed seasonal dependencies, including an increase in TCC at higher temperatures. The investigation of different locations within the WDS obtained a relation between higher retention time and decreasing HNA amount as well as increasing TCC. This relation was confirmed by Schönher et al. (2021).

Since Hammes et al. (2012) realized the possibility of an automated online flow cytometry system, work focus has shifted towards monitoring water quality online. As early-stage research into online monitoring Besmer et al. (2014) showed the feasibility of its application in tap water surveillance by analyzing TCC, HNA, and LNA every 15 minutes up to two weeks. Thus, daily pattern, such as daytime and nighttime were observed. Buysschaert et al. (2018) further tested online flow cytometry as an early warning system within a full-scale water treatment plant. They proved that flow cytometry recorded water quality changes during the treatment process which were not recorded with conventional methods. Favere et al. (2020) continuously monitored the water quality of inflowing and outflowing water from a WDS water tower. Automatic sampling was conducted for several weeks every 40 minutes. Using TCC, operational events such as changes in drinking water sources, e.g., surface water or groundwater, were detected. Besides, changes in the biological stability of the water within the tower were recognized. A similar approach was conducted by Gabrielli et al. (2021), who compared hourly flow cytometry data with daily conventional plate counting approaches. They demonstrated that hourly differences could not be determined using conventional methods due to their lower measurement frequency. Moreover, flow cytometry data showed some correlation with chemical-physical components, such as pH and conductivity.

Literature addressing flow cytometry as an event or contamination detection method is addressed in the following section.

Van Nevel et al. (2017b) investigated general applicable TCC limits for drinking water. Although TCC is affected by several factors such as the type of the source water, the treatment process, and the environmental conditions, they showed that over 95 % of investigated drinking water samples contained less than 5×10^5 cells mL^{-1}. Due to this fact, they recommended this as an alarming limit value for drinking water. Besides, they recommended to rather focus on abnormal changes of fingerprints requiring the establishment of individual water baselines.

More specific online monitoring approaches to contamination detection were conducted by Besmer et al. (2017), Favere et al. (2021), Prest et al. (2013) and Sadler et al. (2020). Prest et al. (2013) tested flow cytometry for its feasibility to assess crucial changes in TCC, HNA, and LNA when mixing drinking water with wastewater or different drinking water. When spiking with drinking water, a volume ratio of < 2 v% of wastewater was detectable. Also, mixing of both drinking waters was detected, although at higher volume ratios. Besmer et al. (2017) determined a wastewater contamination in a through-flow laboratory setup with an equivalent volume ratio of 0.6 v% through TCC, HNA, LNA, and dead/alive discrimination. Sadler et al. (2020) investigated a stirred vessel containing chlorinated drinking water spiked several times with wastewater (< 3 v%). The spiking events were monitored via flow cytometry. Favere et al. (2021) measuring TCC within a flow-through pilot plant fed with water from a treatment plant. Additional to flow cytometry, other bacterial monitoring devices were applied and compared regarding their detection limits. Water quality changes were caused by operational changes, e.g., backwashing, and simulated contamination cases. For the latter, rainwater and groundwater were injected into the system and detected down to 0.1 v%.

2.4.4. Monitoring via Combined (PARAFAC) Fluorescence Spectroscopy and Flow Cytometry

The literature review revealed only two studies in which both methods were applied. One of these was published as part of this work (Schuster et al., 2022a). In the study, an automated online monitoring system was developed in order to determine simultaneously the composition of artificially adjusted waters based on the presence of bacteria and DOM. Thereby, a protein-like PARAFAC component could monitor the TCC. Vucinic et al. (2023) combined both methods offline, excluding PARAFAC analysis, to control karst spring waters, which can then be used for drinking water supply. They highlighted the ability of using both methods to monitor bacterial changes.

2.4.5. Event Detection in Flow Water

Considering drinking water quality monitoring, examples of specific events or contamination detection were discussed in the previous chapter. A more general definition of real-time event

detection for water quality monitoring is given by Zhao et al. (2014). According to them, the measuring points of a continuous monitoring setup are classified into four groups:

- Outlier: A data point in a series of measurements that does not match the previous behavior.
- Event: A series of significantly deviating data points likely due to a contamination event.
- Baseline: Describes water quality in its normal state and is generally stable.
- Baseline change: Describes a persistent change in the mean value of water quality. Baseline changes are usually associated with process changes.

Dynamic and static monitoring approaches are distinguished. Dynamic models are characterized by a sliding threshold interval caused by, e.g., a baseline change. The threshold interval is calculated as a function of the standard deviation of the sliding window. It also uses and additionally can use positive and negative filters to distinguish potential outliers from detected events (Arad et al., 2013). However, static approaches are more convenient considering laboratory and pilot-scale setups since they are more suitable for experiments with limited experimental time and without expected baseline changes (Besmer et al., 2017; Favere et al., 2021). Favere et al. (2021) described the static threshold value for event detection as three times the standard deviation of the mean value (Equation 2-5). It is based on the normal distribution where values outside the threshold interval account for only 0.13 % of the total data points (Howell et al., 1998). Here, $x_{i,min/max}$ is the lower baseline limit (lower threshold) and the upper baseline limit (upper threshold), respectively, of parameter *i*. $\overline{x_i}$ is the average value of parameter *i*, while σ_i is the standard deviation of parameter *i*.

$$x_{i,min/max} = \overline{x_i} \pm 3 \times \sigma_i \qquad \text{(Equation 2-5)}$$

3. Materials and Methods

3.1. Water Samples

This chapter provides a general overview of the water samples that were examined in this work. All water samples originated from Northern Germany, the majority from the city region of Hamburg and were sampled between the years 2021 and 2023. These were primarily untreated groundwater (gw, n = 135) directly from the wells, treated drinking water both at the water-works exits (dw_{wwe}, n = 11) and at various locations in the WDS, e.g., hydrants and household connections (dw_{wds}, n = 24). Individual samples were taken at a sewage treatment facility, named wastewater effluent (wwe, n = 2), and rainwater (rw, n = 2) from roof top rainwater harvesting. Additionally, as part of a large-scale measurement campaign, ground water samples (gw, n = 135) were taken within the supply area from Hamburg.

All samples were taken from their respective locations bubble-free and stored in annealed narrow-neck glass bottles under anaerobic conditions. All samples were analyzed the same day or the day after. Until they were cooled in the fridge at 4–8°C.

Apart from natural water samples, some model solutions play a significant role in the present work. Concerning water quality monitoring experiments, some of the dw_{wds} samples were spiked with an organic nutrient medium. A detailed introduction to these samples is provided in chapter 3.3.2.

Furthermore, this work required the preparation and supply of several standard solutions in order to evaluate DOM characterization methods of investigated water samples. One part of those belong to the group of commercially available Suwannee River natural organic substances of the International Humic Substances Society (IHSS, USA). In particular, Suwannee River fulvic acid (SRFA, 2S101F), Suwannee River humic acid (SRHA, 2S101H), and Suwannee River NOM (SRNOM, 2R101N), which were ordered as solids via the IHSS website (IHSS, 2024). The surface water of Lake Hohlohsee (HSNOM, Germany) was used as another standard representing the full range of aquatic NOM. Further standards were applied to calibrate the fluorescence spectroscopy setup. The respective materials including supplementary information are summarized in the appendix (Table I-I). All standards that were present in particulate form were dissolved in ultrapure water from the Millipore Direct-Q 5 UV System (Merck, Germany). A detailed characterization of all used DOM standards was provided in recent studies (Benecke, 2018; Mantel, 2022; Schulz, 2020).

An explicitly naming requires a data set of samples which was called *base data* from here onwards. In the experimental course of this work, this data set was created to support PARAFAC fluorescence spectroscopy data analysis. The *base data* consisted of fluorescence data of two times 36 water samples (n_{total} = 72). These 36 samples mainly included various drinking water samples (dw_{wds}). Moreover, some drinking water samples of the *base data* were artificially

modified by spiking with slightly varying amounts of different water e.g., varying drinking water, rainwater, or wastewater effluent. The *base data* was not only used to characterize drinking water, but also to identify external influences on drinking water quality such as rainwater or wastewater infiltration. A list of the origin and mixing of each sample of the *base data* is provided in the appendix (Table I-II).

3.2. Analytical Methods

It was necessary to use a variety of chemicals and chemical solutions in order to perform all the analytical methods in this work. A list of the respective materials is provided in the appendix (Table I-III).

3.2.1. Fluorescence Spectroscopy – DOM Fingerprinting

Fluorescent properties of water samples were analyzed using the Aqualog fluorometer (HORIBA, Japan) with its combined Aqualog software version 4.2. Either a 10 mm quartz cuvette or a 10 mm quartz through-flow cell were used. The water samples were exposed to monochromatic light (150 W ozone-free xenon arc lamp) with excitation wavelengths (λ_{ex}) between 200 and 600 nm at 3 nm increments. The detected emission wavelengths (λ_{em}) spanned from 211.71 to 620.92 nm at 3.27 nm increments. The gain of the charged-coupled device detector was set at medium. The integration time was set at 1 s. Every pair of λ_{ex}, λ_{em} results in a corresponding fluorescence intensity in arbitrary units (a.u.). These intensities were stored in a so-called excitation-emission matrix (EEM).

All blank samples were obtained by analyzing ultrapure water samples, and were measured in the quartz cell every experimental day, respectively.

Fluorescence Spectroscopy – Automation Unit

As described in the following sections, specific hardware and software adjustments were required to enable continuous and automated analysis.

The hardware was adjusted as part of this work by adding the opportunity of automatic and subsequent transferring of up to three different samples to the flow-through cuvette of the Aqualog. This hardware part of the device is further called Aqualog-Sipper. Its flow scheme and a screenshot of its customizable setting options are provided in the appendix (Figure I-3 and Figure I-4). With Aqualog-Sipper integrated software, the flow could be controlled by adjusting the peristaltic pump performance, as well as the time to rinse the tubes and flow-through cuvettes for each analyzed sample. As soon as the rinsing time of each sample was reached, Aqualog-Sipper immediately closed all valves and the water present in the flow-through cuvette was ready to be measured. After the time set for the measuring process, the sample was automatically removed from the flow-through cuvette and drained into a waste container. The system cleaned according to the instructions using ultrapure water and/or sodium hypochlorite solution

(1 % active chlorine). The setup allowed automatic water filtration before the measurement, although it was not required for the water samples investigated in this work.

Fluorescence spectroscopy – Measurement

Actual measurements were conducted via the Aqualog 4.2 software integrated 'Sample Queue Method – Q'. The scan settings were set as previously described. All postprocessing steps of 'Sample Queue Method – Q' were deactivated to carry them out in the subsequent PARAFAC analysis. The combined sampling, measuring, and draining process required approximately 6 minutes to guarantee appropriate results.

Fluorescence Spectroscopy – Data Evaluation

Fluorescence spectroscopy raw data was processed using the PARAFAC multi-way decomposition method with the drEEM toolbox version 0.6.5 in MATLAB version R2021b (Murphy et al., 2013). The toolbox performed all needed steps to develop a multi-component FDOM model of the analyzed water samples which was previously explained in chapter 2.2.3 (*(I) Raw Data Import, (II) Data Preprocessing, (III) Data Exploration, (IV) Model Validation*). This process was performed according to the procedure explanation by Murphy et al. (2013) and Wünsch (2024). The mentioned literature contains further detailed explanation of all functions of the drEEM toolbox, thus here only the crucial steps are described. All water specific treatment steps were contributed and discussed in Eberhard (2022), J. Huber (2022), Schuster et al. (2022a) and Wullenweber (2022). The MATLAB code is provided in appendix (Figure I-5).

(I) Raw Data Import

Different sampling data had to be imported into the MATLAB script. For each sample, this implied importing the following three files: absorbance spectra sample ('- Abs Spectra Graphs.dat'), fluorescence spectra sample ('- Waterfall Plot Sample.dat'), and fluorescence spectra blank ('- Waterfall Plot Blank.dat'). (Code lines 9–22, Figure I-5)

(II) Data Preprocessing

By using Aqualog in combination with the Aqualog-Sipper, the data preprocessing focused on the removal of background noises, correction of IFEs, and Raman and Rayleigh scattering. Contrary to Murphy et al. (2013), normalization of the data set was carried out in step *(III)*. Particular parts of the EEM which either contained lots of background noises or was not expected to show any FDOM signals were removed. This concerned emission wavelengths $\lambda_{em} < 240$ nm and $\lambda_{em} > 580$ nm as well as excitation wavelengths $\lambda_{ex} < 250$ nm and $\lambda_{ex} > 550$ nm (code lines 90–91, Figure I-5). While IFEs were corrected by normalizing all EEMs by the Raman peak intensity of the blank at $\lambda_{ex} = 351$ nm, the area of Rayleigh and Raman scattering was determined manually (code lines 65–84, Figure I-5). Since there was no recognized scattering within the 2nd order Raman range, only 1st order Raman as well as 1st and 2nd order Rayleigh were defined. 1st and 2nd order Rayleigh were removed because of interfering

FDOM signals in these areas. It was different for the 1st order Raman area, where values were interpolated for further processing. (Code line 96, Figure I-5).

(III) Data Exploration and (IV) Model Validation

First, errors in all EEMs were automatically localized, removed and data in this range was interpolated (code line 105, Figure I-5). Then, an outlier test was performed to generate preliminary models with different numbers of components and to identify samples with an unusually high leverage (code line 112, Figure I-5). The correlation between each component was tested. In case of high correlation, the data set was normalized (code line 117, Figure I-5). Subsequently, the normalized data set was again tested for outliers and correlations. The spectra of all component models were compared to identify possible components that were only created due to background noise (code line 127, Figure I-5). The last step was to generate the sum of squared errors for each component model (code line 132, Figure I-5). In general, a lower sum of squared errors may indicate the model with the highest probability of having the right component number (Wünsch, 2024). Based on these preprocessing steps, a first idea about the component number could be estimated.

Afterwards, an initial appropriate model calculation of the data set was carried out by random initialization with the condition of non-negative values. This implied that the model scores and loading should be greater than zero. The model was calculated from 10 model calculations with a convergence criterion of 10^{-6} (code line 138, Figure I-5). All previously described steps of data exploration, excluding the outlier test, were performed to evaluate the model's appropriateness. Additionally, the particular core consistency of the different component models was checked (code line 151, Figure I-5). According to this evaluation, previously described steps of data exploration were repeated in an iterative process or a final component number was determined.

A split-half validation analysis was then carried out in the latter case to continue the overall process. Therefore, an overall PARAFAC model with a defined number of components was derived from the normalized data set (code line 165, Figure I-5). Afterwards, the normalized data set was randomly distributed into two splits and a model was generated for each split (code lines 171–172, Figure I-5). To conclude, spectra of each split and the overall model were compared (code line 173, Figure I-5). In case of the same spectral behavior of both splits, the overall model was considered as validated, renormalized and used for further processing. The gained model data was exported via Excel, containing information about each component's spectra and the sample specific fluorescence score in Raman units (r.u.) of each component (code line 184, Figure I-5).

3.2.2. Flow Cytometry – Bacterial Fingerprinting

The bacterial content of all water samples was analyzed using the CyFlow™ Cube 6 (Sysmex, Germany) flow cytometer. The flow cytometer contained a synthetic quartz flow cuvette with

a centric flow channel (capillary diameter: 350 µm × 200 µm). The laminar transport through the cuvette was ensured via the closed flow cytometer system with a pressure below 500 mbar containing the investigated sample covered by a so-called sheath fluid, here ultrapure water. The sample delivery was controlled by an integrated computer-controlled precision syringe pump with adjustable flow rates between 0.1 and 19.9 µL s^{-1}.

The flow cytometer was equipped with a 60 mW blue diode laser at a rigid wavelength of 488 nm. Four detectors were used in current work to detect emitting and scattering light. Fluorescence detectors FI1 (green) and FI3 (red) detected light in the specific wavelength ranges of 536 ± 40 nm and > 630 nm, respectively. Non-fluorescent scattered light was detected via a forward scatter detector (FSC) and a sideward scatter detector (SSC).

The CyFlow™ Cube 6 integrated software CyView™ version 1.8.0.87 allowed accurate calibration, cleaning, measurement, and maintenance steps. Every experimental day, cleaning and calibration steps had to be carried out. The cleaning was performed by consecutive rinsing of the equipment with sodium hypochlorite, decontamination solution, cleaning solution commercially available from Sysmex, and ultrapure water. Afterwards, the system was calibrated using the Sysmex product "Calibration Beads 3.0 µm, ready-to-use". This solution contained 3 µm solid latex microspheres that can be excited by the light of the used flow cytometer. It served as a reference standard for the right laser alignment.

Performing the appropriate measurement of TCC, HNA, and LNA required 1 mL of the aliquot in the sample tube. The aliquot consisted of 990 µL of the sample to be analyzed and 10 µL of the working solution containing the fluorescent dye. 1 mL working solution consisted of 10 µL SYBR® Green I nucleic acid gel stain (SG) and 990 µL sterile filtered 10 mM, pH 8, Tris(hydroxymethyl)aminomethane (Tris). SG was a commercially available product (Sigma-Aldrich, USA) in which the fluorescent dye is dissolved in dimethyl sulfoxide (DMSO) with factor 10^{-4} (Dragan et al., 2012). Corresponding spectra information is displayed in the appendix (Figure I-2). In the event of an abnormally high cell count in the investigated water, the aliquot in the sample tube had to be diluted before mixing with SG. For this purpose, sterile-filtered Evian natural mineral water was used as a dilution medium. After concluding mixing with or without dilution, the aliquot was incubated for 13 minutes at 37°C before being analyzed via flow cytometry. This preparation and staining process was adapted from the optimal procedure compiled by Ho et al. (2020).

After the experimental day, the flow cytometer had to be cleaned by the same cleaning procedure as before the measurements.

Flow Cytometry – Automation Unit

Automating flow cytometry required appropriate pre-treatment equipment, which was performed using the OC-300 add-on (onCyt Microbiology AG, Switzerland). A sketch of the au-

tomation unit flow scheme is provided in the appendix (Figure I-6). The device conducted sampling, mixing the sample with the working solution containing a fluorescent dye, the incubation procedure (37°C, 13 min), and transferring to the flow cytometer. Transportation was performed via an accurate syringe pump system. 350 µL of the undiluted water sample was mixed with 350 µL of the working solution containing fluorescent dye. Accordingly, fluorescent dye's concentration in the working solution was adjusted for automation. For 100 mL of working solution, 20 µL of SYBR® Green I nucleic acid gel stain (SG) (Sigma-Aldrich, USA) was mixed with 100 mL of sterile filtered Tris (10 mM, pH 8) containing sodium thiosulfate (50 mM). After measurement, the sample was automatically drained into a connected waste container.

The measurement frequency was determined using the implemented dashboard software cyOn-300 v3.12 (onCyt Microbiology AG, Switzerland). It was pertinent to note that there must be at least 15 minutes between two consecutive measurements to ensure the aforementioned incubation time of the sample. The appropriate cleaning steps of the automation unit, the flow cytometer, and all connecting tubes were automatically performed between measurements. This implied rinsing the following substances in a proper order: sodium hypochlorite solution (1 % active chlorine), Tris-buffer (10 mM, pH 8) containing sodium thiosulfate solution (50 mM), and ultrapure water. The presence of sodium thiosulfate was needed to quench former hypochlorite disinfection steps (Boal & Patsalis, 2017). All solutions had to be present in advance to guarantee a successful and automatic process. In addition, an overall cleaning process had to be carried out before and after the end of the measurement's series following an appropriate cleaning protocol.

Flow Cytometry – Data Evaluation

Analyzing flow cytometry data was performed in two ways. Regarding a non-automized system, the device integrated software CyView (Sysmex, Germany) was utilized. Different histograms and density plots could be displayed for scatter signals from the actual sample that have been detected by the software. To establish a more sophisticated data analysis, especially in regard to future automated monitoring challenges, the open-source toolbox FlowKit was applied in this work (White et al., 2021). To define a consistent gate regarding the discrimination between non-cells and cells (TCC), as well as HNA and LNA cells the fingerprint of Evian mineral water was used according to Ho et al. (2020). Once the gates were defined for various counts (TCC, HNA, and LNA), they were applied to all further water samples. Figure 3-1 shows the fingerprint of an exemplary water sample including all boundaries. Schuster (2024) provides detailed documentation of the code using the FlowKit toolbox.

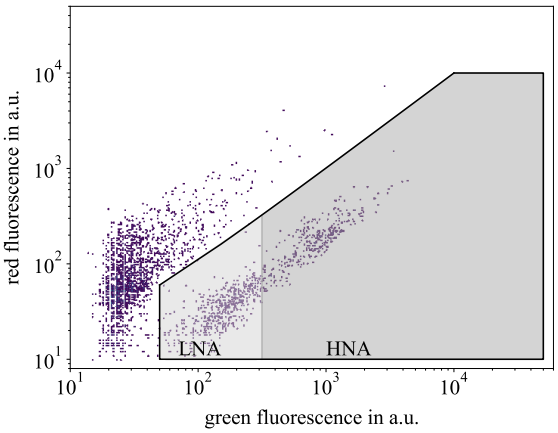

Figure 3-1: Flow cytometry gates in a 2-D density plot. Distinguishment between non-cells and cells (TCC in black bordered gate), LNA, and HNA cells.

3.2.3. Further Analytical Techniques

Total organic carbon (TOC) and dissolved organic carbon (DOC) were determined using a TOC analyzer (Shimadzu, Japan). For TOC, unfiltered water samples were measured, while for DOC analyses, water samples were filtered using a 0.45 µm syringe filter first according to DIN EN 1484 (2019).

Furthermore, DOM (0.45 µm filtered sample) was characterized using liquid chromatography (LC) in combination with an organic carbon detector (OCD) and an ultraviolet detector (UVD) working at 254 nm. The LC-OCD-UVD method was performed according to S. A. Huber et al. (2011). All required solutions are summarized in appendix (Table I-III). All DOM fraction concentrations – biopolymers, humic substances, building blocks, low molecular-weight acids, and low molecular-weight neutrals – were calculated by their occurring peak intensities via the software ChromCalc (DOC-Labor Dr. Huber, Germany).

Assimilable organ carbon (AOC) was determined from unfiltered water samples according to the flow cytometry procedure introduced by Hammes and Egli (2005). On a sampling day, 18 aliquots, 20 mL each, of every water sample were filled in autoclaved borosilicate glass bottles, immediately closed, and put on a shaker at a constant temperature of approximately 30°C. Three glass bottles were taken immediately (0 days) as well as 3, 7, 14, 21, and 28 days later, and the TCC was measured via flow cytometry. The difference between the mean value of the highest TCC during the 28 days (TCC_{max}) and the initial TCC (TCC_0) was converted to carbon equivalents by using the established correlation 10^{-7} μg_{AOC} $cells^{-1}$ (Hammes et al., 2006)

The 3630 IDS multiparameter measuring device was used to determine standard parameters of pH, conductivity, temperature, and dissolved oxygen content of water samples.

The total iron concentration was determined using the test kit Spectroquant® Crack Set 10 (Merck, Germany). This was initially used to decompose the complex bound iron before the total iron content was analyzed through a photometric method (extinction at 565 nm). The samples to be analyzed were acidified for measurement.

3.3. Experimental Setups

Water samples were analyzed offline or online. In this work “offline” entails sampling with subsequent measurement in the laboratory (chapters 3.3.1 and 3.3.3). “Online” in this context describes sampling and measuring directly on site (chapter 3.3.2).

3.3.1. Laboratory Water Characterization – Fingerprinting

For laboratory water fingerprints, samples were collected from different locations as described in chapter 3.1. Therefore, all samples were taken bubble-free and stored in annealed narrow-neck glass bottles under anaerobic conditions. Between sampling and offline analysis, all samples were stored dark and cold at approximately 4–8°C. The measurements were done the sampling day or the day after. Figure 3-2 represents an overview of the corresponding analyzing methods regarding different water samples.

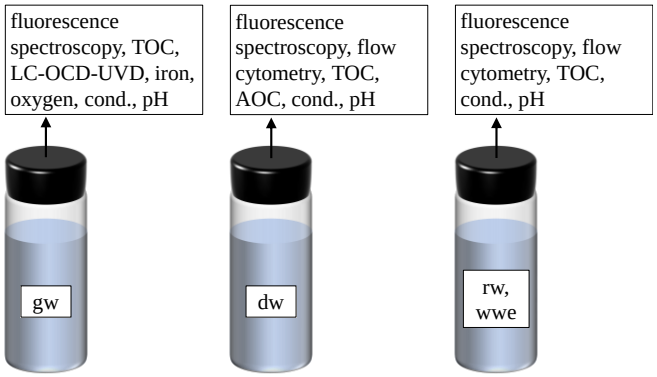

Figure 3-2: Respective offline analyzing methods to characterize groundwater (gw), drinking water (dw), wastewater effluent (wwe), and rainwater (rw).

Drinking water did not require any pretreatment process, however wastewater effluent, rainwater, and some groundwater samples needed an appropriate filtration and/or dilution step before being analyzed. The filtration was generally conducted using a 0.45 µm syringe filter. All samples with DOC concentrations higher than 5 mg L^{-1} were diluted in respective amounts of ultrapure water from the Millipore Direct-Q 5 UV System (Merck, Germany).

Special consideration is necessary when analyzing groundwater samples part of a large-scale measurement campaign to identify changes in humic substances in the groundwater of the corresponding wells. The reduced character of groundwater requires a sophisticated treatment due to its high sensitivity to oxygen. The contact between reduced groundwater and oxygen leads to an increase in the redox potential which causes an immediate start of the iron precipitation process (Worch, 1997). Due to the possible adsorption of DOM molecules on iron flocs, oxygen uptake can also affect the DOM constitution in water. This should be avoided regarding the organic characterization via fluorescence spectroscopy and LC-OCD (Schulz, 2020).

Thus, groundwater samples were taken with a syringe immediately after opening the sampling bottle and dispensed into the fluorescence spectroscopy through-flow cell. A sketch showing the principle is attached in appendix (Figure I-7). To avoid iron precipitation during the long-

term LC-OCD analysis (≈ 120 min), groundwater samples were taken directly in an appropriate LC-OCD sample container oxygen-free at the sampling point and analyzed subsequently. The Master's thesis of Eberhard (2022) provides further detailed information regarding groundwater sample treatment.

3.3.2. Automated Laboratory Experiments for Monitoring Drinking Water Quality

The first part of the following section highlights the proof of concept of the online feasibility of fluorescence spectroscopy and flow cytometry measurements with consecutive data analysis – 'Proof of Concept for Automated Drinking Water Quality Monitoring'. This setup was initially introduced by Schuster et al. (2022a) and Schuster et al. (2022b).

The second part, 'Detection of Crucial Changes in Drinking Water Quality via Automated Monitoring', emphasizes the automation of data evaluation and visualization. This includes the feasibility of using PARAFAC fluorescence spectroscopy and flow cytometry to detect changes in drinking water quality. A schematic overview of the physical automatic monitoring setup is shown in Figure 3-3. Additional information regarding the automation coding structure can be found in the Master's thesis of J. Huber (2022).

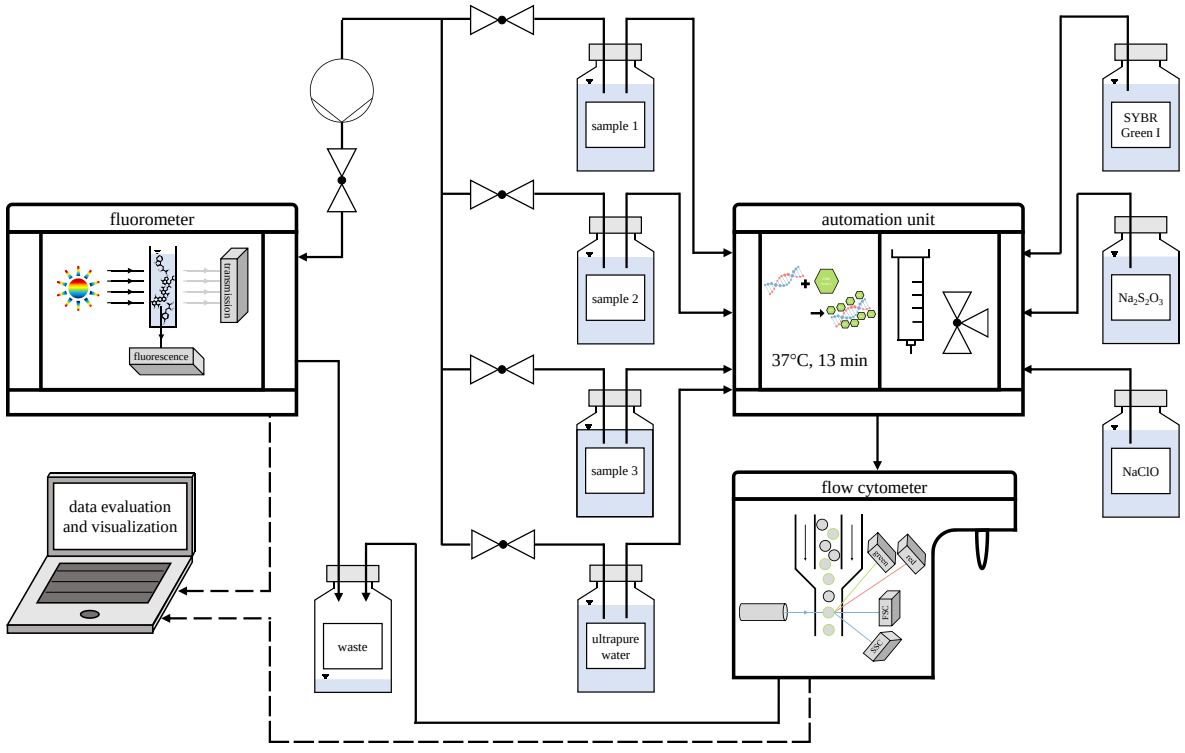

Figure 3-3: Schematic overview of the laboratory setup for automated drinking water quality monitoring experiments.

Proof of Concept for Automated Drinking Water Quality Monitoring

The following experiments describe the parallel, continuous, and automatic drinking water sampling and analysis. Two experimental rows were conducted investigating three samples each,

which led to six experiments in total (i)–(vi). One sample consisted of one glass bottle containing 1 L of drinking water (dw_1) which was filled from the laboratory tap immediately before the experiments were executed. While experiment (i) monitored a sample of sole dw_1, the dw_1 of experiments (ii)–(vi) was spiked with increasing concentrations of the organic nutrient medium Luria-Bertani (LB) at the beginning of each experiment (t = 0 h). Table 3-1 shows the respective spiked LB medium volumes for each experiment. LB medium is known as a conventional cultivation medium for *E. coli* and mainly consists of oligopeptides and amino acids (Sezonov et al., 2007). Here it was applied to serve as an organic nutrient medium for all types of present bacteria in dw_1. LB medium is synthesized by dissolving 5 g of yeast extract, 10 g of peptone, and 10 g of NaCl into 1 L of ultrapure water and adjusting the pH to 7.

The experiments were performed for 70 hours at room temperature (22 ± 1°C) under continuous stirring. Automatic flow cytometry sampling was conducted every 90 minutes and fluorescence spectroscopy every 180 minutes, respectively. In the meantime, all flow cytometer connected tubes and the cuvette were automatically disinfected (1 % sodium hypochlorite), quenched (sodium thiosulfate) and rinsed (ultrapure water). Fluorometer connected tubes and the cuvette were rinsed with ultrapure water only.

Table 3-1: Luria-Bertani (LB) medium added to respective dw_1 samples (1 L) to perform online monitoring experiments 'Proof of Concept for Automated Drinking Water Quality Monitoring'.

dw_1 experiment	spiked volume of LB medium in µL
(i)	0
(ii)	14
(iii)	71
(iv)	142
(v)	286
(vi)	429

Detection of Crucial Changes in Drinking Water Quality via Automated Monitoring

This part focuses on automating parallel measurement and data analysis. The data analysis part aimed to gain a deeper understanding of how online monitoring setups can be structured with respect to software toolboxes. To evaluate the developed software toolboxes and define the sensitivity of online monitoring parameters, automated experiments on monitoring drinking water quality changes were performed.

For this purpose, three samples (1 L) of the same drinking water (dw_2) were monitored in three different experiments (vii)–(ix) for eight hours each at room temperature (22 ± 1°C) under continuous stirring. Contrary to previously described experiments containing dw_1, dw_2 originated from a WDS hydrant. Water was sampled from the hydrant a week before and stored appropriately until the experiment began. During the experiments, dw_2 was spiked with increasing concentrations of different waters, namely wastewater effluent (wwe_1), rainwater (rw_1), and varying drinking water (dw_3). Automatic flow cytometry and fluorescence spectroscopy sampling

was conducted every 15 minutes. In the meantime, all flow cytometry tubes and the cuvette were automatically disinfected (1 % sodium hypochlorite), quenched (sodium thiosulfate) and rinsed (ultrapure water). Fluorescence spectroscopy tubes and the cuvette were rinsed with ultrapure water only. The general procedure for monitoring experiments was as follows:

(I) Preparation and cleaning procedure for devices, tubes etc.
(II) Introductory phase of at least 15 continuous dw_2 measurements
(III) 1st spiking event (approx. 0.1 v%) followed by at least four continuous measurements
(IV) 2nd spiking event (approx. 1 v%) followed by at least four continuous measurements
(V) 3rd spiking event (approx. 10 v%) followed by at least four continuous measurements
(VI) End of the experiment and final cleaning of devices, tubes, etc.

(I) was conducted according to the procedures described in chapter 3.2. *(II)* was performed to stabilize the system and define a static baseline for each parameter (Equation 2-5). For water quality monitoring, a parameter specific baseline represents the limit of detection for each parameter. While monitored measurements within upper and lower baseline limits were considered as no parameter respective quality change, monitored measurements outside upper and lower baseline limits were considered as either outliers, events, or baseline changes.

To avoid any effects by decreasing the volume of the aliquot, the volume of dw_2 was kept constant at 1 L by refilling the sample bottle with the same dw_2 after each measurement. Thus, the subsequent spiking events *(III)–(V)* were properly adjusted and quantified. After each spiking event, the current status of water quality was monitored for at least four measurements before the next spiking event was performed. The respective amounts of spiking water were chosen based on comparable laboratory contamination experiments published by Besmer et al. (2017), Favere et al. (2021) and Sadler et al. (2020). The conditions of the three experiments that were tested are summarized in Table 3-2.

Table 3-2: Three laboratory experiments of spiking drinking water (dw_2) with three deviating waters. The experiments were part of 'Detection of Crucial Changes in Drinking Water Quality via Automated Monitoring'.

dw_2 experiment	spiking water	spiking amount in v%
(vii)	wastewater effluent (wwe_1)	0.1, 1, 10
(viii)	rainwater (rw_1)	0.1, 1, 10
(ix)	drinking water (dw_3)	0.1, 1,10

3.3.3. Pilot Plant Drinking Water Quality Experiments

The trials introduced in this chapter serve to transfer and evaluate the previous findings from the laboratory setup into the more realistic case of a water flowing pipe system. The pilot plant pipe system used in this work is called the pilot plant in further course of this work. Detailed explanations regarding the system setup and performed experiments can be found in the Master's thesis of Wullenweber (2022).

Pilot Plant Setup

Initially, the pilot plant was designed as an *open system* to perform flushing tests of a piping system, as described by Guzinska (2018). As part of the current work in 2022, the system was configured to enable the system to be closed into a complete pipe loop system, *closed system*. For this purpose, the tank including the pump is bypassed and the flow is realized by two centrifugal pumps connected in parallel. A sketch-like flow chart of the *closed system* is given in Figure 3-4.

The pilot plant was supposed to simulate just a small part of a real WDS including realistic instruments, dimensions, and flow characteristics. The *closed system* consisted of approximately 90 % cast iron pipes with a DN200 diameter. The filled pilot plant with a pipe length of 152 m contained approximately 4400 L of water. During water quality monitoring experiments, the two centrifugal pumps produced stable flow conditions with a flow rate of 58 ± 4 L min^{-1} (Re ≈ 5700), equivalent to a flow rate of approximately 0.5 m s^{-1} in DN200 pipes, and an operating pressure of about 1.1 ± 0.1 bar. The pump characteristics are attached to appendix (Figure I-8). During these conditions, the water took approximately 75 ± 5 min to complete one loop in the *closed system*. Regarding the ambient temperature of the pilot plant, it should be mentioned that there was a gradient here, as 1/5 of the pilot plant pipe network is installed indoors, and the remaining part outdoors. For this reason, the indoor area was temperature-stable (23 ± 3°C) while the outdoor part depended significantly on the outside temperature.

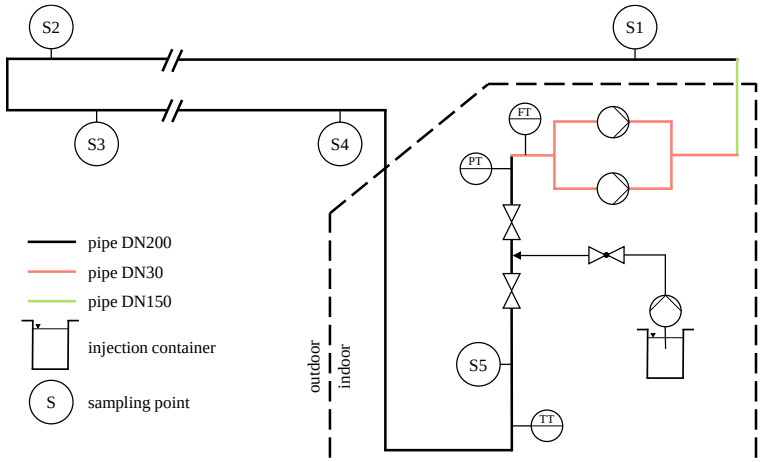

Figure 3-4: Flow chart sketch of the *closed* pilot plant, adjusted for water quality monitoring experiments.

Preliminary Tests

Before starting the experiments, the pilot plant was extensively cleaned through a high-pressure cleaner to remove deposits, biofilms, and anthropogenic substances from inside the pipes. Subsequently, as well as between all experiments, the pilot plant was cleaned using the established air-water-flushing process (v = 0.5 m s^{-1}) according to Korth and Donarth (2015). This was performed using the *open system* by filling the tank with water from a nearby hydrant. After the appropriate cleaning of the pilot plant, the following preliminary tests were performed:

- Investigating the water temperature profile within the pilot plant regarding indoor and outdoor conditions.
- Defining the sampling depth within the pipe (0, 50, and 100 mm).
- Defining the sampling point location (S1–S5).
- Using a tracer (NaCl) to test pilot plant flow conditions.

The temperature profile showed a significant influence of the outside temperature. After running the pilot plant system for approximately 150 min, the temperature difference within the system could be considered negligible. Thus, a preparation or stabilization phase of 150 min was recommended before starting the water quality monitoring experiments. Considering the sampling depth, the water quality of the flowing system was analyzed at the pipe center (100 mm), at 50 mm, and at the pipe wall (0 mm). No deviations were observed between those sampling depths which is why sampling during the water quality monitoring experiments was chosen to be conducted at the pipe center (100 mm). Water quality was compared between sampling locations S1–S5. It was proven that there were no water quality deviations between all sampling locations after approximately two hours. Thus, S1 (Figure 3-4) was chosen for the water quality monitoring experiments. All results of those preliminary tests are summarized in the appendix (Figure I-9).

In another preliminary experiment, an injected NaCl solution was used as a tracer to test the period of time for which injected water could be observed at sampler S1. Since water flows in the loop of the *closed system*, the NaCl solution was expected to pass S1 several times. The measured peak amplitude of the NaCl solution signal at S1 confirmed the water flow rate in the *closed system*. However, due to the progressive mixing in the pipes of the water in the pilot plant and the NaCl solution, the signal amplitude decreased between the first and second time passes by S1. Based on this preliminary result, the sampling interval for the following water quality monitoring experiments was appropriately adjusted. The results of this preliminary trial are shown in appendix (Figure I-10).

Water Quality Monitoring Experiments

The main part of this chapter's investigations are the water quality monitoring experiments, conducted in the *closed system*. The experiments were aimed at describing fingerprinting methods PARAFAC fluorescence spectroscopy and flow cytometry in realistic applications and determining limitations and further advantages compared to conventional monitoring methods.

The system was filled (approximately 4400 L) with drinking water from the hydrant (dw_4) the day before the actual experiments started. On the experimental day, the centrifugal pumps were started in the morning after the stabilization phase of approximately 150 min, the pilot plant was ready to use. The idea of the water quality monitoring experiments was to simulate a scenario of an unexpected and abnormal water quality change, e.g., due to small pipe leakages, seasonal variations, well circuit changes, or drinking water reservoir contaminations. Therefore,

a different water sample of about 44 L was injected through a manual pump into the pilot plant (injection container Figure 3-4). The injection procedure took between 20 and 26 minutes. The average injection volume flow was thus 2 ± 0.5 L min^{-1}. Wastewater effluent (wwe_2), rainwater (rw_2), and varying drinking water (dw_5) were used as injection waters in this work. After the injection, water quality was monitored at sampling location S1. Therefore, samples were taken at intervals of 3 to 15 minutes over an experimental duration of approximately 210 min. The sampled water was directly filled into autoclaved glass vials and stored cold and dark until analyzed offline in the laboratory. After each water quality monitoring experiment, the water was drained, the pilot plant was rinsed properly using dw_4 and was thus ready for the subsequent experiment. An overview of the three performed water quality monitoring experiments is summarized in Table 3-3.

Table 3-3: Monitoring the quality of drinking water (dw_4) through three pilot plant injection experiments. 1 v% of the total pilot plant volume (4400 L) of each water was injected under pilot plant flow conditions.

dw_4 experiment	injecting water
(x)	wastewater effluent (wwe_2)
(xi)	rainwater (rw_2)
(xii)	drinking water (dw_5)

Defining a parameter specific baseline including its upper and lower limits is crucial when detecting water quality changes. The baseline of each parameter was determined using Equation 2-5. The mean and standard deviation values required for this equation were used from samples taken the last hour of the stabilization phase. This was just before injection was performed. Regarding the detection of water quality changes during the experiments, monitored measurements within upper and lower baseline limits were considered as no parameter respective quality change. In contrast, monitored measurements outside upper and lower baseline limits were considered as either outliers, events, or baseline changes.

4. Characterization of Drinking Water

This part screens several drinking water samples and their individual constitution of DOM and present bacteria. Therefore, samples from different locations within a real WDS (dw_{wds}) and samples from different waterworks exits (dw_{wwe}) were investigated. These were characterized via PARAFAC fluorescence spectroscopy, flow cytometry, TOC analysis, AOC analysis, as well as conductivity analysis, and pH analysis. However, not all drinking water samples were analyzed by each analytical method. Individual characteristics and eventual parameter relations and correlations are discussed in the following chapters.

4.1. An Overall Six-Component PARAFAC Model

A generated PARAFAC model, especially the spectra of its components, is strongly dependent on the samples considered for modelling (Stedmon et al., 2011). For example, a model valid for a particular set of drinking water samples does not have to fit all future drinking water samples. This is even less for different types of water e.g., surface water or wastewater. In general, PARAFAC-related work involves collecting a certain amount of data and generating the model subsequently (Sciscenko et al., 2022).

Fluorescence data from all collected drinking water samples (n = 30) were merged to generate a PARAFAC model. To stabilize the PARAFAC model generation, fluorescence data from 72 water samples, called *base data*, were added to the data set. The *base data* consists of diverse drinking water samples of which some were artificially modified by spiking with slightly varying amounts of different water e.g., varying drinking water, rainwater, or wastewater effluent (Table I-II). By adding the *base data*, the aim was to create a PARAFAC model that was as sophisticated and reliable as possible. In particular to describe and characterize external influences on drinking water quality, e.g., rainwater or wastewater infiltration. Moreover, the addition of *base data* samples enhanced a comparison of PARAFAC fluorescence spectroscopy results achieved in the current chapter with the results of online monitoring experiments carried out later in this work (chapter 5.3).

Using the combined data set of all collected drinking water samples and *base* data, the PARAFAC modelling procedure was performed through iteration loops according to the explanations in chapters 2.2.3 and 3.2.1. The analysis of spectral loadings, sum of squared errors and core consistency were applied as the main criteria to determine the correct number of components. Models with three to seven components were generated for this purpose. Based on the spectral loadings, the three- and four-component models could be excluded. Considering the emission spectrum, one component showed either a second peak (three-component model) or a clear peak shoulder (four-component model), both of which indicated that the demonstrated component cannot be correctly represented by the low number of components (Murphy et al., 2013). The lowest sum of squared errors indicates the most probable number of components in a PARAFAC model (Wünsch, 2024). There was a significant decrease in the sum of squared errors

between the five- and six-component model, while the difference between the six- and seven-component model was negligible for most samples (> 90 %). In terms of core consistency, no significant changes were observed between component 6 and 7, respectively, indicating a rather six-component structure (Bro & Kiers, 2003). To avoid overdetermination of the majority of the samples, a six-component model was chosen as the exact solution for the analyzed samples. However, the fact that some samples are probably better represented by a seven-component model should be noted in the following discussion.

The obtained PARAFAC model identifying six fluorescent components is from here on called the overall six-component PARAFAC model. The fluorescence spectrum (EEM), which is the fingerprint, of each component of the overall six-component PARAFAC model is displayed in Figure 4-1. The different components are classified as C1–C6.

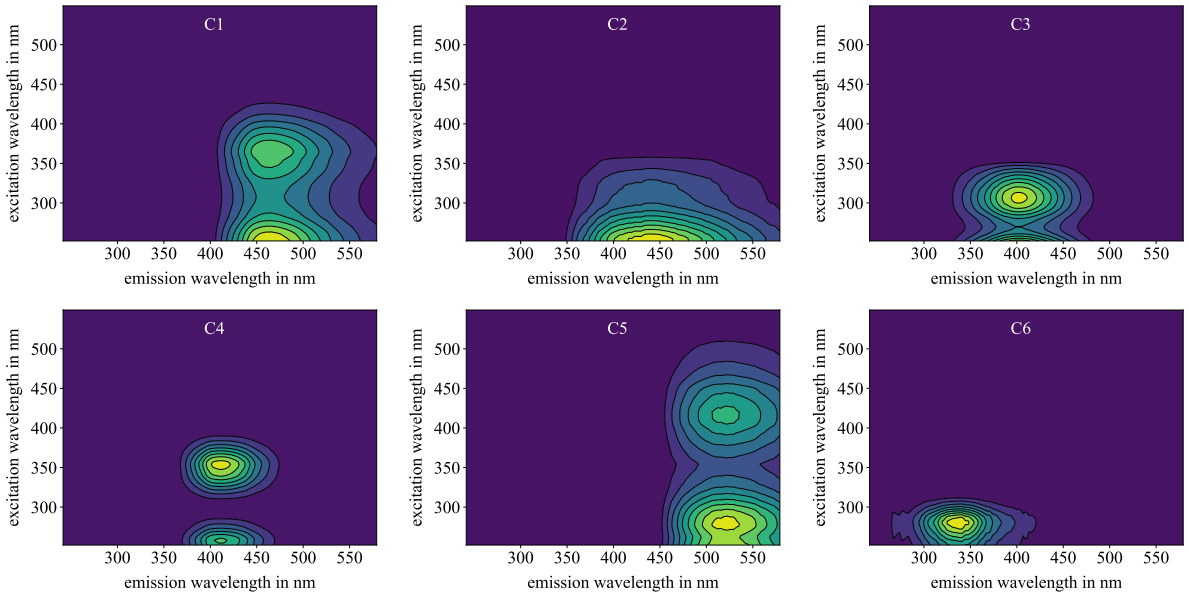

Figure 4-1: Fluorescence spectroscopy EEM fingerprints of each component (C1–C6) of the overall six-component PARAFAC model.

While the primary excitation peak of all components appears in similar wavelength areas (< 300 nm), the emission peak varies over the entire spectrum. According to Wagner (2014) a shift in red emitted light can be correlated to an increase in the molar weight of the respective fluorophore/component. With this information, the respective components can be categorized into their molecular sizes: C6 < C3 ≈ C4 < C1 ≈ C2 < C5. With regard to the peak nomenclature introduced in chapter 2.2.2, this is consistent with the molar weights of amino acid- or tryptophan-like C6 (B and T peak) (Coble et al., 1998; Parlanti et al., 2000; Stedmon & Markager, 2005), fulvic acid-like C1 and C4, and humic-acid-like C5 components (Barsotti et al., 2016; Coble et al., 2014). Tryptophan has a molecular weight of approximately 0.2 kg mol^{-1}, whereas fulvic acids are located in the range of 0.5–2 kg mol^{-1}, and humic acids are typically 1–10 kg mol^{-1} (Thurman et al., 1982).

To gain further information about the six components, the overall six-component PARAFAC model was uploaded to the open-access *OpenFluor* online library (Murphy et al., 2014b). Based on the excitation and emission wavelengths (λ_{ex} and λ_{em}) of each component's peak, *OpenFluor* enables comparison with all related published work. Based on these comparisons, a general characterization of the components was performed and is summarized in Table 4-1. The components are classified into allochthonous substances (DOM that is transported from external sources) and autochthonous substances (DOM that is produced in situ), components with terrestrial or microbial origin, and humic-like or protein-like components. Moreover, it is shown in which natural and engineered aquifers the respective component has been found during former studies.

Table 4-1: Overview of the overall six-component PARAFAC model including spectral peak wavelengths, substance or substance-like designations, main occurrence in natural and engineered water systems and corresponding references for all components. The comparison with the literature was performed via the *OpenFluor* online library (Murphy et al., 2014b).

component	λ_{ex} peak in nm	λ_{em} peak in nm	substance	occurrence	references
C1	252/370	465	humic-like, fulvic acid-like, allochthonous, terrestrial	seawater river water lake water	[1][2] [3][4] [5]
C2	252/325	445	humic-like, allochthonous and autochthonous, microbial and terrestrial	tap water river water lake water seawater	[6][7] [8][9] [10] [11]
C3	252/303	400	humic-like, autochthonous, microbial	river water seawater wastewater drinking water	[4][9] [12] [13] [14]
C4	258/350	415	humic-like, fulvic acid-like, microbial and terrestrial	river water seawater wastewater drinking water	[4] [2][15] [13][16] [14]
C5	275/420	520	humic-like, humic acid-like, terrestrial	river water drinking water soil	[4][8][17] [14] [18]
C6	280	337	amino acid-like, tryptophan-like, autochthonous microbially produced	seawater lake water river water drinking water	[1][2] [5][19] [4][20] [8][14]

[1] Gonçalves-Araujo et al. (2015), [2] DeFrancesco and Guéguen (2021), [3] Bernal et al. (2018), [4] Retelletti Brogi et al. (2020), [5] Du et al. (2021), [6] Hambly et al. (2015), [7] Shutova et al. (2014), [8] Moona et al. (2021), [9] Peleato et al. (2017), [10] Jutaporn et al. (2020), [11] Drozdova et al. (2022), [12] Zhou et al. (2019), [13] Murphy et al. (2011), [14] Wells et al. (2022), [15] Stedmon and Markager (2005), [16] G.-H. Yu et al. (2010), [17] Wünsch and Murphy (2021), [18] Kida et al. (2021), [19] Lapierre and del Giorgio (2014), [20] Graeber et al. (2021)

The following sections focus on a detailed literature review regarding the displayed results in Table 4-1. The focus here is on the detailed characterization of the substance of each component. This is particularly relevant to subsequently presented results of this work.

C1 was mainly found in marine waters (DeFrancesco & Guéguen, 2021; Gonçalves-Araujo et al., 2015) and other surface waters such as rivers (Bernal et al., 2018; Retelletti Brogi et al., 2020) and lakes (Du et al., 2021). In terms of drinking water treatment and production, only two studies highlighted a similar PARAFAC component such as C1 (Bernal et al., 2018; Wells et al., 2022). All authors agree that C1 can be assigned to allochthonous terrestrial humic-like FDOM. More detailed Wells et al. (2022) described C1 as a fulvic acid-like component.

C2 was observed without any specific preferences in various waters such as tap water (Hambly et al., 2015; Shutova et al., 2014), river water (Moona et al., 2021; Peleato et al., 2017), lake water (Jutaporn et al., 2020) and seawater (Drozdova et al., 2022). According to these references, C2 describes an allochthonous and autochthonous humic-like compound of microbial and terrestrial origin, respectively. These ambiguous descriptions can also be caused by the overlapping spectral properties between C2 and C4. Most of the mentioned literature works with a four- or five-component PARAFAC model, which can lead to both components being represented by only one component in their specific models. All authors indicated C2 as a main constituent regarding the fluorescence intensity in their waters. Furthermore, Fellman et al. (2010) and Ishii and Boyer (2012) described C2 component as a fewer aromatic or conjugated bond compound with lower hydrophobicity.

C3 was observed in various waters as well, such as river water (Peleato et al., 2017; Retelletti Brogi et al., 2020), sea- or coastal water (Zhou et al., 2019), treated wastewater (Murphy et al., 2011), and drinking water (Wells et al., 2022). All authors agreed that this component belongs to a group of autochthonous microbial humic-like compounds. A direct relation to microbial activity was identified by Kinsey et al. (2018) who showed that C3 can be produced by bacterial reprocessing of fresh phytoplankton-derived organic matter. Moreover, Fellman et al. (2010) described C3as a compound with fewer aromatic structures or conjugated bonds and lower hydrophobicity. Further, C3 is described as a general critical component in water treatment considering its effects on DBPs, membrane fouling, nitrification, and denitrification (Wells et al., 2022).

On the one hand, C4's character is described by the previously mentioned similarity to C2 (microbial and terrestrial humic-like component). In addition, the excitation and emission spectra are in intermediate between C1 and C3. C4 was identified in several waters, such as river water (Retelletti Brogi et al., 2020), marine or coastal water (DeFrancesco & Guéguen, 2021; Stedmon & Markager, 2005), treated and untreated wastewater (Murphy et al., 2011; G.-H. Yu et al., 2010), and drinking water (Wells et al., 2022). There is no convincing argument regarding the allochthonous or autochthonous character of C4. Murphy et al. (2011) utilized C4 as a

wastewater or nutrient enrichment tracer and described its similarity to syringaldehyde, although with a shift in the excitation spectrum.

C5 was mainly identified occurring in river water (Moona et al., 2021; Retelletti Brogi et al., 2020; Wünsch & Murphy, 2021) but also in drinking water (Wells et al., 2022) or soil samples of a mangrove forest (Kida et al., 2021). All authors described C5 as a terrestrially derived humic-like compound with rather humic acid-like character. The compound can be roughly described by its macromolecular structures of aromatic conjugated substances (Osburn et al., 2016) which can be related to lignin phenol concentration (Walker et al., 2013).

Apart the humic-like behavior of C1–C5, C6 belongs to the group of protein-like, or more precisely tryptophan-like, substances. Its appearance was reported in marine water (DeFrancesco & Guéguen, 2021; Gonçalves-Araujo et al., 2015), lake water (Du et al., 2019; Lapierre & del Giorgio, 2014), river water (Graeber et al., 2021; Retelletti Brogi et al., 2020), drinking water (Moona et al., 2018; Wells et al., 2022) and artificial solutions (Catalán et al., 2021; Du et al., 2019). C6 is often described as freshly microbially produced protein-like, but also as a microbially derived humic-like material (Fellman et al., 2008). Coble (1996) observed the presence of protein-like components associated with recent biological production. A study by Lapierre and del Giorgio (2014) highlighted its relationship with BDOC, although no conclusion could be drawn about whether C6 functions as a substrate for microorganisms or as a biolabile molecule tracer. Gonçalves-Araujo et al. (2015) suggested that C6 seems to be an indicator of the bioavailability of DOM.

The fluorescence scores, symbolizing the fluorescence intensities in Raman units generated by the PARAFAC analysis, of the respective components of all fluorescence spectroscopy investigated drinking water samples (n = 30) are displayed in Figure 4-2. Samples were taken from different locations within the WDS (dw_{wds}, n = 19 samples) and from different waterworks exits (dw_{wwe}, n = 11 samples).

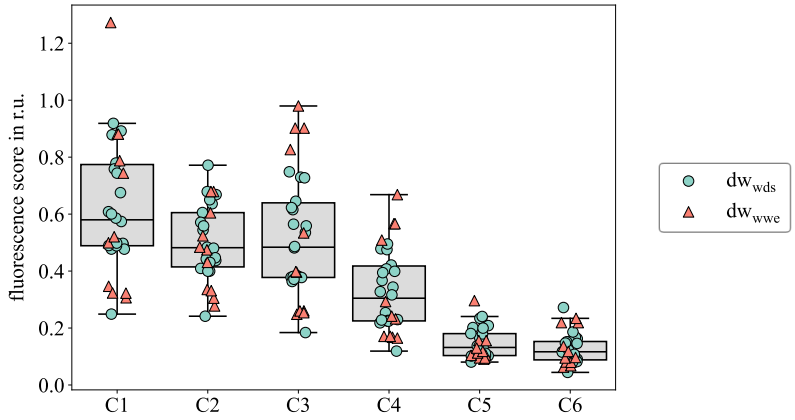

Figure 4-2: Fluorescence scores of the respective six components (C1–C6) of investigated drinking water samples at various locations within the water distribution system (dw_{wds}), n = 19, and at waterworks exits (dw_{wwe}), n = 11. The joint data evaluation (n = 30) is presented in the respective boxplots.

Considering the absolute values, the fluorescence scores of the respective components show a large scatter, especially for C1–C4. There is no recognizable trend in the dependencies of the fluorescence scores on the sampling point (dwwds and dwwwe). This confirms the findings of (Gabrielli et al., 2023) and (Philibert et al., 2022) who highlighted that the fluorescence score of drinking water was mainly affected by the type of water treatment and the respective untreated source water, and did not show any significant variation within the WDS.

Apart from absolute fluorescence scores, a recommended way of analyzing PARAFAC results of water samples is to describe the respective proportion of the fluorescence score of one component (Murphy et al., 2013). To reveal an idea about the relative proportion of each component, boxplot results (Figure 4-2) containing all water samples were used. Therefore, each component's interquartile range (IQR) between quartile one (Q1 – lower horizontal line) and quartile three (Q3 – upper horizontal line) was considered in relation to the accumulated range of all components' fluorescence scores (boxplot explanation in appendix Figure II-1). Since the calculated range is based on only 50 % of the data (IQR), the following fluorescent proportions of each component are only approximate values and can deviate considerably for each individual drinking water sample.

C1: 24–37 %, C2: 20–29 %, C3: 18–30 %, C4: 11–20 %, C5: 5–9 %, and C6: 4–9 %.

Based on these results, only a minority (< 10 %) of the total FDOM in drinking water counts as protein-like FDOM (C6), while the remaining compounds belong to humic-like FDOM (C1–C5). Regarding the amount of humic-like hydrophobic acids in drinking water, the results show a dominant occurrence of fulvic acid-like components C1 and C4 compared to the humic acid-like component C5 (approximate ratio of 6:1). What emerges from these investigations is that both the fluorescence score of each component and its relative amount of the total FDOM can be utilized to uniquely characterize a drinking water sample.

Comparing the fluorescence score ranges of each component and their relative ratios with related studies investigating drinking water related samples, it should be considered that results strongly depend on the number of components generated by PARAFAC modelling. Bernal et al. (2018) and Huang et al. (2015) reported high concentrations of components C1 and C3 in groundwater, although they used a four-component model. The results from Figure 4-2 carefully confirm these observations for drinking water which in this case is entirely derived from groundwater. Hambly et al. (2015) described C2 as the main constituent (approximately 45 %) of tap water in their five-component model. However, C4 is missing in their model which was described above with similar spectra to C2. This presumably leads to signals from both components being combined together into a single component would have shown the highest fluorescence score. Wells et al. (2022) identified fulvic acid-like components (C1 and C4) as dominant in drinking water followed by C3, C6, and C5. Apart from C6, their relative amount of fluorescence scores agrees with the results shown here. However, the fluorescence score of C6 in Wells et al. (2022) is slightly higher, which might also be explained by the limitations of

their four-component model. Shutova et al. (2014) identified C1 and C2 as components with relatively high fluorescence score for investigated water samples before drinking water treatment. During the treatment steps, all fluorescence scores were reduced. However, they observed better treatability for removing terrestrial FDOM (C1, C5) than for processed or microbially-delivered FDOM (C3, C6). Although only non-disinfected drinking water samples were analyzed in the current study, the fluorescence score of each component may play a role if disinfectants are used within the WDS (Fernández-Pascual et al., 2023). In the case of drinking water disinfection, the intensities of all components should be as low as possible to minimize the potential of DBPs formation.

Highlights Chapter 4.1

- All investigated drinking water samples were successfully characterized by the generated overall six-component PARAFAC model (six components C1–C6). Due to the component specific fluorescence scores, each drinking water sample was described individually.
- Using the provided data set, called *base data*, the six-component PARAFAC modelling procedure was successfully stabilized. Thus, the consistent and sophisticated generation of the overall six-component PARAFAC model was ensured.
- All investigated drinking water samples mainly consist of terrestrial and microbial humic-like (C2, C3) or fulvic acid-like (C1, C4) compounds, while humic acid-like (C5) and protein-like (C6) compounds play a subordinate role (general proportion in drinking water samples: C1, C2, C3 > C4 > C5, C6).
- There are no general differentiations between the DOM constitution of drinking water samples from waterworks exits and drinking water samples from different locations within the WDS.

4.2. Flow Cytometry and Others

Figure 4-3 represents the results of flow cytometry analyses of 30 drinking water samples from different locations within the WDS (dw_{wds}, n = 19) and at waterworks exits (dw_{wwe}, n = 11). In addition to the absolute numbers, the proportion of HNA cells in TCC is displayed (%HNA). The respective median numbers are 1.4×10^5 cells mL^{-1} for TCC, 5.2×10^4 cells mL^{-1} for HNA, 9.0×10^4 cells mL^{-1} for LNA and 39 % for %HNA.

Concerning TCC and HNA, there are respectively two dw_{wwe} samples which can be considered outliers compared to the remaining samples. The corresponding LNA values for these dw_{wwe} samples are the two closest to the upper limit and above, respectively. Comparing flow cytometry and fluorescence spectroscopy, it should be mentioned that the most deviating outlier from flow cytometry is the exact same dw_{wwe} sample as the identified C1 outlier in Figure 4-2. Apart from these two outliers, all other samples show uniform distribution regardless of their origin

(dw_{wds} or dw_{wwe}). The TCC range of 10^3 to 10^5 cells mL^{-1} reflected also by Berney et al. (2008), Ho et al. (2020) and Kötzsch and Sinreich (2014) could thus be confirmed for the drinking waters investigated in this study. According to the proposed TCC limit value of 5×10^5 cells mL^{-1} by Van Nevel et al. (2017b), all investigated waters except one outlier would be below this value and thus indicate good drinking water quality with respect to this parameter. The distribution of HNA cells proportions shows a similar behavior to absolute numbers. Apart from one outlier, all drinking water samples are in a similar range.

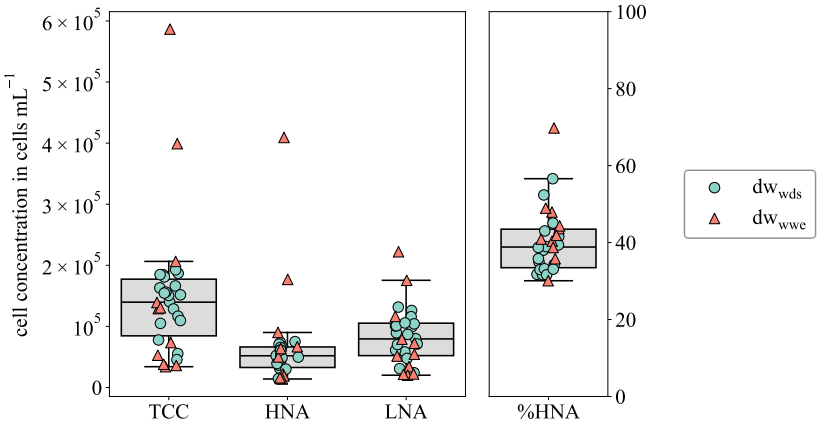

Figure 4-3: Flow cytometry results (total cell count – TCC, high nucleic acid cell count – HNA, low nucleic acid cell count – LNA, proportion of HNA cell count – %HNA) of investigated drinking water samples at various locations within the water distribution system (dw_{wds}), n = 19, and at waterworks exits (dw_{wwe}), n = 11. The joint data evaluation (n = 30) is presented in the respective boxplots.

There is extensive literature dealing with drinking water HNA and LNA amounts. Y. Wang et al. (2009) isolated LNA cells by 0.45 µm filtration of the water and thus described their size and cultivability. They investigated the LNA/HNA ratio of different water, e.g., surface water, tap water, groundwater, and wastewater treatment plant effluent, and described tap water with a ratio of approximately 47 % HNA cells (here: median value at 39 %). Prest et al. (2014) investigated %HNA at waterworks exits and within the WDS. They observed an increase between both locations of approximately 6 %, with total ranges between approximately 35 and 44 %. Simultaneously they observed an increase in TCC. A direct comparison between the water at a specific waterworks exit and the same water at a subsequent WDS location was not made in the present study. Prest et al. (2014) showed a relation between %HNA increase and a shift in bacterial community composition by combined 16S rRNA gene pyrosequencing. This confirms the former work of Müller et al. (2010) and Vila-Costa et al. (2012) that revealed the same relation between %HNA and bacterial community shifts, but in different environments. The determined values of %HNA in drinking water presented in these studies showed consistency with the results of the current study. Minor deviations may also result from individual gate adjustments. In addition to non-disinfected water systems, Jie et al. (2017) examined drinking water from a chlorinated WDS. Although chlorination was applied after the treatment, they determined LNA cells as a predominant community (approximately 67 ± 17 %) at various locations of the WDS, but with generally low cell numbers (< 3×10^4 cells mL^{-1}).

In addition to fluorescence spectroscopy and flow cytometry, drinking water samples were analyzed to investigate their TOC (n = 35) and AOC (n = 27) concentrations as well as their conductivity (n = 34) and pH (n = 34). The respective results are shown in Figure 4-4.

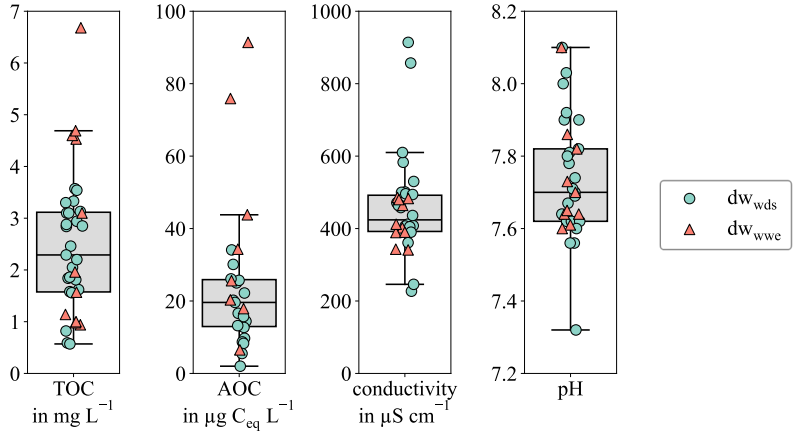

Figure 4-4: TOC, AOC, conductivity and pH results of investigated drinking water samples at various locations within the water distribution system (dw_{wds}) and at the waterworks exits (dw_{wwe}). TOC: n(dw_{wds}) = 24, n(dw_{wwe}) = 11, AOC: n(dw_{wds}) = 19, n(dw_{wwe}) = 8, conductivity: n(dw_{wds}) = 24, n(dw_{wwe}) = 10, pH: n(dw_{wds}) = 24, n(dw_{wwe}) = 10. The joint data evaluation for the respective parameter (n(TOC) = 35, n(AOC) = 27, n(cond) = 34, n(pH) = 34) is presented in the boxplots.

The TOC concentrations of drinking water samples show a wide range from approximately 0.5 to 4.7 mg L^{-1} with a median value of 2.3 mg L^{-1}. This lies within the generally expected range of drinking water (Grohmann et al., 2011). It is evident that most of the dw_{wwe} samples are recognized at the outer edges, however, due to the small sample size, this does not necessarily indicate the general behavior of the samples. The four dw_{wwe} samples with elevated TOC concentrations are identified exactly as those that showed significantly elevated fluorescence scores of C3 and C4 in the PARAFAC analysis in Figure 4-2. This includes the TOC outlier, which was also identified as an outlier regarding C1 and C5 fluorescence scores. This could be an indication that the TOC level of drinking water mainly depends on the FDOM components C3 and C4, but possibly on components C1 and C5 as well. However, regarding dependencies between parameters a more detailed correlation analysis is implemented in the later part of this chapter.

The AOC measurements demonstrate that most of the investigated drinking water samples, excluding two dw_{wwe} samples, are in the range between 2.0 and 44.2 µg C_{eq} L^{-1} with a median value of 19.6 µg C_{eq} L^{-1}. The two outliers cannot be related to outliers recognized other parameters. Differentiation between the data distribution of dw_{wds} and dw_{wwe} samples is not observed based on the investigated waters. Since the AOC method is based on cell counting with its respective triplicated determination at several days (chapter 3.2.3), the final AOC of one particular sample can sometimes show a rather high standard deviation. These sample individual standard deviations are not visualized in Figure 4-4. The median value of all individual standard

deviations is about ± 17 %. Comparing median values of determined AOC and TOC concentrations of all water samples confirms the relatively low amount of AOC contributing to total DOM (Van der Kooij, 1990).

Z. Chen et al. (2018) collected several data of publications from the last two decades and revealed that the AOC of the most treated drinking waters (some include disinfection processes) is less than 500 µg C_{eq} L^{-1}, and over 80 % have an AOC below 200 µg C_{eq} L^{-1}. All samples studied in current work were consistent with this range. Even the estimated and stated value for biostable water ≤ 10–100 µg C_{eq} L^{-1} by Favere (2022), LeChevallier et al. (1996) and Van der Kooij (1992) was not exceeded in the water samples measured here.

As a result of the various factors affecting the AOC, a differentiation between AOC of dw_{wwe} and dw_{wds} is discussed. The AOC of dw_{wwe} is highly dependent on the particular treatment method. Thayanukul et al. (2013) observed lower AOC levels for waters treated via coagulation and sand filtration than waters treated with ozonation and chlorination. Since only waters without ozonation and chlorination were investigated in this work, this could explain the relatively low AOC levels compared to the literature. The origin of the water also plays a crucial role. In this work, treated drinking water samples came exclusively from groundwater sources. Pick et al. (2019) showed that drinking water from groundwater systems in general demonstrates rather low AOC values and can therefore be more likely be classified as biostable. AOC within the WDS is reported to be slightly decreased because of bacterial growth and associated AOC consumption in the pipe system (W. Liu et al., 2002). There may be a similar trend for the data in Figure 4-4, but there are too few data points recorded at waterworks exits to make more reliable conclusions. Furthermore, Pick et al. (2021) investigated the interaction between AOC and the biofilm of pipes within a WDS. On the one hand, AOC in water influences the biofilm growth rate, but on the other hand AOC can also be released from the biofilm. This interaction makes it difficult to define generally valid threshold for AOC and to determine biostability on the basis of AOC alone. Other parameters contributing to the biostability of the water are disinfectant residuals, temperature, pH, hydraulic conditions, residence time, stagnation, and the general condition of the WDS (Z. Chen et al., 2018).

For a more detailed analysis of AOC levels, especially regarding the particular cell amount contributing to bacterial growth, Figure 4-5 shows the bacterial growth curves of six exemplary waters. Three of them (dw_{wds-a}, dw_{wds-b}, and dw_{wds-c}) are samples taken within the WDS, while the remaining three (dw_{wwe-a}, dw_{wwe-b}, and dw_{wwe-c}) represent water samples from waterworks exits. Furthermore, the AOC data in Figure 4-5 is grouped into three quantiles (low < 33 % ≤ medium < 66 % ≤ high) to show a respective growth curve. This leads to quantile limit values of low AOC < 13 µg C_{eq} L^{-1}, and medium AOC < 25 µg C_{eq} L^{-1}, respectively.

There are slightly different growth behaviors between the tested drinking water samples. Especially samples with low AOC ranges (dw_{wds-a} and dw_{wwe-a}) show relatively steady linear growth over at least 14 days. At higher AOC levels, the typical bacterial growth phases – lag phase,

exponential growth, stationary phase and decline phase – are visible. The maximum TCC is generally reached between days 3 and 14. Regarding the LNA and HNA distribution over the growth period, HNA cells contribute the most to TCC growth. The contribution of HNA cells has already been described by Kämmler (2023), G. Liu et al. (2013) and Park et al. (2016). Park et al. (2016) and G. Liu et al. (2013) described HNA cells as more active cells, considering the positive ATP correlation, in the total bacteria population and recommended to focus on HNA monitoring to ensure biostability in WDS. Y. Wang et al. (2009) showed by isolating LNA and HNA cells, that LNA cells grow with lower specific growth rates, less changes in average cell sizes, and lower reproduction count. However, due to the small cell size (< 0.45 µm) the biomass specific production of LNA is similar to HNA cells. Thus, only monitoring HNA cells would ignore significant cell communities that can growth and therefore might affect water quality changes as well.

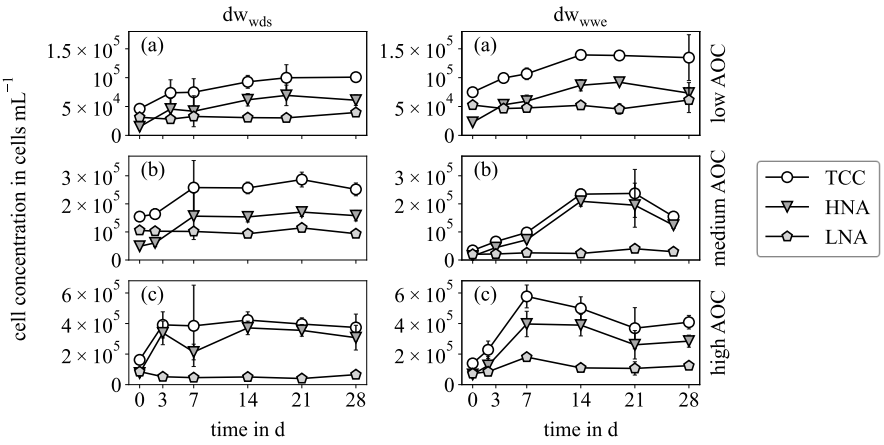

Figure 4-5: Exemplary bacterial growth curves of investigated drinking water samples at various locations of the water distribution system (dw_{wds-a}, dw_{wds-b}, and dw_{wds-c}), and at the waterworks exits (dw_{wwe-a}, dw_{wwe-b}, and dw_{wwe-c}). The AOC calculation was performed according to Hammes et al. (2006). All samples were distinguished into low, medium, and high AOC ranges with the following concentrations: dw_{wds-a} 5.5 ± 0.8, dw_{wds-b} 13.2 ± 2.3, dw_{wds-c} 26.1 ± 5.4, dw_{wwe-a} 6.5 ± 0.8, dw_{wwe-b} 20.3 ± 8.5, dw_{wwe-c} 43.8 ± 13.5 µg C_{eq} L^{-1}.

Regarding the remaining water quality parameters presented in Figure 4-4, all investigated drinking water samples were below the German drinking water limit value for conductivity, < 2790 µS cm^{-1} at 25°C, and within the pH range, $6.5 \leq pH \leq 9.5$, respectively (TrinkwV, 2023). The median value of conductivity was 424 µS cm^{-1}, showing two outlier samples, whereas median value of pH was determined at 7.7.

So far, drinking water parameters have been shown and discussed individually. All parameters of dw_{wds} and dw_{wwe} samples, except the outliers, were demonstrated to be in a similar range, which is why no distinction was made about the sampling location regarding the following analysis. In an attempt to monitor the quality of drinking water, it would be beneficial to know possible relationships between parameters. In order to calculate a correlation matrix, the values of all parameters should be available for all examined samples. Since imputations by either one constant or mean value are not expedient in this case, all insufficient samples were removed

from the data set. The remaining data set consist of n = 23 water samples which can be considered a relatively low number for determining significant correlations. However, the correlation analysis is performed here to reveal an idea about possible relation between investigated parameters, especially from PARAFAC analysis, and its references in literature. Absolute values should be viewed with caution.

The results of an analysis that was conducted using linear Pearson correlation analysis are shown in Figure 4-6. The specific correlation coefficient *r* as well as the significance coefficient *p* of the correlation matrix are attached in the appendix (Table II-1). According to Ratner (2009), a correlation coefficient of 0 indicates no relationship while 1 represents an ideal positive relationship. Values between 0 and 0.3 indicate a weak positive, between 0.3 and 0.7 a moderate positive, and between 0.7 and 1 a strong positive relationship, relatively. The same applies inversely (negative correlation) to negative values. Cross-hatched squares in Figure 4-6 mark correlation coefficients (*r*) with $p > 0.05$ which are not considered statistically significant.

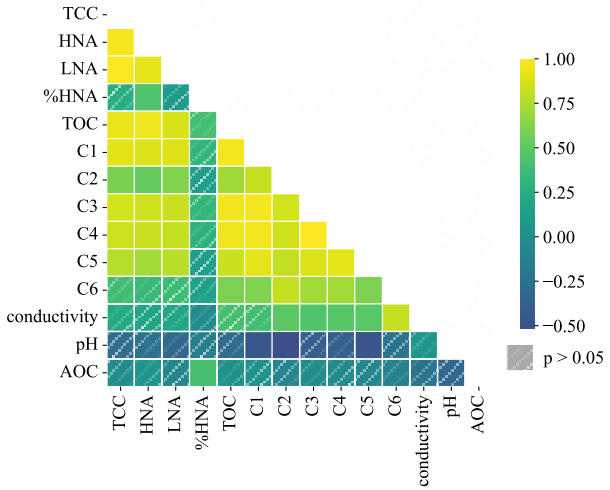

Figure 4-6: Linear correlation (Pearson method) between all measured parameters of n = 23 samples from various location at the water distribution system (dw_{wds}) and at the waterworks exits (dw_{wwe}). Color bar indicates the range of correlation coefficient *r*. Cross-hatched squares indicate insufficient linear correlations ($p > 0.05$).

The correlation analysis follows the same procedure as the individual analysis. First, inner fluorescence spectroscopy correlations between particular components and inner flow cytometry correlations are discussed, respectively. Followed by the discussion between PARAFAC fluorescence spectroscopy and flow cytometry parameters and finally the correlation of both method parameters with other parameters.

C1, C2, C3, C4, and C5 show strong positive correlations between each other, with the highest correlation between C1 and C3 ($r = 0.97$), C1 and C4 ($r = 0.97$), as well as C3 and C4 ($r = 0.99$). The rather high correlation between these three components can be explained based on their similar EEM peak positions (Figure 4-1) and their respective described substances (Table 4-1). Both C1 and C4 were characterized as fulvic acid-like components. Similar correlations between these four components were observed by Retelletti Brogi et al. (2020). C5 shows the

highest correlation with C1 (r = 0.91), C3 (r = 0.88), and C4 (r = 0.92) which explains their terrestrial character since C5 is known as a pure terrestrial compound. C6, indicated as a microbially produced protein-like compound, shows a strong positive correlation with C2 (r = 0.82), C3 (r = 0.72), and C4 (r = 0.72). C2, C3, and C4 are humic-like components described as partly microbial humic-like substances, thus suggesting a stronger correlation of these compounds with microbially produced C6 is plausible. Since C2 shows a relatively high correlation to C6 at a relatively low concentration to other components, it can be assumed that C2 contains a significant amount of microbial character. Retelletti Brogi et al. (2020) explained the correlation between C6 and C3 with its autochthonous origin. However, they worked with a five-component model, which can lead C2 properties being transferred to other components (e.g., C3).

Regarding flow cytometry, both LNA (r = 0.98) and HNA (r = 0.97) show a strong positive correlation with TCC. A correlation between flow cytometry data and pH and conductivity, observed by Gabrielli et al. (2021), could not be determined for the analyzed data, or it was insufficiently significant ($p > 0.05$).

TOC shows a strong positive correlation with TCC (r = 0.93), HNA (r = 0.94) and LNA (r = 0.88). Furthermore, TOC shows a strong positive correlation with all fluorescence components except C6. However, C1 (r = 0.96), C3 (r = 0.96), and C4 (r = 0.95), which are the components that also contribute the most to the total FDOM (Figure 4-2), exhibit the significantly higher correlation. This confirms the findings of Retelletti Brogi et al. (2020) observations of fluorescence components correlation with DOC in river water. The same components were identified with strong positive correlation to the DOC of most of the drinking water originating from different sources, investigated by Shutova et al. (2014). Furthermore, they indicated the ratios C3/C1 and C3/C6 in source waters were informative regarding DOM treatability via conventional methods. These methods include such as coagulation-sedimentation and coagulation-membrane filtration processes.

There are several publications investigating C6 fluorescence intensity and the possible correlations with it. Lapierre and del Giorgio (2014) described a strong correlation between C6 and biodegradable DOC (BDOC) and a negative correlation with photochemically degradable DOC (PDOC) in watersheds, lakes, rivers, and wetlands. According to them, it was not possible to determine whether C6 is driving BDOC as a substrate for cells or acting as a tracer for covarying biolabile molecules. Lapierre and del Giorgio (2014) found less relation between CDOM and C6 in low CDOM environments like drinking water. Thus, even if BDOC or PDOC were analyzed for drinking water investigated within this work, it would be unlikely to correlate with C6. In the absence of PARAFAC modeling, but using peak-picking, C6 or tryptophan-like components (T-peak) were generally assumed to correlate with microbial activity or contamination, especially with *E. coli*. Especially correlations between T-peak intensity and *E. coli* were stated by several authors which led to a recommendation of fluorescence spectroscopy as an additional

method to evaluate the microbial contamination of diverse waters (Bedell et al., 2020; Dapkus et al., 2023; Nowicki et al., 2019; Sorensen et al., 2018; Ward et al., 2020, 2021). However, all work showed a detection limit above the limit for drinking water (0 *E. coli* 100^{-1} mL^{-1}) which is why it cannot be transferred to the samples examined in this work with low or no *E. coli* contamination. Furthermore, the mentioned authors worked with artificial waters (Bedell et al., 2020), karst basin waters (Dapkus et al., 2023), and ground- or drinking water which have a high probability of *E. coli* contamination (Nowicki et al., 2019; Sorensen et al., 2018; Ward et al., 2020, 2021) and are thus difficult to compare to the current results. Vucinic et al. (2023) identified strong correlations between three parameters in karst springs: T-peak fluorescence, TCC, and fecal bacteria. According to the current study, the correlation between T-peak or C6 fluorescence and TCC has not been confirmed. A relatively strong correlation ($r = 0.82$) is shown between C6 and conductivity. This relation was not stated in the references mentioned. It should, however, be considered when it comes to monitoring water quality online.

Last, AOC is a parameter requiring an appropriate amount of time of at least two weeks to be measured (Figure 4-5). The development of a faster methodology for predicting drinking water bacterial stability would be desirable. Regarding the investigated parameters in this work, only one significant correlation was found between AOC and %HNA, although with a low moderate positive correlation ($r = 0.41$). The amount of HNA cells in drinking water might thus be a possible predictor to estimate the AOC level. However, rather low AOC samples are investigated here. These results are slightly contrary to S. Wang et al. (2022) who used fluorescence spectroscopy and AOC method to characterize treated coal chemical industrial wastewater. According to their findings, fluorescence intensities in T-Peak region, which is almost equivalent to C6 component, were strongly correlated with AOC intensities. Additionally, Jie et al. (2017) observed a positive correlation between the LNA growth rate (under AOC method conditions) and the respective AOC concentration. This was not tested here and would also not generate any significant added value in terms of the analysis duration.

In any case, it should be noted that the provided correlation data contain 23 drinking water samples. An extension of the data set, also in the direction of a larger dispersion, e.g., different water bodies etc. might lead to deviations of the observed correlations.

Highlights Chapter 4.2

- Analyzed drinking water samples were successfully and individually characterized by flow cytometry fingerprints (TCC, and %HNA).
- Remaining parameters (TOC, pH, and conductivity) were in the typical range of drinking water, while all samples could be described as biostable (AOC ≤ 10–100 µg C_{eq} L^{-1})
- TOC showed a strong positive correlation with TCC (r = 0.93) and HNA (r = 0.94) as well as PARAFAC components C1 (r = 0.96), C3 (r = 0.96), and C4 (r = 0.95). These PARAFAC components contributed to the majority of the total FDOM in the investigated drinking water samples.
- Strong positive correlations between the PARAFAC components C1, C3, and C4 ($r \geq 0.97$) confirmed the findings from chapter 4.1 regarding their similarities in relation to origin (terrestrial and microbial humic-like) and substance characteristics (fulvic acid-like).
- C6 did not show any correlation with flow cytometry parameters. Regarding investigated drinking water samples, it was not feasible to estimate the bacteria numbers and constitution using C6.

5. Automated Water Quality Monitoring System

Aiming at automatic monitoring of DOM and bacterial related drinking water quality in a future WDS requires several aspects of automation. Chapter 5.1 focuses on the hardware part, like online sampling and actual water sample measurement. An independent hardware setup is the basis for a reliable, coupled data evaluation process. This chapter presents partially published results (Schuster et al., 2022a; Schuster et al., 2022b). Chapter 5.2 deals with the integration of automated data analysis of PARAFAC fluorescence spectroscopy and flow cytometry. Especially PARAFAC analysis requires a solid environment to generate comparable modelling results. The online monitoring results were integrated into an online dashboard which enables future integration into real surveillance processes. Contributed codes are published on GitHub (Schuster, 2024). Chapter 5.3 validates the combined developed hardware and software setup of PARAFAC fluorescence spectroscopy and flow cytometry by laboratory experiments. Changes in drinking water quality were directly caused by adding increasing amounts of different waters to the monitored drinking water.

5.1. Laboratory Setup for Automated Water Quality Monitoring

The laboratory monitoring setup presented in Figure 3-3 was developed to allow simultaneous online monitoring of three different water samples via fluorescence spectroscopy and flow cytometry. As described in chapter 3.3.2 ‘Proof of Concept for Automated Drinking Water Quality Monitoring’ six drinking water samples with increasing concentrations of LB medium were prepared using drinking water from the same tap (dw_1). The experiments were performed in two consecutive parts. While experiment (i) monitored dw_1 only, experiments (ii)–(vi) monitored dw_1 containing different LB medium concentrations. The LB medium was spiked to the respective dw_1 samples as t = 0 h and measurements continued automatically for approximately 70 hours. With a TOC level of approximately 7 g L^{-1}, LB medium significantly affected the TOC level of dw_1 in the respective experiments (ii)–(vi) (TOC_{dw1} = 0.55 mg L^{-1}). An overview of the starting conditions of each experiment is provided in Table 5-1.

Table 5-1: Overview of experimental waters, investigated via a newly developed automatic monitoring laboratory setup. All drinking water samples were taken from the same tap (dw_1).

dw_1 experiment	LB medium in v%	TOC in mg L^{-1}	amount of protein-like FDOM in %
(i)	-	0.55	4
(ii)	0.001	0.68	6
(iii)	0.007	1.07	21
(iv)	0.014	1.52	29
(v)	0.029	2.33	47
(vi)	0.043	3.15	58

Data evaluation by fluorescence spectroscopy and flow cytometry measurements was performed subsequently. This means that PARAFAC modeling was carried out subsequently with

measurements from all six experiments (n = 96). Contrary to the PARAFAC results described in chapter 4.1, *base data* was not added to the modelling procedure. LB medium contains fluorophores which are not present in natural waters and thus, the overall six-component PARAFAC model, encouraged by the addition of *base data*, would have underestimated the actual number of components for experiments (ii)–(vi).

According to the described PARAFAC procedure in chapter 3.2.1 seven components were identified as appropriate for generating the PARAFAC model. Since this is a specific seven-component PARAFAC model, a detailed discussion of the seven components equivalent to chapter 4.1 is not provided in this context. However, a summary table is attached in the appendix (Table II-2) and corresponding discussion can be found in Schuster et al. (2022a).

The presence of LB medium containing oligopeptides and amino acids as its predominant constituents led to more diverse FDOM fingerprints. In general, three protein-like components (P1, P2, P3) and four humic-like components (H1, H2, H3, H4) were identified for the seven-component PARAFAC model. To begin with the protein-like components, P1 (λ_{ex} = 280 nm, λ_{em} = 337 nm) indicates a tryptophan-like component which is equivalent to C6 of the overall six-component PARAFAC model (chapter 4.1). P2 (λ_{ex} = 280 nm, λ_{em} = 350 nm) describes a pure tryptophan component (Wünsch et al., 2015). P3 (λ_{ex} = 275 nm, λ_{em} = 302 nm) describes a tyrosine-like component (Paerl et al., 2020). The humic-like components (H1: λ_{ex} = 252/303 nm, λ_{em} = 403 nm, H2: λ_{ex} = 252/393 nm, λ_{em} = 463 nm, H3: λ_{ex} = 261/357 nm, λ_{em} = 453 nm, H4: λ_{ex} = 252 nm, λ_{em} = 526 nm) resemble a mixture of the five humic-like components from the overall six-component PRAFAC model in chapter 4.1.

Regarding the total FDOM of each sample, the addition of LB medium to dw_1 has more of an effect on protein-like compounds. Therefore, the amount of protein-like fluorescence in the total fluorescence is presented in Table 5-1 as well. Comparing the proportion of protein-like components with drinking waters that were investigated in chapter 4.1 (4–9 %), it becomes clear that only (i) with 4 % and (ii) with 6 % are in drinking water related ranges.

The fluorescence score of each protein-like component in all experiments over the online monitoring duration of 70 hours is presented in Figure 5-1. On the contrary, humic-like components did not show any significant effects, neither by adding LB medium nor by monitoring over the experimental duration. This is why these monitoring results are not discussed in the following section but shown in the appendix (Figure II-2).

Due to the fluorescence score at the beginning of the experiments (t < 25 h), there is an apparent positive correlation between the increasing LB medium concentration in dw_1 (i) → (vi) and the respective fluorescence scores of components P2 and P3. In contrast to that, the fluorescence score of P1 is not significantly impacted by LB medium concentration. Pure dw_1 (i) does not show any occurrence of the sole tryptophan component P2. This is one reason why this component was not found in natural drinking waters by the overall six-component PARAFAC

model in chapter 4.1. However, pure dw_1 (i) even shows the appearance of the tyrosine-like component P3. It is possible that due to the low proportion of P3 in natural drinking water, components P1 and P3 are combined into one component C6 in the overall six-component PARAFAC model. The different distinction might be the reason that most drinking water PARAFAC characterization related articles only identify a single protein-like component (Moona et al., 2021; Wells et al., 2022).

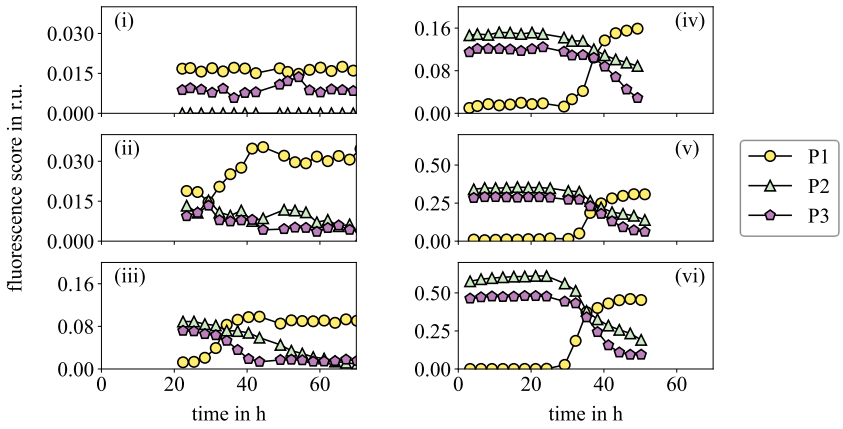

Figure 5-1: Time course of the fluorescence score of the three protein-like components P1, P2, and P3 in six different (i)–(vi) online monitored experiments. The components obtain the following excitation and emission maxima: P1 λ_{ex} = 280 nm, λ_{em} = 337 nm, P2 λ_{ex} = 280 nm, λ_{em} = 350 nm, and P3 λ_{ex} = 275 nm, λ_{em} = 302 nm. The conditions of the respective experiments are presented in Table 3-1 and Table 5-1.

Considering the curve behavior of the fluorescence scores of all six experiments, there is a similar tendency for the dw_1 samples with added LB medium (ii)–(vi). While an exponential increase in the fluorescence scores of P1 is observed between 25 and 40 hours, the intensities of components P2 and P3 decrease at the same time. Thus, the maximum fluorescence score of P1 depends on the initial LB medium concentration. Furthermore, it seems that the maximum fluorescence score of P1 is reached before the corresponding P2 and P3 scores reach a minimum. Based on these stand-alone results, a change in the FDOM characteristics within the water samples can be noted over the investigated period. In order to evaluate the results qualitatively, the results of the flow cytometry analysis conducted in parallel should be considered and are depicted in Figure 5-2.

The TCC at the starting phase (t < 25 h) of each experiment was relatively stable for all experiments. It ranged between 2 and 3×10^4 cells mL^{-1} with an almost equal amount of HNA and LNA cells. Compared to investigated drinking waters (Figure 4-3) the TCC of dw_1 tends to be in the lower range and the %HNA tends to be in the upper range, respectively. However, a slight increasing TCC trend could be identified for the initial 25 hours (*I*). In the subsequent period between 25 and 40 hours (*II*), the TCC shows an exponentially rise behavior for experiments (ii)–(vi), and a slight increasing behavior for experiment (i). After approximately 40 hours, the TCC is almost stable (*III*). According to the definitions of Cypionka (2010) and Madigan et al. (2015) these states are perfect representation of the three phases of bacterial growth:

(I) *Lag phase*: Metabolically active bacteria but no intensive growth yet.

(II) *Exponential growth*: Due to high nutrient concentration, bacteria are under favorable conditions for constantly repeating exponential growth.

(III) *Stationary phase*: The nutrient sources are exhausted, which inhibits the growth of bacteria.

According to Sezonov et al. (2007), LB medium is a carbon-limiting growth medium. Thus, the amount of LB medium is mainly responsible for the growth behavior of drinking water bacteria. This is reflected partly by the final TCC (*III*), which is higher with increasing initial LB medium concentration. Another finding is that the maximum growth rate of bacteria increases with a higher concentration of LB medium (*II*). Besides monitoring TCC, Figure 5-2 indicates significant changes in the distribution of HNA and LNA cells during bacterial growth. While the HNA amount stays rather constant with a slight tendency to rise during the lag phase (*I*), it shifts significantly towards HNA cells during exponential growth (*II*). The more active contribution of HNA cells was already described regarding the AOC determination of growth curves in chapter 4.2. Moreover, the behavior confirms the former recommended preferred monitoring of HNA cells within a WDS to observe bacterial growth (G. Liu et al., 2013; Park et al., 2016).

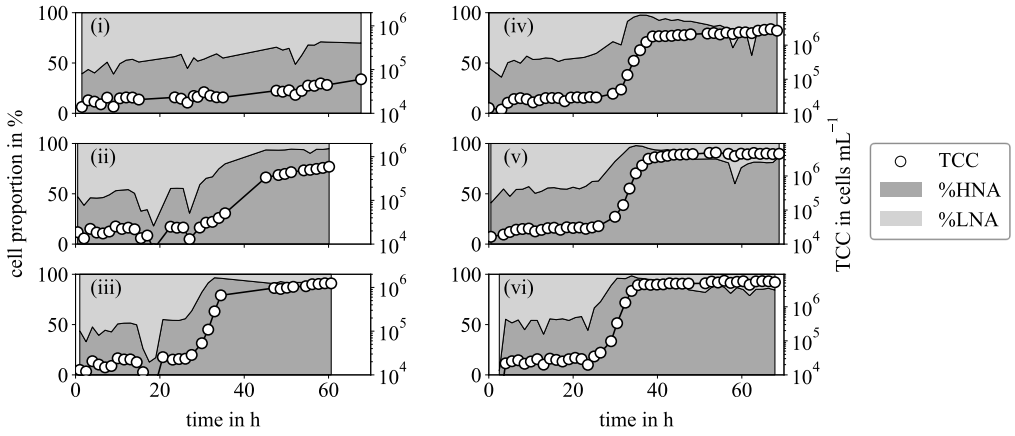

Figure 5-2: Time course of the flow cytometry analysis (TCC, %HNA, and %LNA) of six different online monitored experiments (i)–(vi). The conditions of the respective experiments are demonstrated in Table 3-1 and Table 5-1.

A further HNA cell discrimination was carried out in Schuster et al. (2022a) by adding another classification into HNA_{min} and HNA_{max} cells. Therefore, the flow cytometry HNA gate is divided into two areas of almost equal size. Just before the respective exponential growth phases (between 15 and 25 hours), a remarkable shift towards cells with a very high amount of nucleic acid (HNA_{max}) was observed. This also highlights the opportunity for further detailed flow cytometry fingerprinting and can be considered for future online monitoring. The corresponding results are attached to the appendix (Figure II-3).

Comparing fluorescence spectroscopy (Figure 5-1) and flow cytometry (Figure 5-2) results revealed a strong correlation between component P1 and TCC. However, it must be noted that the correlation might be affected by the strong influence of the provided organic nutrient rich

LB medium. Drinking water samples without added organic nutrients (chapter 4) did not show a correlation between the P1 equivalent C6 and TCC. This might be an indicator that P1/C6 is not directly correlated to TCC, but rather to general active cells in terms of cell growth and cell division. The correlation between this activity and tryptophan-like component P1 confirms the statements of several publications which only used the peak-picking method described in chapter 2.2.2 (Bedell et al., 2020; Dapkus et al., 2023; Nowicki et al., 2019; Sorensen et al., 2018; Ward et al., 2020, 2021). Moreover, since P1 and P3 are represented as one component in the overall six-component PARAFAC model, it could be possible that an increasing behavior of P1 is equalized with a decreasing behavior of P3. A correlation between P1/C6 and %HNA in drinking water could also not be observed in chapter 4.2. Nevertheless, eventual relations should be considered for the following experiments.

A brief mention should be made here of why the application of the seven-component PARAFAC model was not sufficient to analyze drinking water without injected LB medium. It was shown that the fluorescence scores of P2 and P3 provided information regarding organic nutrient concentrations, adjusted by LB medium. These concentrations could be correlated to the cell growth and thus TCC and P1. However, since P2 was not measured in sole dw_1 at all, it is very likely that this component is not part of drinking water in general. It is also likely that due to the spectral similarities of P1 and P3 and the low fluorescence intensities in drinking water compared to humic-like substances, both can only be identified as a combined component. Thus, as long as drinking water does not vary significantly from drinking waters investigated in this study (chapter 4.1), PARAFAC analysis might only identify one protein-like component.

In total, parallel water quality monitoring via PARAFAC fluorescence spectroscopy and flow cytometry was proven to be a promising combination. The general adaption of the setup generates added values regarding the future monitoring process, although the data analysis was performed subsequently to experimental measurements. The appearance of monitored outliers further requires individual parameter baselines. These baselines would enable the definition of an expected range for each parameter value. Regarding long-term monitoring of drinking water, a scenario of a sudden deviation of only one parameter value might be an indication of an evolving water quality change. These aspects are addressed in the following chapters.

Highlights Chapter 5.1

- Hardware for parallel and continuous fluorescence spectroscopy and flow cytometry monitoring was successfully introduced and showed the potential of high-resolution water quality monitoring.
- Bacterial growth was observed via flow cytometry. It could be correlated with the manually added nutrient concentration which was possible to characterize and quantify via PARAFAC fluorescence spectroscopy.
- Three displayed protein-like PARAFAC components (P1, P2, P3) of a seven-component PARAFAC model revealed insights into nutrient conversion and bacterial cell growth.
- The seven-component PARAFAC model is probably not applicable to analyzing drinking water without manually added organic nutrients. There might be, however, a limitation to the overall six-component PARAFAC model in regards to differentiation of protein-like components (equalization of tryptophan- and tyrosine-like components during the growth of bacteria).
- Besides TCC, %HNA might indicate the current state of bacteria and could function as an indicator of sudden bacterial growth.

5.2. Software Tools for Automated Water Quality Data Analysis

After the successful development of the laboratory hardware environment, this chapter deals with the challenge of data analysis of PARAFAC fluorescence spectroscopy and flow cytometry in near real-time. So far, data analysis has been performed after experiments were completed with the full generated data set. With this novel approach, the data is processed immediately and automatically after measurement according to Figure 3-3. The processes which are aimed to automating in this chapter are displayed in Figure 5-3. Generally, both PARAFAC fluorescence spectroscopy (I) and flow cytometry (II) data analysis, are processed in separate procedures and subsequently merged into the data visualization process (III). The coding scripts of the respective procedures are open access and available via GitHub (Schuster, 2024) "https://github.com/waterjonas/drinking-water-quality-online-monitoring/tree/main". The *'README.md'* file contains all required information for the installation and execution process.

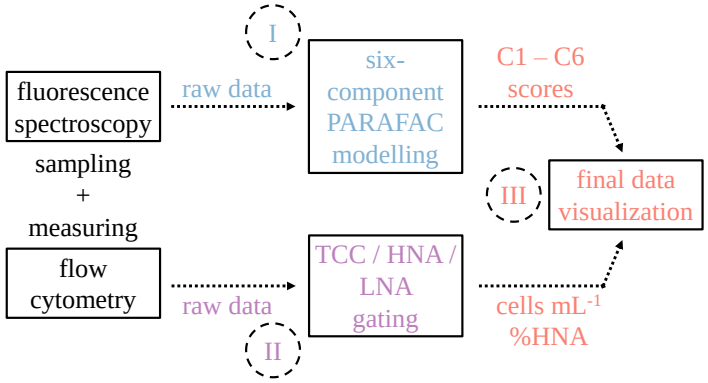

Figure 5-3: Schematic overview of automated monitoring processes. Processing Fluorescence Spectroscopy Raw Data (I). Processing Flow Cytometry Raw Data (II). Data Visualization via Dashboard (III).

5.2.1. Processing Fluorescence Spectroscopy Raw Data (I)

Fluorescence spectroscopy measurements are performed using the Aqualog fluorometer in combination with the Aqualog-Sipper automation unit as described in chapter 3.2.1. The raw data of the measured sample is stored in a selected folder. The raw data of sample absorbance, blank fluorescence and sample fluorescence are saved in respective '*-.dat'* files for each sample including the time stamp.

Here, the general idea of data analysis is to characterize every continuously measured water sample by the same fluorescence components. As it was explained in chapter 4.1, the overall six-component PARAFAC model is a powerful tool for DOM characterization of drinking water. The seven-component PARAFAC model, presented in chapter 5.1, is not applicable due to the overdetermination of the water as a whole. The challenge is to generate a model that is consistent and thus identifies the same spectral properties of the same six components each time it is created. A consistent and validated data set regarding the overall six-component PARAFAC model describes the *base data*. It contains 72 different real waters and model solutions which cover a wide range of different DOM constitutions. An overview of the *base data* samples and their origins is provided in appendix (Table I-II). In a continuous monitoring process,

the most recent Aqualog fluorometer reading is added to the *base data*, any previous readings are removed and the PARAFAC model is generated based on these 73 samples (*base data* + current sample). This procedure repeats automatically after each measurement. Meanwhile the *base data* remain unchanged. Due to the high volume of consistent base data compared to the one new sample, the PARAFAC model assures reproducibility of the same six components regardless of the added water sample.

To realize this approach, two MATLAB scripts were further developed, '*I_I_start_control.m'* and '*I_II_PARAFAC.m'*, which are available online (Schuster, 2024). Script '*I_I_start_control.m'* controls the folder where raw data is automatically saved after each measurement. As soon as three new files – sample absorbance, blank fluorescence, and sample fluorescence – appear, the function `eventhandlerChanged` is called which activates '*I_II_PARAFAC.m'*. This script performs the proper PARAFAC analysis using the established drEEM toolbox (Murphy et al., 2013). In the first step between code lines 7 to 26 in '*I_II_PARAFAC.m'* (Schuster, 2024), the size of the folder containing raw sample data is checked. Whenever a new measurement is conducted, the older readings are removed from this folder. This ensures the exact amount of 73 samples (*base data* + current sample) required for PARAFAC modelling. The remaining PARAFAC modelling process is mainly based on the procedure introduced in chapter 3.2.1, according to Murphy et al. (2013) and Wünsch (2024), containing minimal changes concerning its automatization.

The process is exactly like the manual process until it comes to checking file alignment, regarding the respective absorbance, blank and sample file. This is normally done by the user. In general, the three files of one sample match except for the last three digits. Thus, the remaining digits of the three files are checked for matching. The recognized number of samples is compared with the actual number of files. Within the script, it is realized using the `try` and `catch` instructions (code lines 57 to 61 in '*I_II_PARAFAC.m'* (Schuster, 2024)). Under `try`, two conditions are checked via the `isequal` function and if one fails, `catch` responds with an error message to the user. The remaining steps were executed according to the settings described in chapter 3.2.1. Completing these steps does not require user input or confirmation. The final modelling is performed via the `randinitanal` function as shown in Figure 5-4.

```
128 model = randinitanal(Xpre,6,'starts',10,'constraints','nonnegativity','con-
vgcrit',1e-6);
```

Figure 5-4: Line 128 of MATLAB script '*I_II_PARAFAC.m'* showing PARAFAC modelling function `randinitanal` including its parameters.

In addition to the automation procedure, one objective of this work was to analyze drinking water quality as rapidly as possible in near real-time. Apart from the actual measurement duration (approx. 7 min), a time-consuming step is the calculation duration of the PARAFAC model using the `randinitanal` function which depends highly on computing power. Within this function, the duration is mainly affected by both the number of random initializations and the

convergence criterion. The former showed a much higher influence. According to Murphy et al. (2013), it is conceivable that a random solution is calculated due to an insufficient number of initialized models. This is similar if the convergence criterion is too large. If the calculated model converges with too few iterations, the result will be less accurate. The accuracy and duration of the here established six-component PARAFAC modelling was tested for convergence criteria 1e-4, 1e-5, 1e-6, 1e-7, and 1e-8 and for numbers of random initializations 5, 10, 20, 30, 50, and 100. The combination of 10 random initializations with a convergence criterion of 1e-6 showed rapid calculation time as well as adequate accuracy. The calculation time was reduced to 9 % of the maximum possible calculation time. This is achieved by a convergence criterion of 1e-8 and random initializations of 100. Due to no significant changes observed by varying the convergence criterion below 1e-6, this was chosen as the appropriate criterion (Wünsch, 2024). The relative time dependency and an exemplary accuracy dependency are shown in the appendix (Table II-3 and Figure II-4).

The PARAFAC step of model validation, described as *(IV) Model Validation* in chapter 2.2.3 and 3.2.1, was not performed within the automated process. If the model cannot be validated, this would very likely affect the position of the excitation and emission peaks as well. Since this aspect will be verified in the final data visualization part 'Data Visualization via Dashboard (III)', a successfully validated model is assumed here.

As soon as the PARAFAC model is successfully generated, the fluorescence scores (in Raman units r.u.) of all six components of each sample are saved in the Excel file '*model_data_6.xlsx'*.

5.2.2. Processing Flow Cytometry Raw Data (II)

Flow cytometry measurements were performed using the flow cytometer CyFlow^{TM} Cube 6 and the automation unit OC-300 (chapter 3.2.2). The raw data of one sample is saved in particular flow cytometric '*-.fcs'* files for each sample including a time stamp. The raw data is saved in subfolders according to the integrated measurement software settings (cyOn-300 v.3.12). Each subfolder can contain a maximum of 48 '*-.fcs'* files before another subfolder is created.

The data analysis is performed via the Python script '*II_fcm.py'* and is available online (Schuster, 2024). First the user is asked to enter information about the sampling procedure, e.g., sampling mode and dilution factor. Then, the user has to select the folder or subfolder structure of the raw data, as well as the time duration between two measurements, and the name and location of the received results. Regarding 'Data Visualization via Dashboard (III)', the final results are saved in an excel file named '*fcm_results.xlsx'*. After entering all parameters, the user must wait until the first online measurement is performed by the devices to start the script after confirmation.

Regarding accurate and consistent counting of cells and discrimination between HNA and LNA cells, the gate parameters are defined in the code lines between 104 and 117. This is shown in

Figure 5-5. These boundaries are based on the evaluation process of analyzed samples, described in chapter 4.2.

```
104 # TCC gate
105 x2 = 50
106 x4 = 10000
107 x5 = 50000
108 y1 = 10
109 y2 = 60
110 y5 = 10000
111
112 # HNA/LNA boundary on x-axis,
value is minimum in histogram between
peak regions for mineral water (Evian)
113 x_LNA = 315.0
114
115 # lienar diagonal equation TCC gate
116 m = (y5 - y2)/(x4 - x2)
117 b = y2 - (m * x2)
```

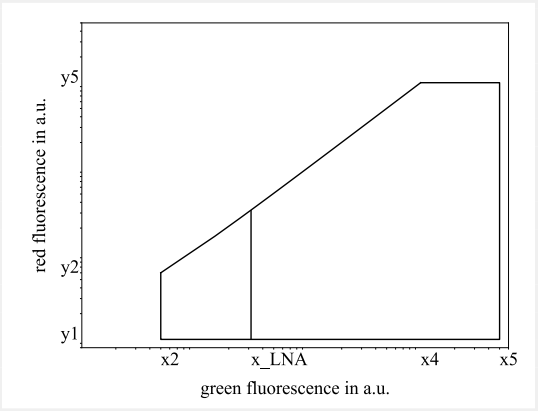

Figure 5-5: Lines 104 to 117 of the Python script '*II_fcm.py'* defining the TCC gate and LNA/HNA boundary. The respective gate is plotted including the boundaries for TCC gate (`x2`, `x4`, `x5`, `y1`, `y2`, `y5`) and LNA/HNA gate (`x_LNA`).

These boundaries are flexible and must be readjusted when using a different flow cytometer. Regarding automated data analysis, the samples ('*-.fcs'* files) are read in by applying the open source toolbox FlowKit (White et al., 2021). With sample information, the TCC, HNA, LNA in cells mL^{-1} as well as the %HNA is calculated. These results are exported, including information about the file name, sample number, data, and time as an additional row to the excel file '*fcm_results.xlsx'*. Furthermore, the flow cytometry fingerprint of the current sample is plotted inline to recognize any significant changes. The script then pauses for the set time between two measurements before restarting with another reading.

5.2.3. Data Visualization via Dashboard (III)

The last part of the automation step merges the previously analyzed PARAFAC fluorescence spectroscopy and flow cytometry results. It visualizes them in an integrative dashboard that can be accessed online. The coding script *'III_app_online_dashboard.py'* is available online (Schuster, 2024). A schematic overview of the process is displayed in Figure 5-6.

The script starts by asking the user about general information like the number of samples analyzed simultaneously, the location of the five Excel files (red colored boxes in Figure 5-6) and the update interval of the final dashboard. It is recommended to choose an update interval > 20 s to guarantee appropriate computing time. Last, the script displays a website to access the dashboard.

The file *'fingerprints_model6.xlsx'* provides the excitation and emission spectra of the six components based on the evaluation of *base data* (chapter 3.1). These are responsible for considering the consistency of the PARAFAC model (chapter 4.1). This file remains constant during

the monitoring process. *'model_data_6.xlsx'* is updated automatically by 'Processing Fluorescence Spectroscopy Raw Data (I)' and contains the fluorescence score of the 72 *base data* samples plus the current sample while *'scores.xlsx'* stores all samples measured up to this point including the *base data* samples. The latter file is used to plot the data. Regarding the flow cytometry part, *'fcm_results.xlsx'* interacts with the earlier described steps of 'Processing Flow Cytometry Raw Data (II)' whereas the final plotting information is gathered in *'fcm_scores.xlsx'*. At the start of every update interval, the bottom rows of both *'model_data_6.xlsx'* and *'fcm_results.xlsx'* are compared with the bottom rows of *'scores.xlsx'* and *'fcm_scores.xlsx'*, respectively. A deviation indicates a new reading which immediately starts the following process.

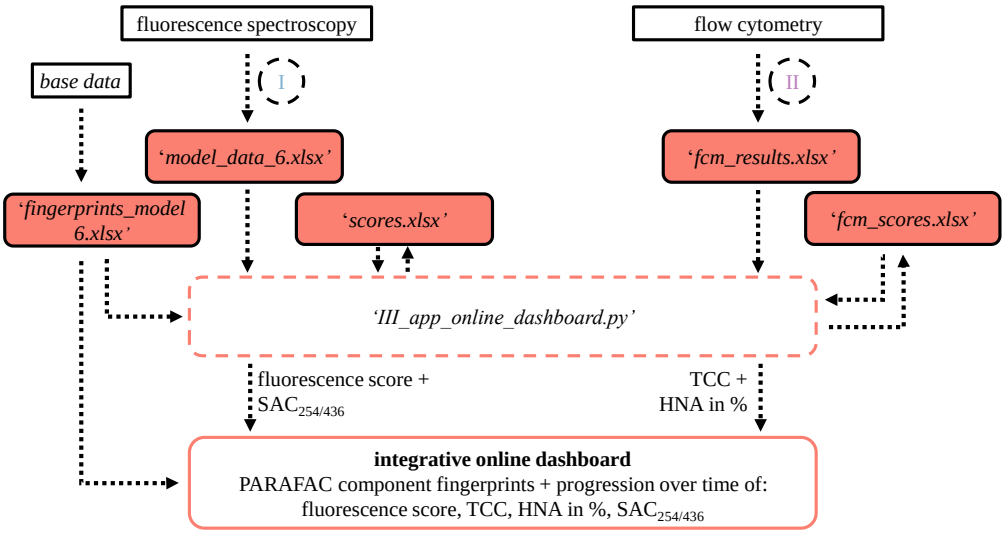

Figure 5-6: Schematic overview of automated monitoring process part 'Data Visualization via Dashboard (III)'.

Since it is not guaranteed that the column positions of C1–C6 are the same for *'model_data_6.*xlsx' and *'fingerprints_model6.xlsx'* the alignment of the components in both Excel files must be reviewed after each measurement. For this purpose, the positions of the excitation and emission maxima in the column of each component are compared with each other (code lines 259–655 in *'III_app_online_dashboard.py'*). According to these comparisons, a revised version of *'model_data_6.xlsx'* is saved containing the correct component alignments. Figure 5-7 shows the relevant excitation and emission peak ranges for each component. Due to an overlap of their 1st excitation peaks in the range between 250 and 280 nm, some components are compared by their 2nd excitation peak (C1, C3, C4). With marked excitation and emission ranges, each component can be individually identified.

After completing the comparison and rewriting process of *'model_data_6.xlsx'*, the fluorescence scores of the latest sample including its timestamp and sample number are added to *'scores.xlsx'*. This file is further processed to plot the data. Besides displaying the absolute values of the fluorescence scores, the baseline region can also be seen. Baseline properties will

be discussed in detail in the following chapters. The number of events to be considered for the baseline calculation can be selected individually (code lines 794–884 in *'III_app_online_dashboard.py'*).

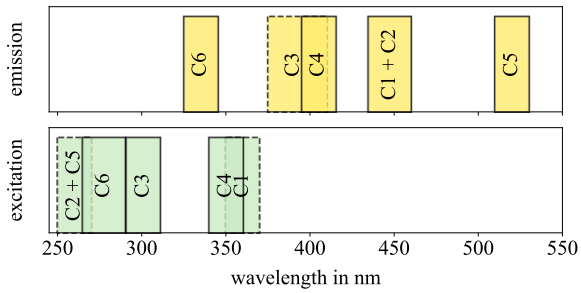

Figure 5-7: Excitation and emission peak regions for C1–C6 of the overall six-component PARAFAC model.

The procedure is similar regarding the evaluation of flow cytometry plots (code lines 1892–1943 in *'III_app_online_dashboard.py'*) even though the data provided from Excel file *'fcm_results.xlsx'* does not require further adjustments.

The script also enables plotting SAC values measured by the Aqualog fluorometer. While there is no focus on this parameter, it shows the flexibility of the dashboard when it comes to considering the integration of additional parameters.

Due to the interactive character of the dashboard, an appropriate illustration of its features cannot be provided in this work. However, the four following screenshots give an idea of the dashboards surface (Figure 5-8). For a detailed overview and evaluation of an entire data set it is recommended to follow the published *'README.md'* on GitHub (Schuster, 2024). This contains an exemplary data set which can be applied on a trial-wise.

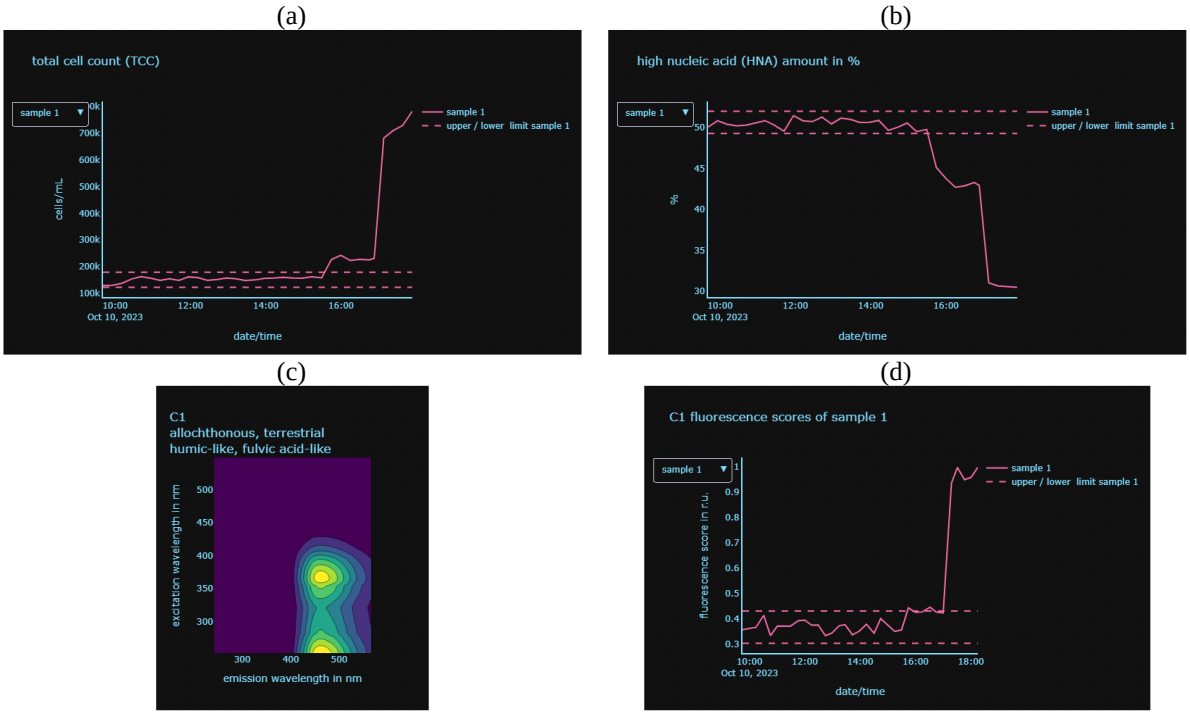

Figure 5-8: Exemplary screenshots of the interactive online dashboard to visualize fluorescence spectroscopy and flow cytometry data. (a) total cell count (TCC), (b) high nucleic acid amount of the TCC (%HNA), (c) fluorescence spectroscopy fingerprint of PARAFAC component C1 (overall six-component PARAFAC model), and (d) fluorescence score of PARAFAC component C1.

Highlights Chapter 5.2

- Successful development of automated PARAFAC fluorescence spectroscopy and flow cytometry data evaluation and visualization software tools through combined MATLAB and Python environments.
- The application of the consistent overall six-component PARAFAC model enables repetitive surveillance of the same six components of the investigated water and allows appropriate comparability.
- Detailed DOM and bacterial fingerprinting of continuously monitored water samples by merging data from both methods and visualizing them into a web-based integrative dashboard.
- The interactive dashboard allows a fast (< 15 min) recognition of outliers and evolving events in water quality.
- The integrative dashboard enables the opportunity to be expanded or adjusted, e.g., by adding different parameters or applying dynamic baselines.

5.3. Detection of Water Quality Changes in an Automated Monitoring System

Experiments presented in this chapter were conducted to evaluate and highlight the possibilities and challenges of combining both methods PARAFAC fluorescence spectroscopy and flow cytometry in one automated monitoring setup as earlier described in chapters 5.1 and 5.2. This gives rise to the following research questions:

- Is it possible to detect rapidly evolving changes in drinking water quality of continuous monitored system? Here this change is induced by spiking drinking water (dw_2) with increasing concentrations (0.1, 1, and 10 v%) of varying waters: wastewater effluent (wwe_1), rainwater (rw_1), and drinking water (dw_3).
- How sensitive is each monitoring parameter?

The experimental procedure addressing these research questions was described in chapter's 3.3.2 subsection 'Detection of Crucial Changes in Drinking Water Quality via Automated Monitoring'. Table 5-2 shows the water quality parameter values of both water in each experiment. The fluorescence scores were calculated applying the introduced overall six-component PARAFAC model.

Table 5-2: Water quality parameters values (PARAFAC fluorescence spectroscopy, flow cytometry, and TOC) of waters investigated in the online monitoring laboratory experiments. (i), (ii), and (iii) represent the three different experiments of an induced change in drinking water quality by injecting different amounts of an external waters (wastewater effluent wwe_1, rainwater rw_1, and drinking water dw_3) to the continuously monitored drinking water (dw_2).

dw_2 experiment	water	TCC in cells mL^{-1}	%HNA in %	C1 in r.u.	C2 in r.u.	C3 in r.u.	C4 in r.u.	C5 in r.u.	C6 in r.u.	TOC in $mg\ L^{-1}$
(vii)	dw_2	$1.54 \times 10^5 \pm 4.7 \times 10^3$	51 ± 1	0.365 ± 0.018	0.535 ± 0.026	0.435 ± 0.020	0.325 ± 0.013	0.156 ± 0.006	0.089 ± 0.011	2.4
	wwe_1	$8.21 \times 10^6 \pm 7.2 \times 10^5$	23 ± 0	4.636	0.000	4.512	3.457	1.751	2.613	14.7
(viii)	dw_2	$3.25 \times 10^5 \pm 7.6 \times 10^3$	79 ± 0	0.386 ± 0.004	0.547 ± 0.007	0.458 ± 0.007	0.345 ± 0.004	0.150 ± 0.003	0.108 ± 0.001	2.4
	rw_1	$1.25 \times 10^6 \pm 1.2 \times 10^5$	73 ± 3	0.062	0.045	0.074	0.040	0.030	0.119	0.7
(ix)	dw_2	$2.15 \times 10^5 \pm 2.5 \times 10^3$	61 ± 1	0.394 ± 0.005	0.576 ± 0.004	0.469 ± 0.004	0.350 ± 0.005	0.161 ± 0.003	0.108 ± 0.002	2.4
	dw_3	$5.87 \times 10^5 \pm 8.2 \times 10^3$	70 ± 1	0.537	0.707	0.809	0.600	0.180	0.224	4.7

All parameter values of dw_2 represent the mean values of the first approximately 15 measurements of each experiment. Three times the standard deviation (Equation 2-5) was calculated to define the baseline which is applied in experiments to define whether a measurement can be

considered outlier/event or not. Differently, the injected waters were only measured once (fluorescence spectroscopy) and three times (flow cytometry).

PARAFAC component scores of all dw_2 samples are within the typical range of drinking water (Figure 4-2). However, flow cytometry parameters of dw_2 are rather high for experiment (ix) and significantly high for experiment (viii) compared to the general range of drinking water (Figure 4-3). Despite the fact that the experiments were carried out with dw_2 from the same sampling day, the experiments were performed several days apart. This could have led to bacterial growth and a shift towards HNA communities within the water.

Specifically, TCC in dw_3 is extremely high, even exceeding the recommended drinking water TCC limit (Van Nevel et al., 2017b) (5×10^5 cells mL^{-1}). Fluorescence scores of dw_3 are in the expected range, although C2, C3, C4, and C6 show comparatively elevated fluorescence scores. The high amount of FDOM can also be confirmed by an increase in TOC concentration, compared to dw_2.

Wastewater effluent and rainwater samples were not investigated in this work yet. Hence, a brief discussion is provided in the following part.

There is considerable reservation to be made about the absolute fluorescence scores of wwe_1 and rw_1, which were generated using the overall six-component PARAFAC model. Since their water matrices significantly differ from samples containing *base data*, accurate fluorescence scores determination is challenging. According to Stedmon et al. (2011) a sample which is substantial different from the remaining samples of the data set affects the PARAFAC modelling part regarding the relative fluorescence scores but also the position of the excitation and emission spectra. This is a reason why it is difficult to develop a PARAFAC model which is able to work for several waters, e.g., drinking water, groundwater, rainwater, and wastewater, at the same time. The PARAFAC fluorescence scores of the six components of wwe_1 and rw_1 are therefore only approximate values.

Wastewater effluent (wwe_1) was sampled after a three-stage treatment at a municipal wastewater treatment plant). The FDOM constitution and general water quality of wastewater effluent mainly depends on the respective treatment steps (Carstea et al., 2018). According to Henderson et al. (2009), natural waters, e.g., surface waters, groundwaters etc., are predominated by refractory organic substances peak A and C which are reflected in this work by the overall six-component PARAFAC components C1, C2, and C5 while wastewater is predominant by labile organic substances peak T and B equivalent to components C3, C4, and C6. According to wwe_1, labile organics (C3, C4, and C6) account for approximately 60% of total FDOM. In comparison, the fluorescence scores of drinking water dw_2 C3, C4, and C6 represent approximately 45 % of total FDOM, respectively. Furthermore, Henderson et al. (2009) summarized studies highlighting a strong positive correlation between peak T, here C6, and the biochemical oxygen demand (BOD), which is a parameter representing the bioavailability of

water constituents in wastewater. C6 was further found by H. Yu et al. (2013) to be associated with anthropogenic activities. C3 and C4 were classified as wastewater specific components (Murphy et al., 2011; G.-H. Yu et al., 2010). Murphy et al. (2011) described both components as essential parts of treated wastewater, while C4 is further recommended to use as a wastewater or nutrient enrichment tracer. Moona et al. (2021) demonstrated that C2 can be used as an indicator for bioavailable DOM and thus is successfully reduced or removed in wastewater treatment plants through, e.g., biofiltration. Thereby, the removal efficiency of C2 or labile DOM fractions is significantly affected by different parameters like raw wastewater condition (Moona et al., 2021), filtration contact time (Hallé et al., 2009) but also the filter media, e.g., granular activated carbon (Peleato et al., 2016). Regarding investigated wwe_1 it can be assumed that C2 was almost completely removed and is no longer present, or at least at very low scores so that it cannot be identified.

Wastewater effluent samples generally contain high amounts of TCC (Figure 2-9) (Berney et al., 2008; Ho et al., 2020; Kötzsch & Sinreich, 2014). Here, wwe_1 contains approximately 8.21×10^6 cells mL^{-1}, whereas a minority (approximately 23 %) remain of HNA cells. In comparison to investigated drinking waters (Figure 4-3) both are extraordinarily high and low, respectively. The TOC of the wastewater effluent lies at 14.6 mg L^{-1} which significantly exceeds the TOC of investigated drinking waters (Figure 4-4) but lies within the expected range compared to literature (Figure 2-2) (Dubber & Gray, 2010; Kördel et al., 1997).

Rainwater (rw_1) investigated in this work was taken from a rooftop water tank after a precipitation event. As known from literature, the quality of rooftop runoff is generally dependent on different parameters such as local ones like the rooftop material and the duration since the last precipitation but also the local climate, and atmosphere pollution (Farreny et al., 2011; Förster, 1999). With the exception of C6 which demonstrates approximately 32 % of the total FDOM, all fluorescence scores of rw_1 are in relatively low ranges compared to investigated drinking waters in this work (Figure 4-2). This confirms the findings of Niloy et al. (2021) who identified the tryptophan-like component as the main constituent of all investigated rainwater samples regardless of season and origin. The TCC of rw_1 is approximately 1.25×10^6 cells mL^{-1} which lies in the range of surface waters (Figure 2-9) (Berney et al., 2008; Ho et al., 2020; Kötzsch & Sinreich, 2014) and can directly be compared to rainwater studied by Favere et al. (2021). The TOC of rw_1 is 0.6 mg L^{-1} which is quite low compared to the TOC of dw_2 (2.4 mg L^{-1}) but within the range reported by Li (2017) (Figure 2-2).

To conclude, wwe_1, rw_1 and dw_3 characterize waters that differ from dw_2. They are thus suitable to determine the general applicability and the sensitivity of the parameters regarding an induced water quality change event. The following section shows and discusses the results of the respective experiments. Although those were performed via the developed software tools, including the dashboard visualization described in chapter 5.2, the figures visualized in this section

were generated individually subsequent to the experiments. Schuster (2024) provides an exemplary online available data set for the virtual simulation of one spiking experiment.

5.3.1. Spiking with Wastewater Effluent – (vii)

Wastewater effluent was chosen as a spike medium due to its possible hazardous impact on drinking water quality. According to Efstratiou et al. (2017) and Moreira and Bondelind (2017) a majority of cases with a harmful impact on drinking water consumers were related to protozoan parasites which can be part of wastewater streams. A selected case regarding the relevance of spiking experiments is reported by Moon et al. (2013). Caused by two pipe leakages in nearby drinking water and sewage pipes, a small amount of wastewater could enter the drinking water pipe. This could contaminate the system.

Figure 5-9 shows the results of experiment (vii), highlighting fluorescence spectroscopy (a) – fluorescence scores of the respective overall six-component PARAFAC model components – and flow cytometry (b) – TCC and %HNA – monitoring results over a period of approximately 8.5 hours. The monitored dw_2 was spiked with wwe_1 at three different times. The period until the 1st spike was used to generate the baseline of dw_2 including its upper and lower limits by three times the standard deviation (Equation 2-5). The corresponding spiking points in time and conditions were following: 1st spiking event after 4.4 h with 0.1 v% of wwe_1, 2nd spiking event after 5.9 h with 1 v% of wwe_1, and 3rd spiking event after 7.3 h with 10 v% of wwe_1.

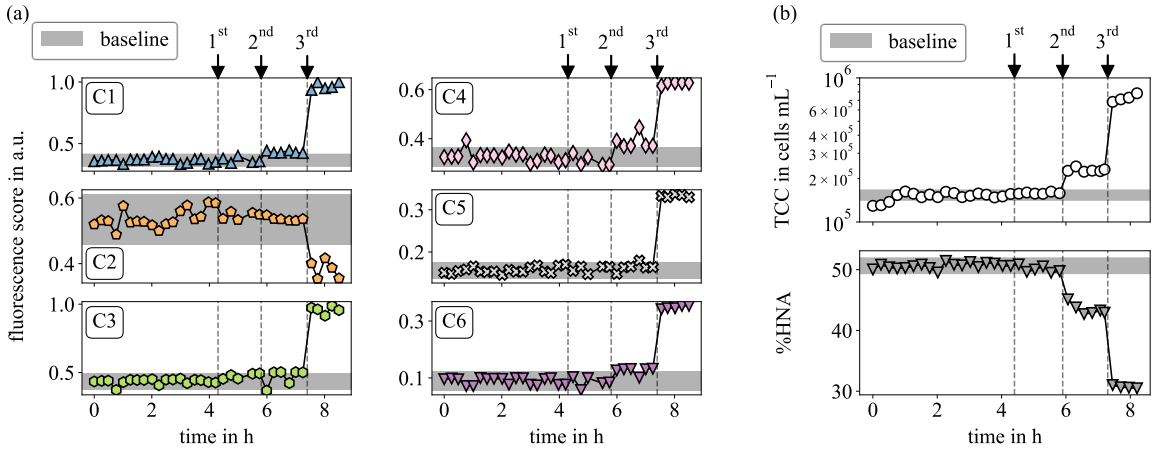

Figure 5-9: Automated online monitoring experiment (vii) results of spiking drinking water (dw_2) at three different times with increasing concentrations 1st 0.1 v%, 2nd 1 v% and 3rd 10 v% of wastewater effluent (wwe_1). Left side (a) shows the PARAFAC fluorescence scores of the respective components C1–C6 while right side (b) shows the TCC and %HNA over the experimental duration.

After the 1st spiking event, fluorescence scores do not show a change in intensity. After the 2nd spiking event, however, an increase or decrease of the parameters value can be observed. Accordingly, the PARAFAC components, C1, C4, C6 and partly C3 exhibit a shift above the baseline. C2 and C5 don't show any significant deviation from the baseline. After the 3rd spiking event, changes in water quality are registered for all six components. Due to the fact that all components, except C2, display higher fluorescence scores of pure wwe_1 compared to dw_2 (Ta-

ble 5-2), the development of the corresponding fluorescence scores can be described as expected. The same trend applies to flow cytometry results. Also, flow cytometry shows no clear trend after the 1st spiking event. However, a divergence upwards for TCC and downwards for %HNA can be observed.

In conclusion, spiking dw_2 with wwe_1 can be detected by parameters C1, C4, C6, TCC, and %HNA somewhere between 0.1 and 1 v%. However, these results are not transferable to any other condition, specifically not to the general spiking of drinking water with wastewater effluent. A detailed discussion regarding the general definition of parameters' sensitivity is provided after the following experiment results in chapter 5.3.4.

5.3.2. Spiking with Rainwater – (viii)

Progressive climate change is causing an increase in the occurrence of heavy rainfall which might affect drinking water quality (Taylor et al., 2013). Rainwater can be considered a major hazardous contaminant either by directly flooding water tanks or by causing runoff on the drinking water distribution systems (treatment plants and pipes) (Favere et al., 2021; Tornevi, 2015). To simulate rainwater contamination events, this experiment was conducted. Experiment (viii) results of spiking dw_2 with rw_1 are shown in Figure 5-10. The period until the 1st spiking event was utilized to generate the baseline of dw_2 and define the upper and lower limits for measurements (three times the standard deviation, Equation 2-5) to be recognized as an event or outlier. 1st rw_1 spiking event was performed after 3.6 h with 0.1 v% of rw_1, 2nd spiking event after 5.6 h with 1 v% of rw_1, and 3rd spiking event after 7.9 h with 10 v% of rw_1, respectively.

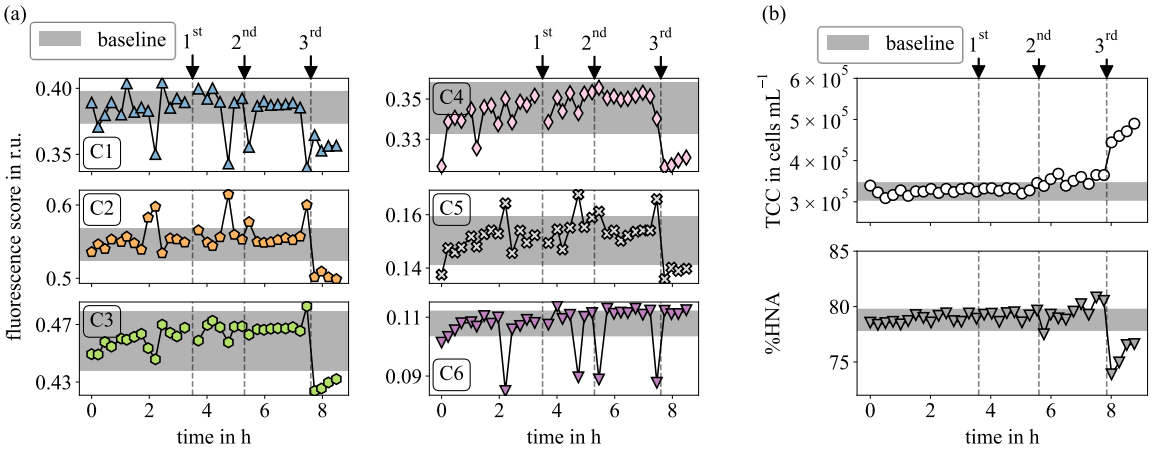

Figure 5-10: Automated online monitoring experiment (viii) results of spiking drinking water (dw_2) at three different times with increasing concentrations 1st 0.1 v%, 2nd 1 v% and 3rd 10 v% of rainwater (rw_1). Left side (a) shows the PARAFAC fluorescence scores of the respective components C1–C6 while right side (b) shows the TCC and %HNA over the experimental duration.

Considering fluorescence spectroscopy results, the baselines of all PARAFAC components appear in the same range as observed in experiment (vii) (Figure 5-9). No significant changes can be observed after the 1st and 2nd spiking event. Even before the 1st spiking event, some outliers were identified. This shows the instability of the PARAFAC model for dw_2 and highlights the need for consecutive measurements to identify evolving changes. Based on the results shown

here, it can be assumed that a significant trend can be reliably guaranteed using PARAFAC monitoring. This is triggered when more than two measurements in a row deviate significantly from the baseline range. After the 3rd spiking event, a substantial and sustainable decrease outside the baseline occurs for C1–C5 fluorescence scores. Since all scores of fluorescence components, except C6, were lower for pure rw_1 than for dw_2, the behaviors of the curves in this manner were to be expected.

Monitoring via flow cytometry shows a remarkably sensitive detection of rw_1 spiking. While no change in both parameters is visible after the 1st spiking event, there is a slight increase of TCC after the 2nd spiking event. However, not all measurements after the 2nd spiking event are outside the baseline range. As a result of the 3rd spiking event, TCC exhibits a significant and continuous increase. The same steady, but less strong, TCC increase was monitored during the experiment (vii). A subsequent continuous increase after a spiking event might indicate the required time for the bacteria to fully disperse in dw_2. %HNA monitoring only shows a significant decrease after the 3rd spiking event. However, for the duration before the 3rd spiking event a slight increasing pattern is observed but cannot be clearly assigned.

Therefore, a general comparison between the flow cytometric baselines of experiment (vii) and experiment (viii) is required. Although the experiments were conducted using drinking water from the same reservoir (dw_2), TCC and %HNA varied significantly (Table 5-2). This may be an indication of a deviation of the bacterial fingerprint between both experiment days. However, dw_2 was stored two days before experiment (vii) was started and four days before experiment (viii) was started, respectively. This time lag could have affected on the state of the bacterial community. The rather high amount of HNA cells during experiment (viii) might indicate a shift in the bacterial community towards HNA cells, whereby it cannot be precluded that this is also changing during the duration of the experiment. Thus, it is not possible to determine whether the slight increase in the %HNA before the 3rd spiking event is caused by the intrinsic shift of LNA and HNA cell communities in dw_2 or due to the spiking event with rw_1 itself.

In summary, spiking dw_2 with rw_1 can be detected by TCC between 0.1 and 1 v%. A range of 1 to 10 v% can be detected using %HNA and C1-C5. However, a more detailed discussion regarding the general sensitivity of each parameter is provided following the result sections in chapter 5.3.4.

5.3.3. Spiking with Varying Drinking Water – (ix)

As the last part of online monitoring experiments, a different drinking water (dw_3) was chosen as a spiking medium to define each parameter's sensitivity regarding moderately varying water quality change. This might be able to transfer towards possible water quality change scenarios which were not considered in this work such as groundwater or river/lake water contamination within the WDS (chapter 2.1). Furthermore this experiment addressed the application of both

methods to monitor water quality changes in mixing zones of WDSs or due to changing well circuits (Prest et al., 2013).

All results from experiment (ix) are shown in Figure 5-11. Excluding outliers, the remaining measurements before the 1st spiking event were used to generate the baseline of dw_2 and define upper and lower detection limits (Equation 2-5). The 1st spiking event was performed after 3.3 h with 0.1 v% of dw_3 after 3.3 h, the 2nd spiking event after 5.3 h with 1 v% of dw_3, and the 3rd spiking event after 7.6 h with 10 v% of dw_3.

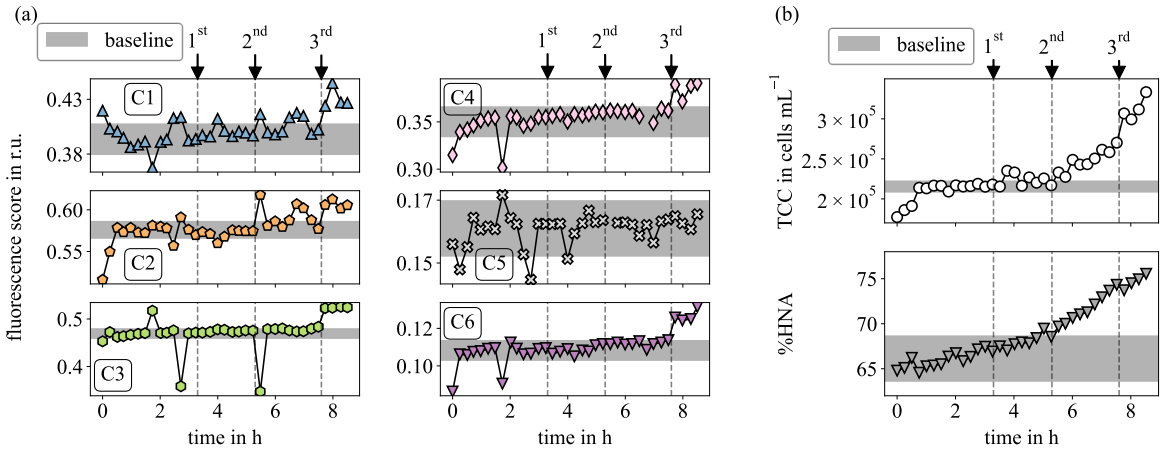

Figure 5-11: Automated online monitoring experiment (ix) results of spiking drinking water (dw_2) at three different times with increasing concentrations 1st 0.1 v%, 2nd 1 v% and 3rd 10 v% of varying drinking water (dw_3). Left side (a) shows the PARAFAC fluorescence scores of the respective components C1–C6 while right side (b) shows the TCC and %HNA over the experimental duration.

Similar to experiment (viii), the components fluorescence scores in Figure 5-11 also show deviations including outliers, even before the 1st spiking event. For the majority of the fluorescence components, there are no visible trends after the 1st and 2nd spiking events. Nevertheless, the fluorescence scores of C1 and C2 show consecutive measurements above the upper baseline value. Due to the fluctuating behavior and the limited number of measurements, a clear trend is difficult to interpret. However, it is different for measurements after the 3rd spiking event. A clear increase in the fluorescence scores can be observed for C1, C2, C3, C4, and C6, yet with varying significance. The fluorescence score of C5 remains constant for the entire experimental period, which could be expected because of the similar intensity of C5 in dw_2 and dw_3 (Table 5-2).

Flow cytometry shows a precise increase in both TCC and %HNA after the 2nd spiking event. However, TCC constantly increases after the 2nd spiking while the %HNA increases over the entire experiment period.

Similar to experiment (viii), the constant TCC increase after spiking events might indicate that the dispersing procedure of both waters requires some time, which could not be provided by the period between the 2nd and 3rd spiking event. There may be evidence that the bacterial community is undergoing a transformation during the experiment, as indicated by the constant shift from LNA to HNA cells before the 1st spiking event. Thus, it is difficult to interpret the results

of bacterial monitoring in relation the induction of different waters. Again, it should be considered that even if dw_2 was sampled on the same day for all three experiments, the time lag between sampling and the starting point of experiment (ix) was the longest at eight days. This allows an intrinsic shift within the bacterial community (from LNA to HNA cells) compared to previous experiments.

In general, flow cytometry results illustrate the necessity of a flexible dynamic baseline as emphasized by Favere et al. (2021). The flow cytometry results of experiments (viii) and (ix) indicate that bacterial community distributions can change over time, even without external spiking events. Hence, the automation code developed and introduced in chapter 5.2 allows for adjusting the number of samples taken into account for the baseline calculation (code lines 780–880 in *'III_app_online_dashboard.py'* in Schuster (2024)). Thus, it is flexible to adjust, e.g., the last 10, 20, or 30 samples for the baseline calculation.

To conclude, spiking dw_2 with dw_3 can be detected by C1, C2, C3, C4, and C6 between 1 and 10 v%. Flow cytometry seems more sensitive. A more detailed evaluation regarding the sensitivity of each parameter in case of an induced water quality change is discussed in the following chapter.

5.3.4. Sensitivity of PARAFAC Fluorescence Spectroscopy and Flow Cytometry Parameters

After demonstrating the results of the online monitoring experiments spiking drinking water (dw_2) with increasing concentrations of varying waters (wwe_1, rw_1, and dw_3), this section addresses two aspects:

- **Comparing Mathematically Estimated with Measured Parameter Values.** Addressing the question of how monitoring methods react in the event of a sudden water quality change, which was simulated here, what are the mathematically estimated values for each parameter after each spiking event?
- **Sensitivity of Monitoring Parameters.** Focus mainly on drinking water analyzed here (dw_2), but is it possible to transfer the sensitivity of each parameter to other drinking waters?

Comparing Mathematically Estimated with Measured Parameter Values

Mathematically estimated values of each parameter after respective spiking events were calculated based on the volume proportion. This was done using the initial water quality parameter values in Table 5-2, assuming ideal mixing. Therefore, Equation 5-1 shows how to calculate the estimated value of one parameter X_e, where X stands for all monitored parameters (C1–C6, TCC, and %HNA). j stands for the volume proportion (0.1, 1, or 10) of the spiked water (wwe_1, rw_1, or dw_3) after the respective spiking event. X_s represents one spiked water's initial parameter value, and X_b represents the baseline the respective values of dw_2. Since baseline values of

dw_2 contain a standard deviation (Table 5-2), a range can also be calculated for the estimated parameter value X_e.

$$X_e = \left(\frac{j}{100} \times X_s\right) + \left((1 - \frac{j}{100}) \times X_b\right) \qquad \text{(Equation 5-1)}$$

The assumed ranges for each parameter value are displayed through boxes in Figure 5-12. In comparison, the respective measured value intervals during experiments (xii), (xiii), and (ix) are displayed. Therefore, the mean value including standard deviation is calculated after each spiking event. By comparing X_e with X_b, the representation is used to determine if the spiked water could be detected under specific conditions, and if so, what range the expected values fall. A comprehensive statement regarding the accuracy of PARAFAC fluorescence spectroscopy and flow cytometry parameters can also be made by comparing expected and measured value ranges.

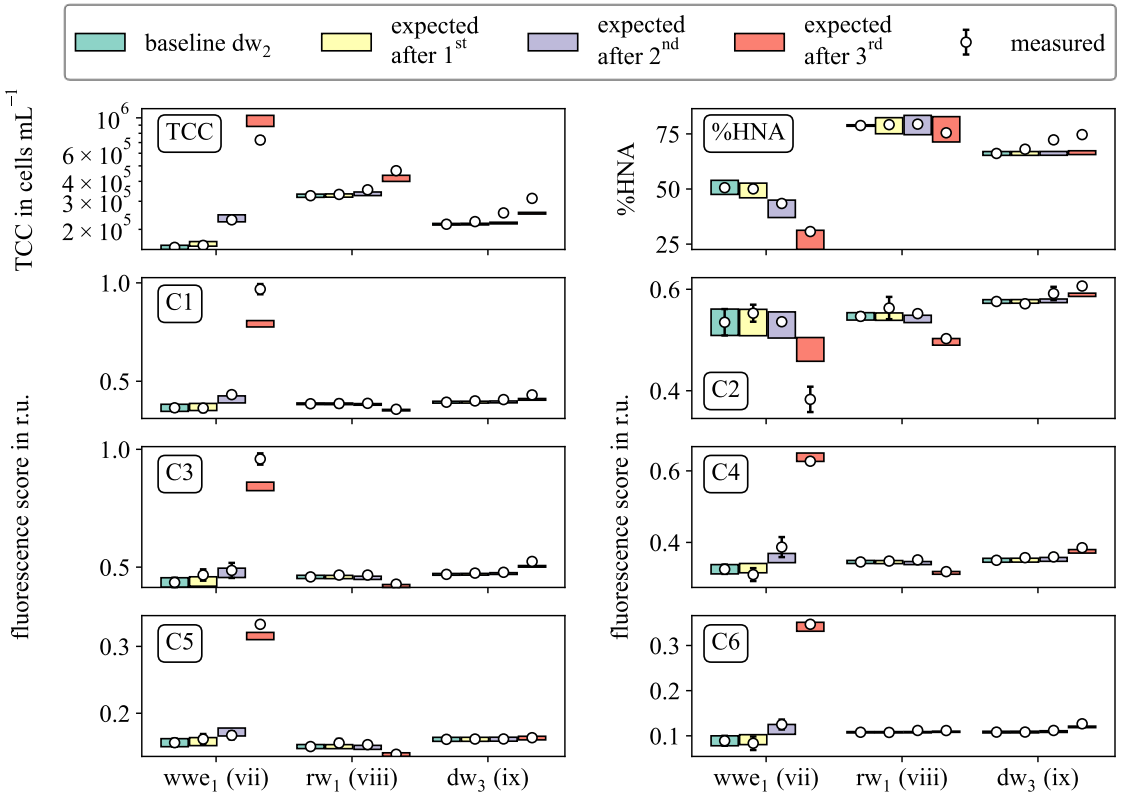

Figure 5-12: Estimated and measured parameter value ranges for conducted spiking events of the three performed online monitoring experiments (vii), (viii), and (ix). Baseline dw_2 represents the parameter range before the 1st spiking event. Spiking event proportions: 1st 0.1 v%, 2nd 1 v%, and 3rd 10 v%.

Regarding the wwe_1 experiment (vii), the estimated range of all parameter values does not deviate significantly from dw_2 baseline after the 1st spiking event, but a significant deviation is to be predicted after the 2nd spiking event. An exception is C2, for which a deviation is expected only after the 3rd spiking event. Comparing the estimated ranges with the measured values reveals that all parameter data match until after the 2nd spiking event. However, after the 3rd spiking event there are some significant discrepancies between estimated and measured parameter

values. This is especially true for C1, C2, C3, C5, and TCC. Regarding fluorescence spectroscopy parameters, deviations between estimated and measured ranges could be linked to the challenge of under- or overdetermining components' fluorescence score intensities. This can appear when adding a sample to the PARAFAC model with significant higher or lower DOM concentration compared to the remaining PARAFAC model samples (here: *base data*). With regard to the PARAFAC modelling of pure wastewater effluent, this effect was discussed previously (wwe_1 in Table 5-2). However, the proportion of 10 v% wwe_1 after the 3rd spiking event might lead to a notably varying DOM constitution. This can therefore affect PARAFAC modelling and corresponding fluorescence scores significantly. As discussed earlier (Figure 5-9), TCC showed increasing behavior after the 3rd spiking event, which could indicate that the final TCC was not reached at the subsequent four measuring points after the 3rd spiking event. Probably, the dispersion of the cells in water would have required more time to stabilize.

Continuing with rw_1 experiment (viii), the detection of rw_1 can mainly be expected after the 3rd spiking event. Only parameters %HNA and C6 do not show any change with increasing proportion of rw_1 due to the similar initial ranges of both parameters (Table 5-2). All measured parameters of experiment (viii) lie within the estimated parameter value ranges.

Regarding dw_3 experiment (ix), the detection of dw_3 can be expected after the 3rd spiking event. Whereas all measured fluorescence spectroscopy parameters can be found in the expected range there is a deviation between measured and expected flow cytometry parameters. A significant increase of %HNA as well as TCC after each spiking event is visible, which was already discussed in context of Figure 5-11. An indication of microbial change regarding bacterial growth and community shift (from LNA cells towards HNA cells) was discussed for both drinking waters dw_2 and dw_3 which might be the reason that the expected ranges do not match the measured values.

Summarizing all three experiments (vii), (viii), and (ix), it can be concluded that all expected water quality parameter changes were successfully detected. However, there were some over- or underdetermining of fluorescence scores caused by the PARAFAC model limitations when dealing with individual measurements whose DOM content deviated significantly from the remaining modelling samples. Furthermore, considering experiment (ix), flow cytometry parameters TCC and %HNA showed the sensitivity of detecting changes in the bacterial community which were not expected. This highlights the added value of using both PARAFAC fluorescence spectroscopy and flow cytometry to monitor the content of DOM and the bacterial constitution appropriately. Moreover experiment (ix) shows that a small change in the bacterial constitution of one water does not necessarily affect the intensities of PARAFAC fluorescence spectroscopy parameters. Dynamic baselines should be emphasized in this context. Even the relatively short experiment durations analyzed here demonstrated that bacterial water quality parameter values could vary over the course of the experiments. These changes must be taken into account when monitoring water quality for longer periods.

Sensitivity of Monitoring Parameters

Regarding the sensitivity of the monitoring parameters investigated in this work (C1–C6, TCC, and %HNA), first, a summary of the dw_2 specific monitoring experiments results is provided, before literature investigating the detection of respective water quality parameters changes is discussed. Last, an approach to defining general drinking water sensitivity limits for each parameter is introduced.

Based on all recorded monitoring experiments of dw_2, it was not possible to detect a spiking proportion of 0.1 v%, not even for very different water matrices such as wastewater effluent (here: wwe_1). A spiking proportion of 1 v% was detectable for wwe_1, although not for the remaining investigated rainwater (rw_1) and drinking water (dw_3). All investigated waters were successfully detected as having a spiking proportion of 10 v%. The conducted spiking experiments showed that the possibility of detecting a change in water quality depends significantly on parameter value deviations between the continuously monitored water (baseline) and the spiked water.

Literature dealing with event detection based on induced water quality changes focuses on microbial parameters and uses flow cytometry. In two separate experiments, Prest et al. (2013) conducted experiments of mixing one drinking water with wastewater and another drinking water, respectively. They detected a change in flow cytometric fingerprint (TCC and %HNA) when more than 2 % of the TCC in the mixed samples originated from the wastewater samples. In comparison with drinking water mixing, detection was possible when more than 40 % of the TCC originated from the added drinking water. A more sophisticated approach was examined by Besmer et al. (2017) and Favere et al. (2021). Besmer et al. (2017) investigated automated TCC monitoring of a drinking water sample that had been spiked with wastewater. They demonstrated the feasibility of detecting a wastewater proportion of 0.6 v%. The applied wastewater contained approximately 100-fold higher TCC than monitored drinking water. A comparative study was carried out by Favere et al. (2021) who detected a fraction of 1 v% wastewater in drinking water by analyzing the TCC. In this study TCC of wastewater contained approximately 32-fold greater TCC of monitored drinking water. In comparison to these studies, in the present work TCC of wwe_1 was approximately 53-fold higher than TCC of dw_2.

Research investigating the sensitivity of PARAFAC fluorescence spectroscopy intensities was conducted by Stedmon et al. (2011). Spiking drinking water with wastewater revealed the detection of a contamination event ranging between 0.1 and 0.25 v%. The wastewater used by Stedmon et al. (2011) contained approximately 7-fold higher TOC than monitored drinking water TOC. In comparison to the present work, the TOC of wwe_1 has been approximately 6-fold higher than the TOC of dw_2.

All studies emphasized the high sensitivity of parameters from both PARAFAC fluorescence spectroscopy and flow cytometry. This is regarding the detection of spiking in drinking water

with small amounts of wastewater. However, the results also show that the possibility of detection heavily depends on the individual parameters (TCC, %HNA, and TOC) of the two analyzed waters. Hence, a general valid approach would be helpful to predict the detection of waters based on their individual parameter values TCC, %HNA, and PARAFAC fluorescence scores of C1–C6. The following section addresses this gap.

The following paragraph focuses on theoretical and fully simulated scenarios for the induction of water into an online monitored system. The online monitored drinking water system is assumed to be perfectly mixed. The ability of online monitored water to recognize changes in parameter values due to induced water depends on two aspects. First, the deviation between one parameter value from the induced water and the respective parameter value of monitored water. Second, the volume proportion of induced water in relation to the total volume of monitored water. Both aspects are addressed in simulated scenarios. In order to make the simulation as general as possible, a typical model drinking water is generated. Therefore, each parameter median value of in this work investigated drinking waters (Figure 4-2 and Figure 4-3) were taken into account to create a theoretical drinking water (dw_t). Its parameter values are summarized in Table 5-3. Regarding the detection limit of each parameter, upper and lower baseline limits are defined by three times the standard deviation (Equation 2-5). The standard deviations were taken from conducted online monitoring experiments samples (Table 5-2). Table 5-3 contains both standard deviations and baseline limits of dw_t.

Table 5-3: Generated water quality parameter values characterizing typical median drinking water (dw_t). Median values are taken from Figure 4-2 and Figure 4-3, standard deviations are taken from Table 5-2, upper and lower baseline limits are calculated based on Equation 2-5.

parameter	median	standard deviation	baseline upper limit	baseline lower limit
C1 in r.u.	0.580	0.01	0.610	0.550
C2 in r.u.	0.482	0.01	0.512	0.452
C3 in r.u.	0.484	0.01	0.514	0.454
C4 in r.u.	0.305	0.01	0.335	0.275
C5 in r.u.	0.132	0.005	0.147	0.117
C6 in r.u.	0.117	0.005	0.132	0.102
TCC in cells mL^{-1}	1.39×10^5	5×10^3	1.55×10^5	1.25×10^5

This water dw_t is taken as the basis for subsequently simulated scenarios and can be assumed as a general valid drinking water. The volume proportions of the induced water simulated in this work were chosen based on previously discussed experiments (vii), (viii), and (ix) and the debated literature. The respective volume proportions were 0.1-, 0.5-, 1-, 2-, 5-, and 10 v%. Regarding the parameter values of induced water, they were chosen relatively to the respective parameter values of dw_t. They were selected in a range to simulate other drinking waters (Figure 4-2, and Figure 4-3) but also with higher deviations in order to simulate different waters such as wastewater, rainwater, groundwater or surface water. The proportion for each parameter value of the induced water is as follows: 0.1-, 0.5-, 1.5-, 2-, 5-, 10-, 20-, and 50-fold, relative to

the respective parameter of dw_t. Considering each possible combination of both variables, volume proportion and relative parameter value, this leads to 48 different scenarios of simulated water quality changes. Assuming proper mixing of dw_t and the induced water allows to apply Equation 5-1 to determine particular parameter values of all 48 theoretically mixed water. Considering that each parameter's detection limit is defined by the upper and lower baseline values, determined parameter values are classified into either inside the baseline limits or outside the baseline limits. This shows the sensitivity of the respective parameter to detect a change in water quality under certain conditions. Table 5-4 shows the results of 48 simulated scenarios. The red coloring of a cell means that the calculated value of mixed water is outside the baseline. This is therefore recognized as a significant change in water quality. Cells colored green mean that the calculated value is inside the baseline which does not imply a corresponding change in drinking water quality.

Table 5-4: Simulated experiment results investigating water quality parameters' sensitivity. Drinking water (dw_t) is spiked with varying volume proportions (0.1-, 0.5-, 1-, 2-, 5-, and 10 v%). Simulated spiked water contains parameter values 0.1-, 0.5-, 1.5-, 2-, 5-, 10-, 20-, and 50-fold, relative to dw_t. Red cells and green cells represent parameter values calculated outside and inside the baseline limit values after spiking, respectively.

		C1	C2	C3	C4	C5	C6	TCC			C1	C2	C3	C4	C5	C6	TCC
0.1 v%	0.1-fold								2 v%	0.1-fold							
	0.5-fold									0.5-fold							
	1.5-fold									1.5-fold							
	2-fold									2-fold							
	5-fold									5-fold							
	10-fold									10-fold							
	20-fold									20-fold							
	50-fold									50-fold							
0.5 v%	0.1-fold								5 v%	0.1-fold							
	0.5-fold									0.5-fold							
	1.5-fold									1.5-fold							
	2-fold									2-fold							
	5-fold									5-fold							
	10-fold									10-fold							
	20-fold									20-fold							
	50-fold									50-fold							
1 v%	0.1-fold								10 v%	0.1-fold							
	0.5-fold									0.5-fold							
	1.5-fold									1.5-fold							
	2-fold									2-fold							
	5-fold									5-fold							
	10-fold									10-fold							
	20-fold									20-fold							
	50-fold									50-fold							

The results provide insight into which conditions of a scenario must be present for a parameter to be recognized as an event outside the baseline. Regardless of the water quality of the incoming or induced water, it is not possible to assess any water quality change for a volume fraction

of 0.1 v% by any parameter. As the volume proportion increases, the probability of detecting a deviation from the baseline also increases. In comparison to the remaining parameters, there seems to be a tendency that parameters C1, C2, and C3 can identify less deviant water.

To put the results into context, the range of all parameters for studied waters in this work is briefly discussed. All PARAFAC fluorescence scores of investigated drinking waters vary generally between approximately 0.5- and 2-fold of dw_t PARAFAC fluorescence scores. Thus, inducing drinking water must have a volume proportion greater than 5 v% to be recognized. On the contrary TCC of all investigated drinking waters vary between approximately 0.1- and 10-fold of dw_t which allows to detect even lower volume proportions (> 1 v%). Regarding the investigation of different water sources such as wastewater effluent and rainwater, the fluorescence scores mainly deviated between approximately 0.1- and 10-fold, relative to dw_t fluorescence scores, while the TCC of wastewater effluent was approximately 50-fold the TCC of dw_t. It is more likely that TCC will detect significant deviations in water, such as wastewater effluent or rainwater, followed by C1, C2, and C3.

The discussion about the sensitivity of parameters provides an initial approach to how changes in water quality could also be recognized in real WDS by applying PARAFAC fluorescence spectroscopy and flow cytometry as long-term monitoring methods. The addressed changes can have various origins, such as seasonal water quality changes, changing well circuits, or contamination events. With regards to detecting an evolving event through online monitoring of PARAFAC fluorescence spectroscopy and flow cytometry, it is indispensable to appropriately define parameter specific baselines and their upper and lower limits which characterize the detection limit of each parameter. All displayed simulated scenarios apply a model of drinking water which is generated from the median values of all investigated drinking waters. Since each water quality parameter can vary significantly in real life, aiming for a general valid approach is still a challenge. Additionally, simulations were performed in perfectly mixed laboratory systems. Compared to a real world WDS change of water quality, a WDS change of water quality shows more complexity, such as the continuous intrusion of an inducing water over an extended period of time.

Highlights Chapter 5.3

- Under examined experimental conditions (1 L water bottle, assuming perfectly homogenization), PARAFAC fluorescence spectroscopy and flow cytometry could detect sudden changes in near real-time. These changes were artificially created by spiking drinking water with, e.g., wastewater effluent and rainwater.
- Combining PARAFAC fluorescence spectroscopy with flow cytometry was necessary to detect any changes regarding DOM or bacterial constitution.
- Baselines of each parameter are crucial to generate to define parameter specific detection limits regarding parameter specific changes. Using three times standard deviation of each parameter to define upper and lower baseline limits was successful.
- The ability to detect a water quality change in case of an induced water event strongly depends on individual water quality parameter values of both monitored and induced water, and the proportion of induced water.
- In a perfectly homogenized system, the parameters TCC, C1, C2, and C3 were shown to be the most sensitive in terms of detection of water quality changes.

6. Transforming the Monitoring System to the Field

The following chapter focuses on the application of PARAFAC fluorescence spectroscopy and flow cytometry to field challenges. It aims to show how both can be integrated into water quality control principles. First chapter 6.1 deals with their sensitivity within a pilot plant flow pipe system. It considers the advantages compared to current water quality monitoring methods, such as TOC, pH and conductivity. Chapter 6.2 highlights an approach to applying a developed groundwater specified PARAFAC model to monitor groundwater characterization and quality control.

6.1. Pilot Plant – Detection of Water Quality Changes in a Flowing System

Contrary to experiments discussed in chapter 5.3, experiments in this chapter were conducted in a more real-world environment and application. The idea was to simulate events of drinking water contamination or general water quality changes, e.g., due to small pipe leakages, seasonal variations, well circuit changes, or drinking water reservoir contaminations. These trials aspire to address more precisely the real cases of drinking water quality changes and contamination events reported by Moreira and Bondelind (2017) and Hrudey and Hrudey (2007).

The pilot plant system was filled with approximately 4400 L the day before performing respective experiments using water from the same hydrant for each trial (dw_4). Inhomogeneous temperature conditions of the containing water were recorded each morning of the experiment due to the different ambient temperature of the pilot plant (indoor and outdoor). The pilot plant's pumps were started prior to the actual experiments in order to address this temperature gradient. According to preliminary experiments (chapter 3.3.3 and Figure I-9), a stable system temperature was reached after approximately 150 min. To achieve a typical WDS flow velocity of approximately 0.5 m s^{-1}, the pumps were adjusted to generate a flow of 58 ± 4 L min^{-1} ($Re \approx 5700$). Regarding sampling depths, which were examined in the preliminary studies, there were no deviations between 0 mm (pipe wall) and 100 mm (pipe center). Consequently, these depths were determined for the actual experiments in the pipe center (chapter 3.3.3 and Figure I-9). Due to regular maintenance and flushing of the pilot plant, the pipes did not show the same biofilm environment as in regular WDS. Consequently, biofilm bacteria mobilization did not play a major role in TCC monitoring during the experiments.

Table 6-1 presents water quality parameters of respective waters for the three executed experiments. Fluorescence scores were calculated applying the overall six-component PARAFAC model using *base data*. Compared to the laboratory experiments in chapter 5.3 the conventional water quality parameters pH and conductivity were determined as well.

Table 6-1:Water quality parameter values (PARAFAC fluorescence spectroscopy, flow cytometry, TOC, pH, and conductivity) of all investigated waters during pilot plant experiments. (x), (xi), and (xii) represent the three different experiments of injecting different waters (wastewater effluent wwe_2, rainwater rw_2, and drinking water dw_5) into the pilot plant water (dw_4).

dw_4 experiment	water	TCC in cells mL^{-1}	%HNA in %	C1 in r.u.	C2 in r.u.	C3 in r.u.	C4 in r.u.	C5 in r.u.	C6 in r.u.	TOC in mg L^{-1}	pH in a.u.	cond. in µS cm^{-1}
(x)	dw_4	$1.82 \times 10^5 \pm 8.8 \times 10^3$	40 ± 2	0.699 ± 0.020	0.878 ± 0.012	0.748 ± 0.010	0.561 ± 0.015	0.301 ± 0.007	0.166 ± 0.013	3.4	7.7	532
	wwe_2	$1.69 \times 10^7 \pm 1.7 \times 10^5$	30 ± 0	5.028	0	4.387	3.671	2.076	2.193	16.4	7.6	880
(xi)	dw_4	$1.79 \times 10^5 \pm 7.7 \times 10^3$	41 ± 2	0.650 ± 0.009	0.810 ± 0.005	0.693 ± 0.008	0.521 ± 0.009	0.290 ± 0.007	0.156 ± 0.003	3.3	7.8	527
	rw_2	$1.03 \times 10^7 \pm 9.7 \times 10^4$	96 ± 1	0.123	0.011	0.107	0.326	0.186	0.508	2.5	6.3	34
(xii)	dw_4	$1.35 \times 10^5 \pm 9.1 \times 10^3$	61 ± 0	0.445 ± 0.004	0.621 ± 0.006	0.522 ± 0.005	0.394 ± 0.003	0.192 ± 0.002	0.124 ± 0.004	2.8	7.9	484
	dw_5	$1.26 \times 10^5 \pm 1.27 \times 10^4$	51 ± 4	0.609	0.785	0.889	0.661	0.209	0.207	4.6	7.8	480

All parameters of dw_4 are mean values considering samples of the pilot plant water before actual injection experiments started. The water quality parameters of dw_4 mainly differ between the first two experiments (x), and (xi) and the last experiment (xii). Especially fluorescence scores, TOC, conductivity and %HNA show a significant increase in the first two experiments. In view

of the fact that DW4 was taken from the same hydrant, it could either be related to the general water quality changes of the distributed water or to internal changes or variations within the pilot plant itself. However, all parameters of dw_4 can be considered stable during experimental times.

Injected waters (wwe_2, rw_2, and dw_5) were measured once (fluorescence spectroscopy) and three times (flow cytometry and conventional parameters). A detailed discussion of wastewater effluent and rainwater parameters is provided in chapter 5.3 which were taken from the same wastewater treatment plant and rainwater water tank, respectively, but on a different day. All parameter values of the wastewater effluent of experiment (x) are in a similar range to wastewater effluent of experiment (vii), but show slight increases. Daily changes of wastewater effluent parameters are known due to, e.g., high seasonal variations and corresponding varying DOM constitution of untreated wastewater (Carstea et al., 2018). Rainwater parameters of experiment (xi) and experiment (viii) are more difficult to evaluate due to the different sampling times and the strong dependency of rainwater quality on the duration since the last precipitation, the climate, and atmosphere pollution (Farreny et al., 2011; Förster, 1999). Experiment (xi) rw_2 contained approximately 10-fold the quantity of TCC, compared with experiment (viii) rw_1. Moreover, cells of rw_2 contains almost exclusively HNA cells. Additionally, organic parameters of rainwater differ between experiment (xi) and experiment (viii). Compared to rw_1, rw_2 showed a significantly higher TOC (2.5 to 0.7 mg L^{-1}) which is shown by the increasing tendency of the respective fluorescence scores as well. Although the TOC of rw_2 is within the typical range of drinking water (Figure 4-4), the fluorescence scores of C1, C2, and C3 are significantly below the typical range of drinking water (Figure 4-2). However, C4 and C5 of rw_2 lie within typical drinking water ranges, while C6 is far above the typical range. This fact shows the main contribution of C4, C5; and especially C6 to the majority of the TOC of rainwater (Niloy et al., 2021). The organic matter load in dw_4 and the drinking water (dw_5) of experiment (xii) is higher for the latter, while TCC and %HNA are comparable. The different characteristics of wwe_2, rw_2, and dw_5 compared to dw_4 enabled them to be used as the medium to be injected in the pilot plant experiments.

The following sections present and discuss the monitoring results of the three experiments conducted. During each experiment, 44 L (approx. 1 v% of the total pilot plant volume) of the respective injection water was induced to the flowing water into the pilot plant at $t = 0$ h. The injection process took approximately 20 minutes.

To distinguish between occurring event and non-event, individual dw_4 parameter baselines were calculated for each experiment (three times the standard deviation, Equation 2-5). Therefore, several samples ($n > 7$) were taken before the injection process was launched ($t < 0$ h). Respective mean values and standard deviations were calculated for each parameter. In further

detail, the upper and lower baseline limits characterize the threshold. This means that all measurements outside the baseline are called events or outliers while all measurements within the baseline as referred to as non-events.

On the basis of the NaCl preliminary test results, the time periods during which the injected water would pass the monitoring point (S1) were determined. During this time, the sampling frequency was increased, up to every three minutes. For the remaining monitored period, samples were taken at S1 with an interval between five and ten minutes.

6.1.1. Injection of Wastewater Effluent – (x)

Figure 6-1 shows the fluorescence spectroscopy (a) and flow cytometry (b) monitoring results of the wastewater effluent (wwe_2) injection experiment (x). The baseline of each parameter is highlighted in the corresponding plot. Considering the general pilot plant setup of a closed loop pipe system, the mentioned flow conditions (≈ 0.5 m s^{-1}), and the performed pre-experiments, estimated times of the injected water passing by the sampling point were calculated. By injecting wwe_2 into the system at time t = 0 h, wwe_2 is expected to pass the sampling point after 0.13 h, 1.38 h and 2.63 h, respectively. The respective times are marked as the 1st, 2nd, and 3rd expected wwe_2 injection peak in the plots.

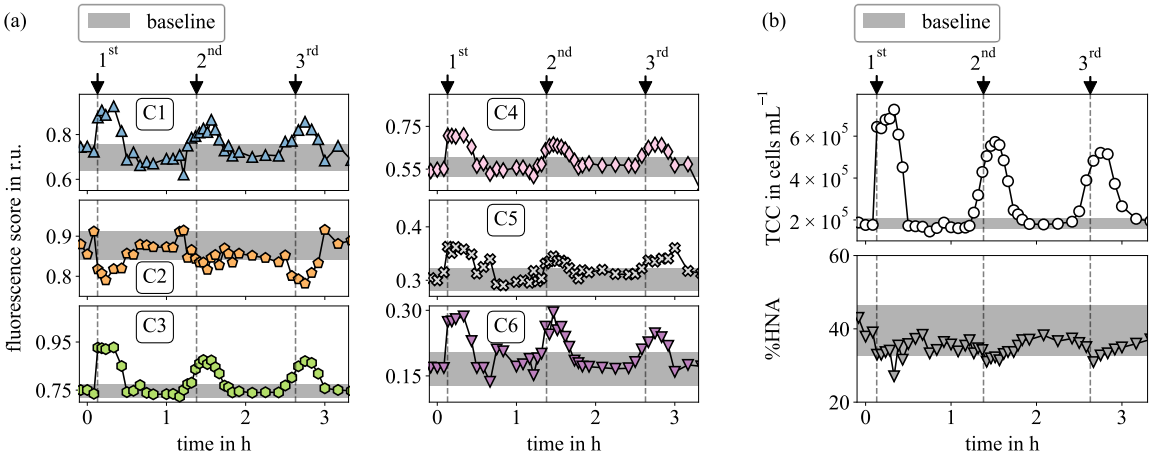

Figure 6-1: Monitored parameters of the pilot plant experiment (x) – wastewater effluent (wwe_2) injection. (a) shows the fluorescence scores of the respective PARAFAC components C1–C6, while (b) shows the TCC and %HNA over the experimental duration. wwe_2 was injected under continuous flow at t = 0. The three estimated times of the injected water passing by the sampling point are marked in the plots (1st, 2nd, and 3rd).

All parameters show a significant deviation from the baseline at the 1st expected wwe_2 injection peak. For all parameters, except C2 and %HNA, there is an increase in the fluorescence score or TCC, respectively. Due to the injection process taking approximately 20 min, an increase or decline is observed for the corresponding period. This is before the values approach the baseline again. The same behavior is observed for the 2nd and 3rd time when the wwe_2 passes by the sampling point, although with rather low amplitude. At the same time, the period of the recorded event peaks becomes a little longer. This confirms the findings from the preliminary NaCl-injection test, where the induced water mixes and homogenizes in the pilot plant water over time and distance. However, the homogenization process seems to take an extended period of

time when it comes to bacterial and DOM concentrations. Considering the slight decline between the 2nd and 3rd TCC peak, a linear approximation can be made as to how long the peak would be recognizable. This is before the peak value is within the baseline and no longer detected as an event. Accordingly, a peak could be recognizable for up to 10 hours, which would be approximately 1.2 km away. However, the change in the peak amplitude of PARAFAC components C1, C2, C3, C4, and C5 does not vary significantly between the 2nd and 3rd peaks. Hence, it is questionable whether there will be any DOM mixing at all under the present conditions. This would enable detection of wwe_2 even for long distances after injection into the pipe system. More likely the homogenization will take place within further system flow.

Regarding the stability of the particular PARAFAC components fluorescence scores, there are some measurements which might be considered as outliers. This is especially between the 1st and 2nd estimated peak. This once again demonstrates the importance of evaluating outliers in real applications but also the determination of an appropriate baseline. Some outliers could be identified as non-events using a broader baseline. This limitation, which was already emphasized in chapter 5.3, becomes clear again in this experimental setup.

To summarize, it can be highlighted that under investigated flow conditions, wastewater effluent contamination is still detectable after approximately 300 m of a pipe system. However, it will also probably be detected for longer distances.

6.1.2. Injection of Rainwater – (xi)

Figure 6-2 shows the fluorescence spectroscopy (a) and flow cytometry (b) monitoring results of the rainwater (rw_2) injection experiment (xi). The baseline for each parameter and the expected time of the 1st, 2nd and 3rd peak was calculated in exactly the same way as for the wwe_2 injection experiment (x) before.

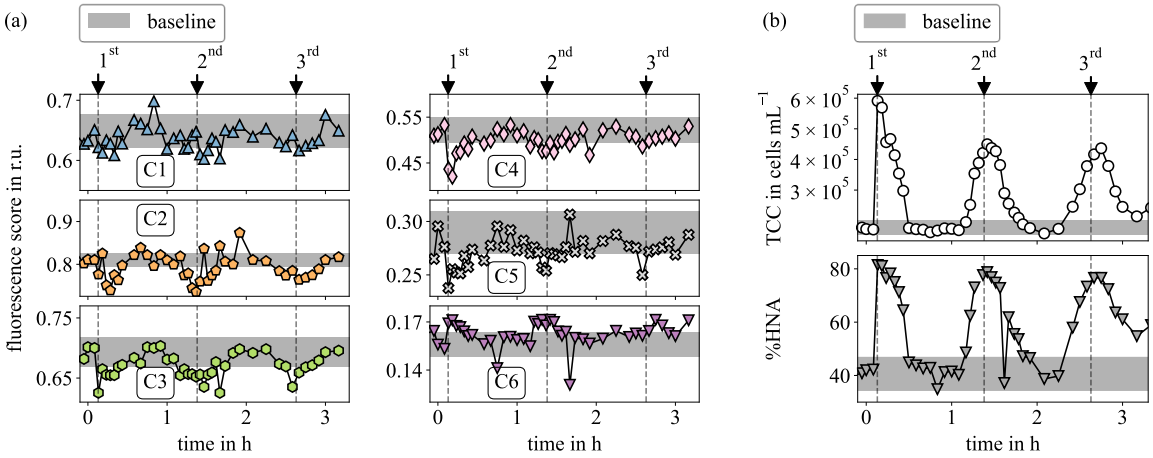

Figure 6-2: Monitored parameters of the pilot plant experiment (xi) – rainwater (rw_2) injection. (a) shows the fluorescence scores of the respective PARAFAC components C1–C6, while (b) shows the TCC and %HNA over the experimental duration. rw_2 was injected under continuous flow at t = 0. The three estimated times of the injected water passing by the sampling point are marked in the plots (1st, 2nd, and 3rd).

All parameters show a clear peak, either increase or decline, when it comes to the 1st expected rw_2 peak. Due to the fact that the fluorescence score of all components, except C6, of rw_2 is significantly lower than the fluorescence scores of dw_4, the monitoring of the scores shows downward-sloping peaks of C1–C5. However, not as clearly compared with the wwe_2 experiment (x), which is due to the lower deviation of the fluorescence scores between rw_2 and dw_4 compared to wwe_2 and dw_4 (Table 6-1). While the fluorescence scores of C1, C3, C4, C5 of wwe_2 were at least 5-fold the scores of dw_4 (C6 approximately 13-fold), the fluorescence scores of C1–C3 of rw_2 were about 0.2-fold and of C4 and C5 nearly 0.6-fold the scores of dw_4, respectively. The rw_2 fluorescence score of C6 was approximately 3-fold the score of dw_4. These relative behaviors were calculated based on the measurements shown in Table 6-1. A detailed discussion considering the relative sensitivity of each parameter in flowing systems is addressed in follow-up to the experiment results.

Besides the observed peaks, the results reveal several events as outliers. As a result, C2 shows a rather narrow baseline which leads to various outliers. It demonstrates the limitations of PARAFAC fluorescence spectroscopy regarding the ability to detect very little deviating water within a flowing system. On the other hand, the results of flow cytometry show a similar pattern to that of the wwe_2 experiment (x). Since the TCC of both injected waters is comparable, the course is very similar.

A difference between both injected waters is the initial amount of HNA cells. As rw_2 contains almost exclusively HNA cells, clear peaks are also recognizable for %HNA monitoring. The approach described for wwe_2 experiment (x) regarding an estimation of the distance where it is still possible to detect an increase in TCC due to rw_2 injection was also applied. Based on the decline between the 2nd and 3rd peak, a TCC event would be recognized at approximately 1.8 km distance. This is significantly longer than the distance calculated in the previous experiment. This shows the uncertainty of this calculation based on only two monitored peaks.

To conclude, the simulated water quality change due to injecting rw_2 was detected using both methods during the investigated period of 3.3 h which is equivalent to a distance of approximately 300 m. However, a lot of detected events via fluorescence spectroscopy must be considered as outliers which entails a corresponding uncertainty for this method and requires further adjustments.

6.1.3. Injection of Varying Drinking Water – (xii)

A different drinking water (dw_5) was injected in the pilot plant system for the last event detection experiment (xii), which differed slightly from dw_4 in terms of DOM, but not in terms of bacterial content. The monitoring results are presented in Figure 6-3 divided into fluorescence spectroscopy results (a) and flow cytometry results (b). As for the previous experiments (x) and (xi) the plots contain respective baselines and a time marker regarding expected peaks of the injected dw_5.

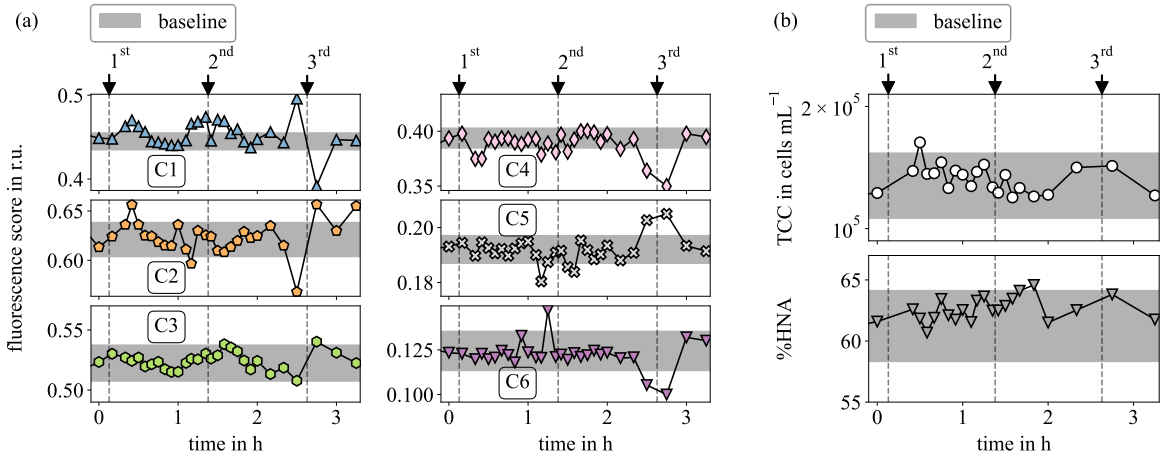

Figure 6-3: Monitored parameters of the pilot plant experiment (xii) – drinking water (dw_5) injection. (a) shows the fluorescence scores of the respective PARAFAC components C1–C6, while (b) shows the TCC and %HNA over the experimental duration. dw_5 was injected under continuous flow at t = 0. The three estimated times of the injected water passing by the sampling point are marked in the plots (1st, 2nd, and 3rd).

Due to the similarity of both waters, a clear identification of the expected peaks is more challenging than in the previously described experiments. Additionally, there were not as many data points monitored at later time points. In particular, the fourth-last and third-last data points of the fluorescence components exhibit very high deviations for all components, which is why these can be considered as outliers and thus hamper an assessment of the 3rd peak region. In the 1st peak region, C1, C2, and C4 show some monitored data outside the baseline range. However, C4 shows a decline in the fluorescence score. This was not expected due to the higher fluorescence score of dw_5 compared to dw_4. A too low baseline threshold could be the reason. This could be similar for C5, where some data points are also outside the baseline and contrary to the trend. Detecting the injected water in dw_5 is only possible with C1, the only fluorescence component following the right trend. These fluorescence spectroscopy results confirm the previous findings that minor deviation detection is challenging and requires an appropriate baseline adjustment. C1 shows, however, the general detection capability of the system, as well as the advantages in comparison with conventional monitoring methods, which will be discussed in the following section. Regarding flow cytometry results, no significant changes were observed during the experiment due to the lack of substantial deviations between both waters.

6.1.4. Comparing PARAFAC Fluorescence Spectroscopy and Flow Cytometry with Established Water Quality Parameters TOC, pH, and Conductivity

This section aims to evaluate the sensitivity to water quality changing events between PARAFAC fluorescence spectroscopy, flow cytometry and established water quality parameters TOC, pH, and conductivity. Considering the three previously discussed experiments (x), (xi), and (xii), Figure 6-4 shows an equivalent. It examines whether the respective method can detect contamination or critically important changes in water quality within a flowing system. Therefore, both flow cytometry parameters, TCC and %HNA are summarized in one marker. The same was done for PARAFAC components which are summarized in one marker as well. In both cases, the marker shows ‘outside baseline’ for a measuring point if only one particular

parameter is outside the baseline at this measuring point. Regarding the remaining conventional water quality parameters, the baseline was determined according to the procedure of fluorescence spectroscopy and flow cytometry (Equation 2-5) by calculating mean values from measuring points before the actual injection experiments started. If the respective data was recorded, Figure 6-4 shows whether the parameter is within or outside the baseline at given points in time.

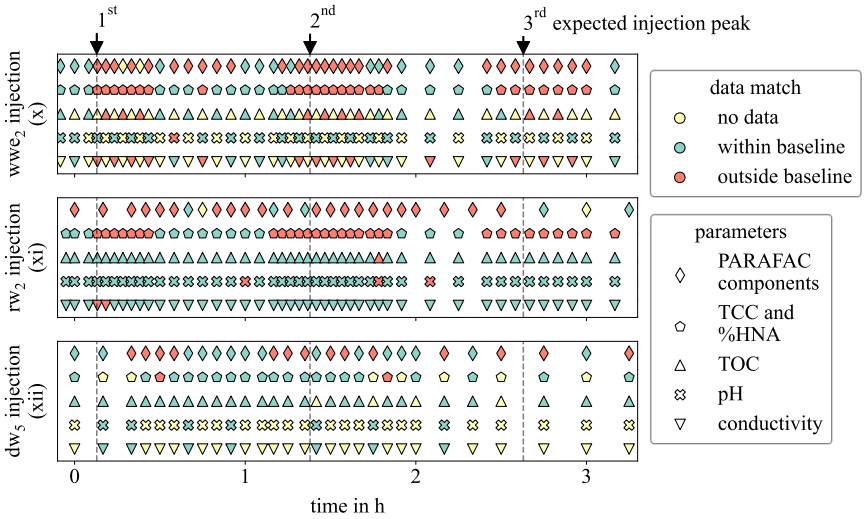

Figure 6-4: All monitored parameters of experiments (wastewater effluent wwe_2 injection – x), (rainwater rw_2 injection – xi), and (drinking water dw_5 injection – xii) are distinguished for either being within the baseline or outside of the baseline, respectively. Thereby, flow cytometry parameters are summarized as TCC and %HNA, while C1–C6 are summarized as PARAFAC components. The results of the TOC, pH and conductivity measurements are displayed as if they were recorded. The respective water was injected under continuous flow at t = 0.

Starting with the wwe_2 injection experiment (x) all methods, except pH, can detect water quality changes in the expected regions. However, PARAFAC components show several 'outside baseline' markers between the 1st and 2nd expected peaks. This is due to the registered outliers of some components (C1, C5, and C6) discussed previously (Figure 6-1). In any case, a deviating injected water such as wwe_2 could be detected using fluorescence spectroscopy, flow cytometry, as well as conventional methods such as conductivity and TOC. Both fluorescence spectroscopy and flow cytometry enable the opportunity of discriminating between measurements 'outside baseline' and 'within baseline', but deeper analysis which gives an idea about the constitution of the monitored water quality change. As described previously, the significant increase in C1, C3, C4, C5, and C6 coupled with a significant decline in C2 indicate the characteristics of inducing wastewater effluent into the drinking water system. This emphasizes the objective of monitoring all components and not only individual components, e.g., C4 which was recommended by Murphy et al. (2011) to use as a wastewater tracing component. Regarding flow cytometry results, the bacterial composition of the monitored water is more diverse than just 'outside baseline' and 'within baseline'. Hence, keeping track of TCC and %HNA as individual parameters allows a deeper characterization of the water quality change and could give an idea about AOC development of the water system (Figure 4-6). This confirms the former recommendation of Santos et al. (2019) to use discrimination between HNA and LNA cells for appropriate water quality monitoring.

In the case of the rw_2 injection experiment (xi), it can be concluded that fluorescence spectroscopy and flow cytometry are the only methods that can reliably monitor changes in the drinking water quality. There are few values outside the baseline for the remaining parameters, but they do not have a consistent behavior, making them unusable for monitoring. This experiment clearly shows the added value of both PARAFAC fluorescence spectroscopy and flow cytometry, sophisticated fingerprinting methods. However, due to several markers considered 'outside baseline', between 1st and 2nd and between 2nd and 3rd expected peaks, the limitations of PARAFAC fluorescence spectroscopy are apparent for this experiment. It may be caused by the very limited baseline range, especially regarding C2, C4, and C6. These components' baseline range is inappropriate and should be larger. Adjustment via long-term baseline monitoring of several hours or days would be required, but was not feasible during the duration of the experiment (8 hours total).

With regard to the differently pronounced C6 deviation in both experiments (x) and (xi), the general correlation between high C6 intensity and potential microbial risk should be viewed with caution. Monitoring C6 itself, as recommended by Ward et al. (2020) and Ward et al. (2021), is not sufficient to assess bacterial contamination. Rather, detailed fingerprinting via both methods is emphasized to reveal information regarding monitored water's bacterial and DOM constitution. Concerning the DOM constitution only, PARAFAC fluorescence spectroscopy is a more powerful and sensitive tool to assess the DOM of drinking water than TOC. These measurements were almost completely within the baseline. Moreover, its higher sensitivity was earlier indicated by contamination experiments (Stedmon et al., 2011) comparing fluorescence results with TOC analysis. In present work, fluorescence spectroscopy was emphasized to be an adequate method for characterizing DOM of drinking water, despite the fact that only a limited proportion of approximately 50–60 % contain fluorescent properties (Leresche et al., 2022; Stubbins et al., 2014). In case of eventual contamination and following accurate removal or disinfection of the contaminated water, PARAFAC monitoring would enable another added value by detailed DOM characterizing and thus evaluate potential formation of DBPs (for a limited set of reaction components) (Fernández-Pascual et al., 2023).

Regarding dw_5 injection experiments (xii), only PARAFAC fluorescence spectroscopy monitoring shows a reliable detected change in water quality during the experiment. This highlights PARAFAC fluorescence spectroscopy not only as a method to identify potential hazardous water quality changes, like inducing wastewater effluent or rainwater. It is also a method for detecting minor changes of the drinking water due to, e.g., mixing two drinking waters or changing well circuits. This confirms the remarks of Heibati et al. (2017) and Heibati (2019) of the necessity of applying continuous DOM characterization within the entire WDS, because of the occurrence of deviations that can affect drinking water quality. Considering the mixing of two drinking waters within the WDS, PARAFAC analysis enables the possibility to characterize

both waters and thus, determine their respective origin due to components' intensity dependence regarding the type of untreated drinking water and its treatment method (Gabrielli et al., 2023; Philibert et al., 2022).

Summarizing the results from all three experiments, PARAFAC fluorescence spectroscopy and flow cytometry facilitate the use as a cumulative parameter to distinguish between 'outside baseline' and 'within baseline'. However, the major advantage is the detailed discrimination into several parameters (TCC, %HNA, and C1–C6) which characterize the constitution of the monitored water. Different waters such as drinking water, rainwater and wastewater effluent show significant and individual differences from their original constitution, which are able to identify in the course of continuous monitoring. Generally, it is possible to set a larger baseline range for PARAFAC components to avoid false 'outside baseline' results. However, this can somehow lower the sensitivity and thus the detection limit of each method. The need for defining accurate baselines through long-term monitoring is therefore essential. Furthermore, baselines must be flexible in light of general and intrinsic changes in the water's constitution. Both methods were proven to have this flexibility and a dynamic baseline for each parameter can be adjusted as discussed in chapter 5.3.

A further aspect emphasized by Gabrielli et al. (2021), is the detection of short bacterial events that cannot be recognized using conventional plating methods. Both methods allow high-frequency monitoring which provides a detailed characterization in near real-time.

An Approach to Define Sensitivities of PARAFAC Fluorescence Spectroscopy and Flow Cytometry Parameters

Defining the sensitivity of each parameter in a dynamic system is not as trivial to determine as it was during the laboratory monitoring experiments in chapter 5.3 (Table 5-4). Because induced water cannot be considered to be perfectly homogenized with the flowing water, detection is largely dependent on the following factors. The deviation between water quality parameter values of the induced water and the flowing water, the total volume of the induced water, the flow rate of the induced water and general flow conditions of the flowing water. In the case of comparatively high flow velocities close to turbulent flow conditions, earlier homogenization of both waters is more likely. This could lead to the parameter sensitivities discussed in Table 5-4. However, under the current executed experimental conditions, only a rough estimate of the mixing behavior and the sensitivity of each parameter could be made.

As shown in the dw_5 injection experiment (xii), fluorescence scores of an injected water should be at least 1.3- to 2-fold higher than those of flowing water in the system. A fluorescence score at least 6.5 times higher than that of water in the system was clearly observed (wwe_2, experiment (x)). According to the rw_2 injection experiment (xi), fluorescence values should be 0.6 times the fluorescence values of water in the system in order to be detected at lower levels. With regard to the TCC, there are less data available, as only waters with a very high TCC content

were injected in experiments (x) and (xi). Despite this, water with approximately 50-fold TCC could be detected. Generally, the detection of a water quality change event always depends on the baseline range of the respective parameter. Uncertainties regarding the baseline range accuracy were discussed earlier.

In summary, these experiments should only serve as an approximation under given circumstances and must be validated in further trials. A similar general estimation of sensitivity of all parameters, such as in chapter 5.3 is not possible in this case. This is due to the increased number of unknown variables. Scenarios using waters with different baseline limits, different flow conditions, and varying amount of injection conditions could be investigated in future. This would increase the general understanding of the sensitivity of fluorescence spectroscopy and flow cytometry parameter in a flowing system.

Highlights Chapter 6.1

- Under performed pilot plant conditions, PARAFAC fluorescence spectroscopy and flow cytometry could detect sudden water quality changes artificially created by injecting, e.g., wastewater effluent and rainwater into the flowing drinking water.
- Similar to laboratory experiments (three times standard deviation), variable-specific baselines were successfully used to define a threshold regarding evolving events and characteristic water quality changes.
- Water quality changes were detectable even 300 m behind the injection point. Further work should consider longer distances to investigate.
- The sensitivity of each PARAFAC fluorescence spectroscopy and flow cytometry parameter depends on several factors such as hydraulic conditions and water characteristics.
- The combination of PARAFAC fluorescence spectroscopy and flow cytometry provides a clear picture of the present state of bacteria and DOM constitution in a real water flowing pipe system.
- The combination of PARAFAC fluorescence spectroscopy and flow cytometry showed high sensitivities and detailed characterization compared to TOC, pH and conductivity analysis.
- Long-term monitoring scenarios require dynamic baselines to avoid false event detection signals.

6.2. Field Trial – Groundwater Characterization via PARAFAC

Groundwater is the main source of drinking water in Germany and almost exclusively in Northern Germany (Statistisches Bundesamt, 2019). To provide high drinking water quality, groundwater quality has to be guaranteed. Its quality and especially DOM and microbial content is generally affected by external parameters, e.g., rainfall events, droughts, agricultural runoffs etc. (Barbieri et al., 2023). Results of the previous chapters highlighted the deviations between DOM characteristics of different drinking water which depend on the DOM characteristic of the respective source water as well as the applied treatment method (Gabrielli et al., 2023; Philibert et al., 2022). A gap exists between established continuous groundwater DOM characterization methods and the ability to anticipate potential evolving changes and adapt treatment procedures for better drinking water quality.

The groundwater samples analyzed in the current chapter were part of a large-scale measurement campaign to characterize groundwater samples within a supply area in Northern Germany. An emphasis was placed on identifying the levels of humic compounds, such as humic and fulvic acids, in the groundwater of various wells. In summary, this chapter aims at:

- Detailed characterization of groundwater DOM via PARAFAC fluorescence spectroscopy with the aim of using it in an online application in the future (as in chapter 5.3).
- Quantification of groundwater DOM by comparing PARAFAC fluorescence spectroscopy with LC-OCD analysis.

However, the total number of groundwater samples (n = 135) was only used for fluorescence spectroscopy analysis. All remaining characterizing methods were applied to a smaller number of groundwater samples.

6.2.1. DOM Characterization of Groundwaters

Figure 6-5 shows an overview of groundwater quality parameters determined in this work. The samples were analyzed in an anaerobic environment to avoid precipitation processes. The mean value of the TOC lied in a similar range to drinking water (Figure 4-4), although there are more groundwater samples identified in a rather high TOC range (> 5 mg L^{-1}). A comparison between TOC and DOC of the groundwater samples showed that a majority (> 90 %) of the groundwaters consist of primary dissolved organic matter (> 90 %) which lies in the observed range of reduced groundwaters (Schulz, 2020). Conductivity and pH measurements did not show any unexpected deviations and can be compared directly to drinking water samples (Figure 4-4). According to Kunkel et al. (2002) rather high conductivities may be an indication of increased well depths, which was not taken into account in this work. The iron content of all groundwater samples was far above the limit for drinking water (0.2 mg L^{-1}) (TrinkwV, 2023). The oxygen concentration of all analyzed samples lied below 1.5 mg L^{-1}. The combination of

relatively high iron content and low oxygen content allows the analyzed samples to be described as reduced groundwater (Kunkel, 2004).

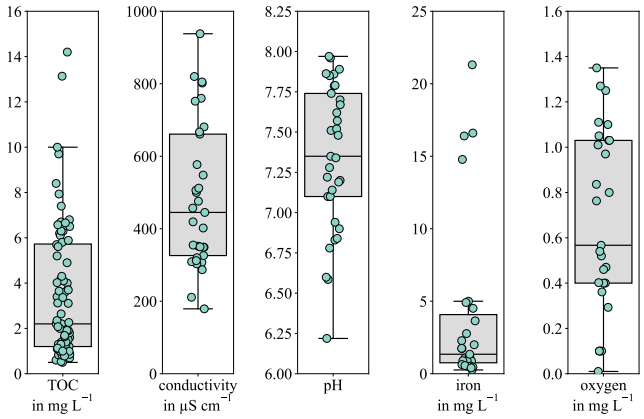

Figure 6-5: Groundwater quality parameters TOC (n = 84), conductivity (n = 33), pH (n = 33), iron content (n = 27), and oxygen content (n = 27) of investigated samples.

Groundwater DOM plays a crucial role in groundwater treatment, both in terms of quantity and quality (Schulz, 2020). Addressing a detailed characterization of groundwater DOM, 39 groundwaters were analyzed using the size exclusion chromatography method LC-OCD. Despite its limitations regarding its analysis duration (over 2 hours) and handling solid precipitation, this method can determine DOM from aquatic samples according to their molecular weight distribution (S. A. Huber et al., 2011). Latter especially is relevant with regard to the reduced character of groundwater. Reduced groundwater shows high reactivity between its containing iron content and possible oxygen exchange (Worch, 1997). Moreover, iron precipitation allows some DOM molecules to adsorb or accumulate onto iron flocs. This means that these corresponding DOM molecules cannot be analyzed by LC-OCD (Schulz, 2020). According to the oxygen content shown in Figure 6-5, it must therefore be considered that the DOM analysis may not be entirely correct for virtually all samples. Thus, the LC-OCD results regarding the quantification and qualification of DOM should be evaluated with caution. The results of the LC-OCD analysis were evaluated via ChromCalc software and are presented in Figure 6-6. It shows the box-plot results of each fraction (biopolymers, humic substances – HS, building blocks, low molecular-weight acids, and neutrals) of the analyzed groundwater.

HS represent the largest proportion of groundwater DOM, followed by building blocks and low molecular-weight neutrals and a virtually negligible proportion of biopolymers and low molecular-weight acids, respectively. The median of HS is approximately 1.2 mg L^{-1} although apparently there are some samples indicating rather high HS concentrations. Since HS represent the main fraction of DOM, a strong positive correlation between TOC and HS was confirmed (r = 0.99, Figure II-5). The strong positive correlation between LC-OCD and TOC analysis can be explained due to the similar method of carbon quantification through the detection of generated CO_2 (Baghoth et al., 2011).

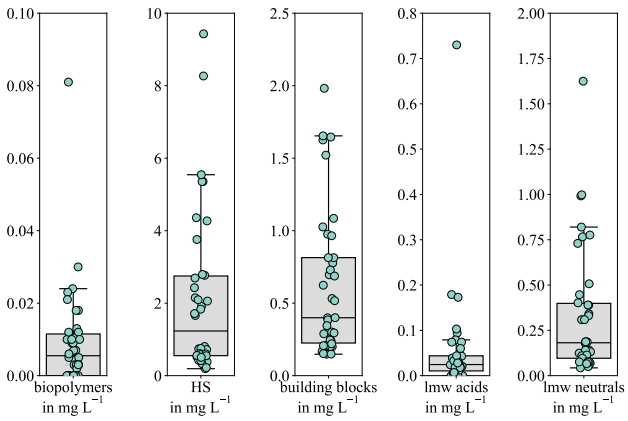

Figure 6-6: DOM characterization of groundwater samples (n = 38) via LC-OCD analysis. DOM is divided into introduced fractions: biopolymers, humic substances (HS), building blocks, low molecular-weight (lmw) acids, and low molecular-weight (lmw) neutrals.

The median proportion of HS in TOC is approximately 55 %, while the range between Q1 and Q3 (IQR) is between 47 and 65 % (Figure II-6). This confirms the findings of Teichmann (2014), according to whom approximately 50 % of the TOC in aquifers consists of HS. However, the varying proportions of individual samples highlight the limitation of only measuring TOC as an estimation of HS content in groundwater.

LC-OCD on the other hand shows the opportunity for detailed DOM characterization. However, it is known as a quite expensive and relative time-consuming method. Additionally, the differentiation between HS fractions, humic acids and fulvic acids, is challenging. Although there are IHSS standards to calibrate the system appropriately (IHSS, 2024), humic acids and fulvic acids appear in the same peak region which makes it almost impossible to identify within natural waters.

The mentioned limitations led to the introduction of PARAFAC fluorescence spectroscopy to characterize groundwater. The rapid measurement and the possibility of anaerobic sampling in an online monitoring setup make this method attractive for future tasks.

6.2.2. A Groundwater Three-Component PARAFAC Model

Contrary to the overall six-component PARAFAC model outlined in chapter 4.1, *base data* samples were not included in groundwater modelling. In the current chapter, a rather simple approach was followed, which was not designed to describe possible external influences on water quality (overall six-component PARAFAC model), but focused on the characterization of humic substances in the groundwater of the corresponding wells. In total, 158 samples were used for PARAFAC modelling consisting of both natural anaerobe sampled groundwaters (n = 135) and model solutions (n = 23) containing HS standards from IHSS (IHSS, 2024). While the DOM contents of natural groundwaters were unknown, model solutions in low- to mid-range TOC (0–5 mg L^1) were prepared and added to stabilize the PARAFAC modelling. Model solutions belong to the group of SRHA, SRFA, SRNOM, and HSNOM.

According to the procedure described in chapter 2.2.3 and 3.2.1 and by Murphy et al. (2013) and Wünsch (2024), the PARAFAC model was evaluated through iteration loops. Similar to the evaluation process in chapter 4.1, analysis of spectral loadings, sum of squared errors, and core consistency were applied as main criteria to determine the appropriate number of components. According to the evaluation process, generating a six-component PARAFAC model was not feasible to determine the humic substances character of groundwater samples. Consequently, successful validation of the PARAFAC model was performed by a three-component model, hereinafter referred to as the groundwater three-component PARAFAC model. The EEM of each component of the groundwater three-component PARAFAC model is presented in Figure 6-7. The MATLAB code revealing these components was basically the same as for chapter 4.1 and is attached to appendix (Figure I-5), although with adjustments regarding the number of components.

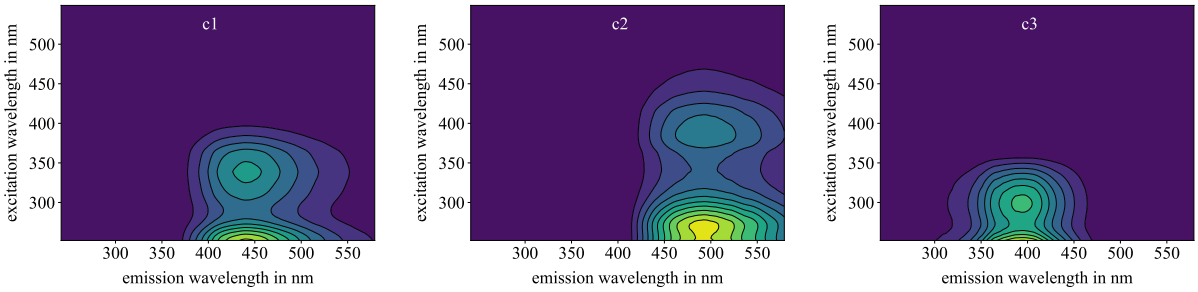

Figure 6-7: Fluorescence spectroscopy EEM fingerprints of each component of the groundwater three-component PARAFAC model.

Since there are some similarities and cross-connections with chapter 4.1, the discussion here will be kept brief. A rough comparison between the groundwater three-component PARAFAC model and the overall six-component PARAFAC model reveals the missing characteristic of the protein-like C6 component in groundwater. However, the spectral properties of c3 are still in the transition range of C6 spectral properties. Regarding the remaining components, it seems that c1, c2, and c3 somehow combine components C1–C5.

The already introduced dependency between emission peak and component size can also be applied to c1, c2, and c3. A shift of the emission peak towards higher wavelengths of emitted light can be correlated to an increase in the molar weight of the particular component (Wagner, 2014). According to DOM molecular weight definitions by Thurman et al. (1982) this would indicate c2 as rather humic acid-like, while c1 and c3 would be as rather fulvic acid-like compounds. To review these statements, the fluorescence intensity scores of SRFA and SRHA were investigated applying the groundwater three-component model and are presented in appendix (Figure II-7). It was shown that SRFA had the highest fluorescence intensity of c1, whereas SRHA showed the highest fluorescence intensity of c2. However, all three components were identified within both HS standards. This shows the complexity of clear differentiation between fulvic acid-like and humic acid-like compounds via fluorescence spectroscopy.

Similar to chapter 4.1 a literature comparison of the compounds was performed using the *OpenFluor* database (Murphy et al., 2014b). The results are briefly summarized in Table 6-2 and described subsequently.

Table 6-2: Overview of groundwater three-component PARAFAC model including spectral peak wavelengths, substance or substance-like designations, main occurrence in natural and engineered waters, and corresponding references for all components. The comparison with the literature was performed via the *OpenFluor* online library (Murphy et al., 2014b).

component	λ_{ex} peak in nm	λ_{em} peak in nm	substance	occurrence	references
c1	252/345	440	humic-like, fulvic acid-like, terrestrial	river water groundwater drinking water	[1][2] [3] [4]
c2	252/390	500	humic-like, humic acid-like, terrestrial	river water groundwater drinking water	[1][2] [3] [4]
c3	252/300	400	humic-like, autochthonous, terrestrial and microbial	river water seawater groundwater drinking water	[5] [6] [7] [3] [4]

[1] Lambert et al. (2016), [2] Murphy et al. (2014a), [3] Gabrielli et al. (2023), [4] Wells et al. (2022), [5] Peleato et al. (2017), [6] Fellman et al. (2010), [7] Zhou et al. (2019)

PARAFAC components c1 and c2 were identified in several studies examining river water (Lambert et al., 2016; Murphy et al., 2014a), groundwater (Gabrielli et al., 2023), and drinking water (Wells et al., 2022). According to these references, c1 represents fulvic acid-like compounds, with rather aliphatic and lower molecular weight than c2. The terrestrial character of c1 was emphasized, partly because the emission spectrum indicates substances formed during lignin degradation (Murphy et al., 2014a). All references characterize c2 with a highly degraded aromatic organic matter content of terrestrial origin and a rather humic acid-like composition. Reviewing the literature listed makes characterization of component c3 more controversial. Peleato et al. (2017) described c3 as a terrestrial humic-like component while Gabrielli et al. (2023) described c3 as similar to protein-like components, potentially derived from microbial processes. Considering the remaining references, c3 might represent a transition compound with humic-like and microbial protein-like characteristics.

In this study, groundwater is the source water, and its DOM constitution affects drinking water quality. Thus, the characterization of groundwater is essential when assessing drinking water quality. The detailed characterization of groundwater samples was done via the three-component PARAFAC model and the respective fluorescence intensity scores of the investigated groundwater samples are visualized in Figure 6-8.

The investigated groundwater samples showed a majority of c1 fluorescence followed by c3 and c2 fluorescence, respectively. The confirms the work of Bernal et al. (2018) and Huang et al. (2015) who also reported c1 and c3 as the components with the largest share in groundwater. With regard to c1 representing fulvic acid-like compounds, this result is consistent with the

general statement that fulvic acid-like components make up the largest fraction of aquatic humic substances, especially in groundwater (Schulz, 2020; Thurman et al., 1982). Due to a missing protein-like component, c3 is thought to play a role when it comes to changes in the bacterial constitution of groundwater. However, a correlation between the fluorescence score of c3 and bacterial constitution was not investigated in this work but should be considered regarding future application.

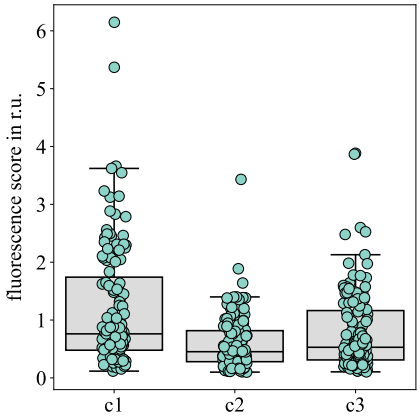

Figure 6-8: Fluorescence scores of respective three components (c1, c2, and c3) of investigated groundwater samples (n = 135).

Besides determining individual fluorescence intensity component scores of groundwater, another approach to characterizing DOM is to work with ratios between fluorescence component intensities. Shutova et al. (2014) and Philibert et al. (2022) highly recommended the monitoring of fluorescence component ratios before and after treatment to reveal information about both treatment efficiency considering DOM removal and/or transformation and the subsequent quality of drinking water. Shutova et al. (2014) even showed that monitoring the ratio between two humic-like components (e.g., here c1/c2) before the water treatment plant had a significant correlation with its treatability, especially applying coagulation-sedimentation and coagulation-membrane filtration techniques. Furthermore, Philibert et al. (2022) showed that a high ratio of c2/c3 leads to strong DOC removal by means of coagulation and ozonation, while a low ratio favors the use of ion exchangers.

In addition to the characterization of the DOM through PARAFAC fluorescence spectroscopy, a possible quantification of the DOM was also investigated in this work. Especially the quantification of HS plays a significant role due to their large proportion in groundwater (Figure 6-6). Due to the mentioned deviations in HS content in groundwaters, DOM quantification based solely on HS content can be incorrect. Since PARAFAC fluorescence spectroscopy generates fluorescence scores in Raman units of each component instead of accurate concentrations, quantification of natural water's DOM is generally challenging using this method (W. Chen & Yu, 2021). Hence, the results of both fluorescence spectroscopy and LC-OCD analysis were combined to reveal correlations between PARAFAC component intensities and HS content quantified by LC-OCD. The highest correlation was achieved for the c1 component (r = 0.97),

which also represents the major compound in groundwater (fulvic acid-like, Figure 6-8). The correlation between HS and c1 is presented in Figure 6-9, while the correlations between HS and c2, and c3, respectively, are attached to appendix (Figure II-8).

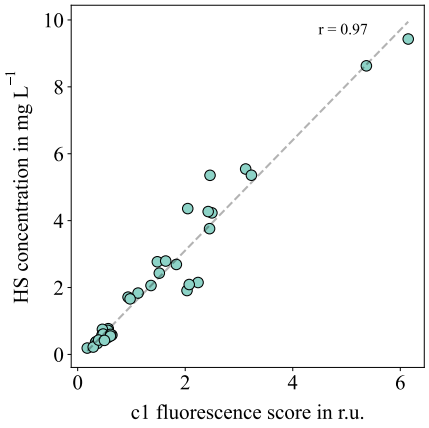

Figure 6-9: Correlation between HS concentration (LC-OCD) and fluorescence score c1 (groundwater three-component PARAFAC model). Includes all groundwater samples analyzed using both methods (n = 38).

Figure 6-9 enables the assumption of HS concentration in groundwater, using the introduced groundwater three-component PARAFAC model. Based on the presented data, some uncertainties should be mentioned. A majority of the investigated groundwaters lie in low HS ranges (< 1 mg L^{-1}) which might affect the accuracy of groundwaters containing high levels of DOM. Based on the results in Figure 6-9, another challenge should also be highlighted, namely correct anaerobic sampling and measurement. Some of the investigated samples already had a slight appearance of reddish precipitates immediately after sampling, which could be related the previously discussed oxygen-iron interactions. Precipitations can significantly affect fluorescence spectroscopy in different ways. On the one hand non-organic particles cause scattering and absorption which might lead to attenuation and redirection of the light and can thus interfere with the fluorescence detector signals (Hulst, 1981). On the other hand, iron precipitates may offer an adsorption surface for organic substances, which can therefore no longer be detected as DOC (Schulz, 2020). Both events can ultimately lead to an incorrect fluorescence score when performing the subsequent PARAFAC analysis. Analyzed samples in which low precipitation was observed can be associated with samples showing an increased deviation from the trend line in Figure 6-9. This emphasizes the importance of accurate sampling under anaerobic conditions to reveal conclusive results.

Respective correlations between PARAFAC components and LC-OCD results were already investigated by Baghoth et al. (2011), Peleato et al. (2016) and Rutlidge et al. (2015), although investigating mainly surface waters. Peleato et al. (2016) examined river water samples with relatively high DOC levels between 4 and 6 mg L^{-1}. Three components which can be compared to c1–c3 of this work, showed a strong positive correlation with LC-OCD characterized HS. Contrary to this work, Peleato et al. (2016) highlighted c2 as strongly correlated. A similar strong correlation between HS and c2 was observed by Rutlidge et al. (2015) and Baghoth et

al. (2011), although investigating cave drip water and treated surface water, respectively. Comparing this literature review with results from the present work generally shows the distinction between groundwater DOM, which tends to be dominated by fulvic acid-like components, and surface water, which has a higher proportion of humic acid-like components.

In order to facilitate future application of the groundwater three-component PARAFAC model in groundwater DOM characterization, appropriate consistency and stability must be ensured. In this work, all samples were merged to generate the PARAFAC model. However, a generally valid data set would be helpful in the future. This known data set could be utilized to characterize unknown samples, just like it was done for drinking water in chapter 4.1 using *base data*. However, this requires a more diverse data set, especially in regions with high fluorescence intensity or HS concentration regions. Otherwise, an exact characterization could not be guaranteed by investigating an unknown sample containing an unusually large amount of DOM. To stabilize the groundwater three-component PARAFAC model, it would be beneficial to include mid- and high-level DOM containing groundwater samples as well as the supporting DOM containing HS standards.

A stable and consistent groundwater three-component PARAFAC model would mean that it can generally be applied in all similar groundwater investigations, ensuring a general comparability between the groundwater samples. However, comparability should be limited to an area with similar aquifer characteristics, such as a certain catchment area of a municipal water supplier. Furthermore, a solid data set enables the automated online application of the groundwater three-component PARAFAC model such as described for drinking water in chapter 5.3.

Highlights Chapter 6.2

- Within a large-scale measurement campaign (n = 135) groundwater samples within a supply area in Northern Germany were successfully characterized through PARAFAC fluorescence spectroscopy.
- The generation of a simple groundwater three-component PARAFAC model provided the opportunity of describing each groundwater individually through each component's fluorescence score. This focused on groundwater containing humic substances. The groundwater three-component PARAFAC model provides additional value to DOC characterization by TOC or LC-OCD analyzers.
- c1 and c2 represent fulvic acid-like and humic acid-like DOM compounds, respectively, while c3 represents a transition compound with humic-like and microbial protein-like characteristics (general mass proportion in groundwater: c1 > c3 > c2).
- A strong correlation (r = 0.97) between HS concentration (via LC-OCD) and c1 (via fluorescence spectroscopy groundwater three-component PARAFAC model) was revealed.
- The groundwater three-component PARAFAC model will benefit from a larger (samples with increased DOC) data set, thus the consistency of the approach can be extended to other applications.

7. Conclusions and Outlook

7.1. Conclusions of this Thesis

Like other industrial sectors, drinking water utilities are undergoing a transformation towards automation and digitalization of water supply and water quality control. Currently, several established methods for determining drinking water quality are specified in the Germany drinking water ordinance (TrinkwV, 2023). However, conventional analytical methods for determining the bacterial constitution of drinking water are time- and labor-intensive and thus challenging to transfer to an automated monitoring system. In this context, there is also a lack of established methods for assessing the detailed characterization of dissolved organic matter (DOM) in drinking water. Monitoring present bacteria and DOM constitution in near real-time is desirable to react appropriately to changes and exclude a health risk for end consumers. Thus, a future monitoring system should be able to reliably detect and characterize the slightest changes in water quality.

The present work addresses the aforementioned challenges through the evaluation, further development and adaptation of innovative fluorescence spectroscopy coupled with parallel factor analysis (PARAFAC) modelling and flow cytometry. The capability of generating rapid and consistent DOM and bacterial fingerprints makes this approach promising for near real-time analysis. The following sections highlight the main conclusions of the three previously defined research objects of this work.

Further development of PARAFAC fluorescence spectroscopy and flow cytometry.

Groundwater-based drinking waters were investigated in this work, with samples taken from waterworks exits and different locations within the water distribution system (WDS). Through PARAFAC fluorescence spectroscopy and flow cytometry, the water was successfully characterized for DOM and bacteria on an individual fingerprinting basis. A sophisticated six-component PARAFAC model (C1–C6) was proven to describe the DOM of the sampled drinking water appropriately. The model was developed using the introduced *base data* set that could be transferred to compare with diverse other water samples in the future. The determination of respective waters' fluorescence scores revealed that the DOM of drinking water mainly consist of terrestrial and microbial humic-like or fulvic acid-like compounds, while humic acid-like and protein-like compounds were present in comparatively low concentrations. Flow cytometry allowed the characterization of present bacteria via total cell count (TCC) as well as the proportion of respective high nucleic acid cells (%HNA). A strong positive correlation was found between total organic carbon (TOC) amount and TCC (r = 0.93) as well as HNA cells (r = 0.94) and between TOC and PARAFAC components C1 (r =0.96), C3 (r = 0.96) as well as C4 (r = 0.95).

In laboratory experiments, contamination scenarios or significant changes in organic and bacterial water characteristics were simulated in a perfectly homogenized system. This was done to evaluate the sensitivity of all parameters. Therefore, one specific drinking water was monitored continuously and spiked with increasing amounts of different waters, such as wastewater effluent, rainwater, and varying drinking water. An evolving change in the monitored organic or bacterial characteristics was counted as recognized if consecutive measurements were outside the defined baseline (three time the standard deviation). It was found that fluorescence spectroscopy and flow cytometry could detect simulated changes in near real-time. Higher deviating waters, e.g., were detected even in lower spiking amounts. It must be noted that the ability to detect changes is strongly dependent on the characteristics of both monitored water and spiked water. In addition, it depends on the spiking proportion. C1–C3 as well as TCC were recognized as the most performant parameters regarding the detection of organic or bacterial characteristic changes, respectively.

Automatization of PARAFAC fluorescence spectroscopy and flow cytometry.

The further development of both hardware and software enables automated measurement of PARAFAC fluorescence spectroscopy and flow cytometry including multivariate data analysis and visualization. The developed setup was successfully validated in a laboratory environment performing different contamination scenarios as previously described. Both software applications were merged in a combined MATLAB and Python environment, and they were uploaded to GitHub along with an exemplary data set ("https://github.com/waterjonas/drinking-water-quality-online-monitoring/tree/main"). The expanded PARAFAC analysis is based on the established open-source drEEM toolbox, published by Murphy et al. (2013) and revised by Wünsch (2024). A data set (*base data*) of water samples' fluorescence data was determined to guarantee a consistent PARAFAC model regarding the repetitive surveillance of the same six components of online monitored drinking water which allows an appropriate comparison. In the context of a long-term monitoring scenario for drinking water, the possibility of automatic adjustment of the drinking water quality baseline is also addressed through the developed environment. Furthermore, the visualization of multivariate data allows rapid identification of outliers and evolving water quality changes. The entire process from automated sampling to visualization can be implemented in near real-time (< 15 min).

Transfer to real-world scenarios.

The first part introduces an initial approach to integrating both methods into a pilot plant pipe system. This is to monitor organic and bacterial changes in drinking water quality in a real-world setting. PARAFAC fluorescence spectroscopy and flow cytometry demonstrated comprehensively high sensitivity and fast results. Thus, the combination of both methods was shown to be a valuable monitoring system to characterize DOM and present bacteria in drinking water in a flowing system. Similar to the laboratory tests, different waters were injected into the flowing pipe system over a short period of time. This was to simulate a contamination case

or a characteristic change of drinking water quality. The induced change of organic and bacterial water characteristics was detectable for strongly varying injected waters, such as wastewater effluent and rainwater, until the end of the experiments (approximately 2.5 h and 300 m pipe distance). However, the detection was demonstrated to be contingent not only on the water quality characteristics of the flowing and induced water, but also on external factors such as hydraulic conditions.

The second part was a large-scale measurement campaign (n = 135) of groundwater samples within a supply area in Northern Germany. It aimed at identifying and characterizing changes in humic substances in the groundwater of the corresponding wells using PARAFAC fluorescence spectroscopy. Therefore, a simpler three-component PARAFAC model (c1–c3), compared to the six-component PARAFAC model, was generated and validated. This model could determine and quantify fulvic acid-like, humic acid-like and transition of humic-like and microbial protein-like compounds in groundwater. A strong correlation (r = 0.97) between the concentration of humic substances (HS), analyzed via LC-OCD, and the fluorescence score of fulvic acid-like compound (c1), emphasized the opportunity to estimate the concentration of HS through the application of PARAFAC fluorescence spectroscopy.

7.2. A Glimpse into the Future

This study presents an evaluation of the application of PARAFAC fluorescence spectroscopy and flow cytometry for automated and continuous drinking water monitoring. The performed experiments provided new knowledge regarding DOM and bacterial fingerprinting in laboratory and pilot plant environments.

However, a legitimate question is how these methods can be implemented in real WDS, especially in large urban areas. The potential for detecting characteristic changes of DOM and present bacteria 300 m behind the injection point was discussed. However, it must be debated how this is suitable for larger WDS with more than 5000 km of pipe system, as in Hamburg. Moreover, long-term monitoring at different waterworks and within the WDS should be considered to monitor real water quality changes due to seasonal and daily changes, or changing well circuits.

A further point to consider is the current lack of direct limit values for the fluorescence scores of individual PARAFAC components as well as for flow cytometry parameters. One promising approach is the specific definition of dynamic baselines for corresponding monitoring stations, which in turn must be defined through long-term investigations. The influence of the earlier described natural changes in water quality on each parameter baseline must be analyzed.

It must be noted that all parameters determined in this work should only be considered as an extension to existing water quality monitoring, as they cannot replace the detection of specific

pathogens, for example. Nevertheless, the potentially high measurement frequency, rapid analysis, and detailed characterization are providing added value compared to conventional laboratory-based methods.

Finally, the transition towards digitalization of the water sector requires automated data analysis. It is anticipated that simulation tools for drinking water quality prediction in WDSs will assume a more prominent role. In particular, multivariate data, such as those generated by fluorescence spectroscopy and flow cytometry, enable the enhancement of such water quality models as well as the training of simulation tools. In order to bridge the gap towards a fully digitalized WDS, it is essential to address the interfaces between water quality monitoring, simulation of water quality, and the corresponding automatic responses in case of an evolving event. The combined setup of PARAFAC fluorescence spectroscopy and flow cytometry developed in the present work offers suitable conditions for this purpose.

References

Allen, M. J., Edberg, S. C., & Reasoner, D. J. (2004). Heterotrophic plate count bacteria—What is their significance in drinking water? International Journal of Food Microbiology, 92(3), 265–274. https://doi.org/10.1016/j.ijfoodmicro.2003.08.017

Andersen, C. M., & Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. Journal of Chemometrics: A Journal of the Chemometrics Society, 17(4), 200–215. https://doi.org/10.1002/cem.790

Anderson, L. E., DeMont, I., Dunnington, D. D., Bjorndahl, P., Redden, D. J., Brophy, M. J., & Gagnon, G. A. (2023). A review of long-term change in surface water natural organic matter concentration in the northern hemisphere and the implications for drinking water treatment. Science of The Total Environment, 858, 159699. https://doi.org/10.1016/j.scitotenv.2022.159699

Andersson, C. A., & Bro, R. (2000). The N-way Toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems, 52(1), 1–4. https://doi.org/10.1016/S0169-7439(00)00071-X

Arad, J., Housh, M., Perelman, L., & Ostfeld, A. (2013). A dynamic thresholds scheme for contaminant event detection in water distribution systems. Water Research, 47(5), 1899–1908. https://doi.org/10.1016/j.watres.2013.01.017

Arkesteijn, G. J. A., Lozano-Andrés, E., Libregts, S. F. W. M., & Wauben, M. H. M. (2020). Improved Flow Cytometric Light Scatter Detection of Submicron-Sized Particles by Reduction of Optical Background Signals. Cytometry Part A, 97(6), 610–619. https://doi.org/10.1002/CYTO.A.24036

Baghoth, S. A., Sharma, S. K., & Amy, G. L. (2011). Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation–emission matrices and PARAFAC. Water Research, 45(2), 797–809. https://doi.org/10.1016/j.watres.2010.09.005

Bahram, M., Bro, R., Stedmon, C., & Afkhami, A. (2006). Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. Journal of Chemometrics, 20(3–4), 99–105. https://doi.org/10.1002/cem.978

Baker, A. (2001). Fluorescence Excitation−Emission Matrix Characterization of Some Sewage-Impacted Rivers. Environmental Science & Technology, 35(5), 948–953. https://doi.org/10.1021/es000177t

Banna, M. H., Imran, S., Francisque, A., Najjaran, H., Sadiq, R., Rodriguez, M., & Hoorfar, M. (2014). Online Drinking Water Quality Monitoring: Review on Available and Emerging Technologies. Critical Reviews in Environmental Science and Technology, 44(12), 1370–1421. https://doi.org/10.1080/10643389.2013.781936

Barbieri, M., Barberio, M. D., Banzato, F., Billi, A., Boschetti, T., Franchini, S., Gori, F., & Petitta, M. (2023). Climate change and its effect on groundwater quality. Environmental Geochemistry and Health, 45(4), 1133–1144. https://doi.org/10.1007/s10653-021-01140-5

Barsotti, F., Ghigo, G., & Vione, D. (2016). Computational assessment of the fluorescence emission of phenol oligomers: A possible insight into the fluorescence properties of humic-like substances (HULIS). Journal of Photochemistry and Photobiology A: Chemistry, 315, 87–93. https://doi.org/10.1016/j.jphotochem.2015.09.012

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(9), Article 9. https://doi.org/10.3390/su12093768

Benecke, J. (2018). Gypsum Scaling During Reverse Osmosis Desalination Characterization and Effects of Natural Organic Matter [Doctoral Thesis]. Technische Universität Hamburg.

Bernal, S., Lupon, A., Catalán, N., Castelar, S., & Martí, E. (2018). Decoupling of dissolved organic matter patterns between stream and riparian groundwater in a headwater forested catchment. Hydrology and Earth System Sciences, 22(3), 1897–1910. https://doi.org/10.5194/hess-22-1897-2018

Berney, M., Vital, M., Hülshoff, I., Weilenmann, H.-U., Egli, T., & Hammes, F. (2008). Rapid, cultivation-independent assessment of microbial viability in drinking water. Water Research, 42(14), 4010–4018. https://doi.org/10.1016/j.watres.2008.07.017

Besmer, M. D., Weissbrodt, D. G., Kratochvil, B. E., Sigrist, J. A., Weyland, M. S., & Hammes, F. (2014). The feasibility of automated online flow cytometry for in-situ monitoring of microbial dynamics in aquatic ecosystems. Frontiers in Microbiology, 5. https://doi.org/10.3389/fmicb.2014.00265

Besmer, M. D., Sigrist, J. A., Props, R., Buysschaert, B., Mao, G., Boon, N., & Hammes, F. (2017). Laboratory-Scale Simulation and Real-Time Tracking of a Microbial Contamination Event and Subsequent Shock-Chlorination in Drinking Water. Frontiers in Microbiology, 8, 1900. https://doi.org/10.3389/fmicb.2017.01900

BMG. (2021). Bericht des Bundesministeriums für Gesundheit und des Umweltbundesamtes an die Verbraucherinnen und Verbraucher über die Qualität von Wasser für den menschlichen Gebrauch (Trinkwasser) in Deutschland (Bundesministerium Für Gesundheit Und Umweltbundesamt Bundesgesundheitsblatt 2013 56:1191-1215).

Boal, A. K., & Patsalis, F. I. (2017). Use of Sodium Thiosulfate to Quench Hypochlorite Solutions Prior to Chlorate Analysis. Journal AWWA, 109(10), E410–E415. https://doi.org/10.5942/jawwa.2017.109.0097

Botes, M., de Kwaadsteniet, M., & Cloete, T. E. (2013). Application of quantitative PCR for the detection of microorganisms in water. Analytical and Bioanalytical Chemistry, 405(1), 91–108. https://doi.org/10.1007/s00216-012-6399-3

Bouvier, T., Troussellier, M., Anzil, A., Courties, C., & Servais, P. (2001). Using light scatter signal to estimate bacterial biovolume by flow cytometry. Cytometry, 44(3), 188–194. https://doi.org/10.1002/1097-0320(20010701)44:3<188::AID-CYTO1111>3.0.CO;2-C

Bro, R. (1997). PARAFAC. Tutorial and applications. Chemometrics and Intelligent Laboratory Systems, 38(2), 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4

Bro, R., & Kiers, H. A. L. (2003). A new efficient method for determining the number of components in PARAFAC models. Journal of Chemometrics, 17(5), 274–286. https://doi.org/10.1002/cem.801

Brown, M., & Wittwer, C. (2000). Flow Cytometry: Principles and Clinical Applications in Hematology. Clinical Chemistry, 46(8), 1221–1229. https://doi.org/10.1093/clinchem/46.8.1221

Buysschaert, B., Vermijs, L., Naka, A., Boon, N., & De Gusseme, B. (2018). Online flow cytometric monitoring of microbial water quality in a full-scale water treatment plant. Npj Clean Water, 1(1), 16. https://doi.org/10.1038/s41545-018-0017-7

Cabral, J. P. S. (2010). Water Microbiology. Bacterial Pathogens and Water. International Journal of Environmental Research and Public Health, 7(10), 3657–3703. https://doi.org/10.3390/ijerph7103657

Carstea, E. M., Zakharova, Y. S., & Bridgeman, J. (2018). Online Fluorescence Monitoring of Effluent Organic Matter in Wastewater Treatment Plants. Journal of Environmental Engineering, 144(5), 04018021. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001360

Carstea, E. M., Popa, C. L., Baker, A., & Bridgeman, J. (2020). In situ fluorescence measurements of dissolved organic matter: A review. Science of The Total Environment, 699, 134361. https://doi.org/10.1016/j.scitotenv.2019.134361

Catalán, N., Pastor, A., Borrego, C. M., Casas-Ruiz, J. P., Hawkes, J. A., Gutiérrez, C., von Schiller, D., & Marcé, R. (2021). The relevance of environment vs. Composition on dissolved organic matter degradation in freshwaters. Limnology and Oceanography, 66(2), 306–320. https://doi.org/10.1002/lno.11606

Cattell, R. B. (1944). “Parallel proportional profiles” and other principles for determining the choice of factors by rotation. Psychometrika, 9(4), 267–283. https://doi.org/10.1007/BF02288739

Chandler, W. L., Yeung, W., & Tait, J. F. (2011). A new microparticle size calibration standard for use in measuring smaller microparticles using a new flow cytometer. Journal of Thrombosis and Haemostasis, 9(6), 1216–1224. https://doi.org/10.1111/j.1538-7836.2011.04283.x

Chen, W., & Yu, H.-Q. (2021). Advances in the characterization and monitoring of natural organic matter using spectroscopic approaches. Water Research, 190, 116759. https://doi.org/10.1016/j.watres.2020.116759

Chen, Z., Yu, T., Ngo, H. H., Lu, Y., Li, G., Wu, Q., Li, K., Bai, Y., Liu, S., & Hu, H. Y. (2018). Assimilable organic carbon (AOC) variation in reclaimed water: Insight on biological stability evaluation and control for sustainable water reuse. Bioresource Technology, 254(December 2017), 290–299. https://doi.org/10.1016/j.biortech.2018.01.111

Cheswick, R., Moore, G., Nocker, A., Hassard, F., Jefferson, B., & Jarvis, P. (2020). Chlorine disinfection of drinking water assessed by flow cytometry: New insights. Environmental Technology & Innovation, 19, 101032–101032. https://doi.org/10.1016/J.ETI.2020.101032

Clark, R. M., Panguluri, S., Nelson, T. D., & Wyman, R. P. (2017). Protecting Drinking Water Utilities from Cyberthreats. Journal AWWA, 109(2), 50–58. https://doi.org/10.5942/jawwa.2017.109.0021

Coble, P. G., Green, S. A., Blough, N. V., & Gagosian, R. B. (1990). Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. Nature, 348(6300), Article 6300. https://doi.org/10.1038/348432a0

Coble, P. G. (1996). Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Marine Chemistry, 51(4), 325–346. https://doi.org/10.1016/0304-4203(95)00062-3

Coble, P. G., Del Castillo, C. E., & Avril, B. (1998). Distribution and optical properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon. Deep Sea Research Part II: Topical Studies in Oceanography, 45(10), 2195–2223. https://doi.org/10.1016/S0967-0645(98)00068-X

Coble, P. G., Lead, J., Baker, A., Reynolds, D. M., & Spencer, R. G. M. (Eds.). (2014). Aquatic Organic Matter Fluorescence. Cambridge University Press. https://doi.org/10.1017/CBO9781139045452

Conmy, R. N., Del Castillo, C. E., Downing, B. D., & Chen, R. F. (2014). Experimental Design and Quality Assurance: In Situ Fluorescence Instrumentation. In P. G. Coble, J. Lead, A. Baker, D. M. Reynolds, & R. G. M. Spencer (Eds.), Aquatic Organic Matter Fluorescence (1st ed., pp. 190–230). Cambridge University Press. https://doi.org/10.1017/CBO9781139045452.010

Cypionka, H. (2010). Grundlagen der Mikrobiologie. Springer. https://doi.org/10.1007/978-3-642-05096-1

Danielsson, L. G. (1982). On the use of filters for distinguishing between dissolved and particulate fractions in natural waters. Water Research, 16(2), 179–182. https://doi.org/10.1016/0043-1354(82)90108-7

Dapkus, R., Fryar, A., Tobin, B., Byrne, D., Sarker, S., Bettel, L., & Fox, J. (2023). Utilization of Tryptophan-like Fluorescence as a Proxy for E. coli Contamination in a Mixed-Land-Use Karst Basin. Hydrology. https://doi.org/10.3390/hydrology10040074

De Oliveira, G. F., Bertone, E., Stewart, R. A., Awad, J., Holland, A., O'Halloran, K., & Bird, S. (2018). Multi-Parameter Compensation Method for Accurate In Situ Fluorescent Dissolved Organic Matter Monitoring and Properties Characterization. Water, 10(9), Article 9. https://doi.org/10.3390/w10091146

DeFrancesco, C., & Guéguen, C. (2021). Long-term Trends in Dissolved Organic Matter Composition and Its Relation to Sea Ice in the Canada Basin, Arctic Ocean (2007–2017). Journal of Geophysical Research: Oceans, 126(2), e2020JC016578. https://doi.org/10.1029/2020JC016578

Derrien, M., Shin, K.-H., & Hur, J. (2019). Assessment on applicability of common source tracking tools for particulate organic matter in controlled end member mixing experiments. Science of The Total Environment, 666, 187–196. https://doi.org/10.1016/j.scitotenv.2019.02.258

DIN EN 1484. (2019). Wasseranalytik—Anleitungen zur Bestimmung des gesamten gelösten organischen Kohlenstoffs (TOC) und des gelösten organischen Kohlenstoffs (DOC); Deutsche Fassung EN 1484:1997. https://dx.doi.org/10.31030/3042067

DIN EN ISO 6222. (1999). Wasserbeschaffenheit—Quantitative Bestimmung der kultivierbaren Mikroorganismen—Bestimmung der Koloniezahl durch Einimpfen in ein Nähragarmedium (ISO 6222:1999); Deutsche Fassung EN ISO 6222:1999. https://dx.doi.org/10.31030/8128078

DIN EN ISO 9308-1. (2017). Wasserbeschaffenheit—Zählung von Escherichia coli und coliformen Bakterien—Teil 1: Membranfiltrationsverfahren für Wässer mit niedriger Begleitflora (ISO 9308-1:2014 + Amd 1:2016); Deutsche Fassung EN ISO 9308-1:2014 + A1:2017. https://dx.doi.org/10.31030/2630794

Ding, S., Deng, Y., Bond, T., Fang, C., Cao, Z., & Chu, W. (2019). Disinfection byproduct formation during drinking water treatment and distribution: A review of unintended effects of engineering agents and materials. Water Research, 160, 313–329. https://doi.org/10.1016/j.watres.2019.05.024

Dragan, A. I., Pavlovic, R., McGivney, J. B., Casas-Finet, J. R., Bishop, E. S., Strouse, R. J., Schenerman, M. A., & Geddes, C. D. (2012). SYBR Green I: fluorescence properties and interaction with DNA. Journal of Fluorescence, 22(4), 1189–1199. https://doi.org/10.1007/s10895-012-1059-8

Drozdova, A. N., Krylov, I. N., Nedospasov, A. A., Arashkevich, E. G., & Labutin, T. A. (2022). Fluorescent signatures of autochthonous dissolved organic matter production in Siberian shelf seas. Frontiers in Marine Science, 9. https://doi.org/10.3389/fmars.2022.872557

Du, Y., Zhang, Q., Liu, Z., He, H., Lürling, M., Chen, M., & Zhang, Y. (2019). Composition of dissolved organic matter controls interactions with La and Al ions: Implications for phosphorus immobilization in eutrophic lakes. Environmental Pollution, 248, 36–47. https://doi.org/10.1016/j.envpol.2019.02.002

Du, Y., Lu, Y., Roebuck, J. A., Liu, D., Chen, F., Zeng, Q., Xiao, K., He, H., Liu, Z., Zhang, Y., & Jaffé, R. (2021). Direct versus indirect effects of human activities on dissolved organic matter in highly impacted lakes. Science of The Total Environment, 752, 141839. https://doi.org/10.1016/j.scitotenv.2020.141839

Dubber, D., & Gray, N. F. (2010). Replacement of chemical oxygen demand (COD) with total organic carbon (TOC) for monitoring wastewater treatment performance to minimize disposal of toxic analytical waste. Journal of Environmental Science and Health, Part A, 45(12), 1595–1600. https://doi.org/10.1080/10934529.2010.506116

DVGW. (2022). Auswirkungen des Klimawandels auf das Wasserdargebot Deutschlands—Überblick zu aktuellen Ergebnissen der deutschen Klimaforschung. DVGW Deutscher Verein des Gas- und Wasserfaches e.V.

Eberhard, O. (2022). Entwicklung einer Methode zur Charakterisierung von Huminstoffen in Grundwässern [Master's Thesis]. Technische Universität Hamburg.

Ebrahimi, D., Kennedy, D. F., Messerle, B. A., & Hibbert, D. B. (2008). High throughput screening arrays of rhodium and iridium complexes as catalysts for intramolecular hydroamination using parallel factor analysis. Analyst, 133(6), 817–822. https://doi.org/10.1039/b719501j

Efstratiou, A., Ongerth, J. E., & Karanis, P. (2017). Waterborne transmission of protozoan parasites: Review of worldwide outbreaks - An update 2011–2016. Water Research, 114, 14–22. https://doi.org/10.1016/j.watres.2017.01.036

Escalas, A., Droguet, M., Guadayol, J. M., & Caixach, J. (2003). Estimating DOC regime in a wastewater treatment plant by UV deconvolution. Water Research, 37(11), 2627–2635. https://doi.org/10.1016/S0043-1354(03)00063-0

Escobar, I. C., & Randall, A. A. (2001). Assimilable organic carbon (AOC) and biodegradable dissolved organic carbon (BDOC): Complementary measurements. Water Research, 35(18), 4444–4454. https://doi.org/10.1016/S0043-1354(01)00173-7

Farreny, R., Morales-Pinzón, T., Guisasola, A., Tayà, C., Rieradevall, J., & Gabarrell, X. (2011). Roof selection for rainwater harvesting: Quantity and quality assessments in Spain. Water Research, 45(10), 3245–3254. https://doi.org/10.1016/j.watres.2011.03.036

Favere, J., Buysschaert, B., Boon, N., & de Gusseme, B. (2020). Online microbial fingerprinting for quality management of drinking water: Full-scale event detection. Water Research, 170, 115353. https://doi.org/10.1016/j.watres.2019.115353

Favere, J., Waegenaar, F., Boon, N., & De Gusseme, B. (2021). Online microbial monitoring of drinking water: How do different techniques respond to contaminations in practice? Water Research, 202, 117387. https://doi.org/10.1016/j.watres.2021.117387

Favere, J. (2022). Biostability of drinking water: From online monitoring to microbial quality management [Doctoral Thesis]. Ghent University.

Fellman, J. B., D'Amore, D. V., Hood, E., & Boone, R. D. (2008). Fluorescence characteristics and biodegradability of dissolved organic matter in forest and wetland soils from coastal temperate watersheds in southeast Alaska. Biogeochemistry, 88(2), 169–184. https://doi.org/10.1007/s10533-008-9203-x

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(6), 2452–2462. https://doi.org/10.4319/lo.2010.55.6.2452

Fenchel, T. M., & Jørgensen, B. B. (1977). Detritus Food Chains of Aquatic Ecosystems: The Role of Bacteria. In M. Alexander (Ed.), Advances in Microbial Ecology (pp. 1–58). Springer US. https://doi.org/10.1007/978-1-4615-8219-9_1

Fernández-Pascual, E., Droz, B., O'Dwyer, J., O'Driscoll, C., Goslan, E. H., Harrison, S., & Weatherill, J. (2023). Fluorescent Dissolved Organic Matter Components as Surrogates for Disinfection Byproduct Formation in Drinking Water: A Critical Review. ACS ES&T Water, 3(8), 1997–2008. https://doi.org/10.1021/acsestwater.2c00583

Ferris, F. G., Szponar, N., & Edwards, B. A. (2021). Groundwater Microbiology: Vol. The Groundwater Project.

Filella, M. (2009). Freshwaters: Which NOM matters? Environmental Chemistry Letters, 7(1), 21–35. https://doi.org/10.1007/s10311-008-0158-x

Folkman, S. (2018). Water Main Break Rates In the USA and Canada: A Comprehensive Study. Mechanical and Aerospace Engineering Faculty Publications. https://digitalcommons.usu.edu/mae_facpub/174

Förster, J. (1999). Variability of roof runoff quality. Water Science and Technology, 39(5), 137–144. https://doi.org/10.1016/S0273-1223(99)00095-5

Frimmel, F. H., & Abbt-Braun, G. (2011). 3.01 - Sum Parameters: Potential and Limitations. In P. Wilderer (Ed.), Treatise on Water Science (pp. 3–29). Elsevier. https://doi.org/10.1016/B978-0-444-53199-5.00048-8

Gabrielli, M., Turolla, A., & Antonelli, M. (2021). Bacterial dynamics in drinking water distribution systems and flow cytometry monitoring scheme optimization. Journal of Environmental Management, 286, 112151. https://doi.org/10.1016/j.jenvman.2021.112151

Gabrielli, M., Pulcini, F., Barbesti, G., & Antonelli, M. (2023). Source to tap investigation of natural organic matter in non-disinfected drinking water distribution systems. Environmental Science: Water Research & Technology, 10(1), 128–143. https://doi.org/10.1039/D3EW00280B

Gao, H., & Zepp, R. G. (1998). Factors Influencing Photoreactions of Dissolved Organic Matter in a Coastal River of the Southeastern United States. Environmental Science & Technology, 32(19), 2940–2946. https://doi.org/10.1021/es9803660

Gjessing, E. T. (1976). Physical and chemical characteristics of aquatic humus. Ann Arbor Science.

Gonçalves-Araujo, R., Stedmon, C. A., Heim, B., Dubinenkov, I., Kraberg, A., Moiseev, D., & Bracher, A. (2015). From Fresh to Marine Waters: Characterization and Fate of Dissolved Organic Matter in the Lena River Delta Region, Siberia. Frontiers in Marine Science, 2. https://doi.org/10.3389/fmars.2015.00108

Graeber, D., Tenzin, Y., Stutter, M., Weigelhofer, G., Shatwell, T., von Tümpling, W., Tittel, J., Wachholz, A., & Borchardt, D. (2021). Bioavailable DOC: Reactive nutrient ratios control heterotrophic nutrient assimilation—An experimental proof of the macronutrient-access hypothesis. Biogeochemistry, 155(1), 1–20. https://doi.org/10.1007/s10533-021-00809-4

Groeneveld, M., Catalán, N., Einarsdottir, K., G. Bravo, A., & N. Kothawala, D. (2022). The influence of pH on dissolved organic matter fluorescence in inland waters. Analytical Methods, 14(13), 1351–1360. https://doi.org/10.1039/D1AY01702K

Grohmann, A. N., Jekel, M., Grohmann, A., Szewzyk, U., & Szewzyk, R. (2011). Wasser: Chemie, Mikrobiologie und nachhaltige Nutzung. In Wasser. De Gruyter. https://doi.org/10.1515/9783110213096

Guillemette, R., Harwell, M. C., & Brown, C. A. (2023). Metabolically active bacteria detected with click chemistry in low organic matter rainwater. PLOS ONE, 18(5), e0285816. https://doi.org/10.1371/journal.pone.0285816

Gunter, H., Bradley, C., Hannah, D. M., Manaseki-Holland, S., Stevens, R., & Khamis, K. (2023). Advances in quantifying microbial contamination in potable water: Potential of fluorescence-based sensor technology. WIREs Water, 10(1), e1622. https://doi.org/10.1002/wat2.1622

Guo, K., Wu, Z., Chen, C., & Fang, J. (2022). UV/Chlorine Process: An Efficient Advanced Oxidation Process with Multiple Radicals and Functions in Water Treatment. Accounts of Chemical Research, 55(3), 286–297. https://doi.org/10.1021/acs.accounts.1c00269

Guzinska, H. (2018). Physikalischer Modellversuch zur Bewertung von Spülverfahren in Trinkwasserleitungen [Master's Thesis]. Technische Universität Hamburg.

Hallberg, G. R. (1987). The impacts of agricultural chemicals on ground water quality. GeoJournal, 15(3), 283–295. https://doi.org/10.1007/BF00213456

Hallé, C., Huck, P. M., Peldszus, S., Haberkamp, J., & Jekel, M. (2009). Assessing the Performance of Biological Filtration As Pretreatment to Low Pressure Membranes for Drinking Water. Environmental Science & Technology, 43(10), 3878–3884. https://doi.org/10.1021/es803615g

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

Hammes, F., & Egli, T. (2005). New method for assimilable organic carbon determination using flow-cytometric enumeration and a natural microbial consortium as inoculum. Environmental Science and Technology, 39(9), 3289–3294. https://doi.org/10.1021/es048277c

Hammes, F., Salhi, E., Köster, O., Kaiser, H.-P., Egli, T., & von Gunten, U. (2006). Mechanistic and kinetic evaluation of organic disinfection by-product and assimilable organic carbon (AOC) formation during the ozonation of drinking water. Water Research, 40(12), 2275–2286. https://doi.org/10.1016/j.watres.2006.04.029

Hammes, F., & Egli, T. (2010). Cytometric methods for measuring bacteria in water: Advantages, pitfalls and applications. Analytical and Bioanalytical Chemistry, 397(3), 1083–1095. https://doi.org/10.1007/s00216-010-3646-3

Hammes, F., Broger, T., Weilenmann, H.-U., Vital, M., Helbing, J., Bosshart, U., Huber, P., Peter Odermatt, R., & Sonnleitner, B. (2012). Development and laboratory-scale testing of a

fully automated online flow cytometer for drinking water analysis. Cytometry Part A, 81A(6), 508–516. https://doi.org/10.1002/cyto.a.22048

Hansell, D. A., & Carlson, C. A. (2014). Biogeochemistry of marine dissolved organic matter (Vol. 2). Academic Press. https://doi.org/10.1016/C2012-0-02714-7

Harshman, R. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-model factor analysis. UCLA Working Papers in Phonetics, 16, 1–84. https://www.semanticscholar.org/paper/Foundations-of-the-PARAFAC-procedure%3A-Models-and-an-Harshman/c5b28cae82b14417f1250e58bb241367248e827d (2024, April 09)

Heibati, M., Stedmon, C. A., Stenroth, K., Rauch, S., Toljander, J., Säve-Söderbergh, M., & Murphy, K. R. (2017). Assessment of drinking water quality at the tap using fluorescence spectroscopy. Water Research, 125, 1–10. https://doi.org/10.1016/j.watres.2017.08.020

Heibati, M. (2019). Applications of natural organic matter optical properties for assessing drinking water disinfection and distribution [Doctoral Thesis]. Chalmers University of Technology.

Henderson, R. K., Baker, A., Murphy, K. R., Hambly, A., Stuetz, R. M., & Khan, S. J. (2009). Fluorescence as a potential monitoring tool for recycled water systems: A review. Water Research, 43(4), 863–881. https://doi.org/10.1016/j.watres.2008.11.027

Ho, J., Nocker, A., Bendinger, B., West, S., & Tiehm, A. (2020). Durchflusszytometrie als schnelle Detektionsmethode für Bakterien: Abschlussbericht. TZW.

Hoefel, D., Grooby, W. L., Monis, P. T., Andrews, S., & Saint, C. P. (2003). Enumeration of water-borne bacteria using viability assays and flow cytometry: A comparison to culture-based techniques. Journal of Microbiological Methods, 55(3), 585–597. https://doi.org/10.1016/S0167-7012(03)00201-X

Holland, J. F., Teets, R. E., Kelly, P. M., & Timnick, Andrew. (1977). Correction of right-angle fluorescence measurements for the absorption of excitation radiation. Analytical Chemistry, 49(6), 706–710. https://doi.org/10.1021/ac50014a011

Howell, D. C., Rogier, M., Yzerbyt, V., & Bestgen, Y. (1998). Statistical methods in human sciences. New York: Wadsworth, 721.

Hrudey, S. E., & Hrudey, E. J. (2007). Published Case Studies of Waterborne Disease Outbreaks—Evidence of a Recurrent Threat. Water Environment Research, 79(3), 233–245. https://doi.org/10.2175/106143006X95483

Huang, S., Wang, Y., Ma, T., Tong, L., Wang, Y., Liu, C., & Zhao, L. (2015). Linking groundwater dissolved organic matter to sedimentary organic matter from a fluvio-lacustrine aquifer

at Jianghan Plain, China by EEM-PARAFAC and hydrochemical analyses. Science of The Total Environment, 529, 131–139. https://doi.org/10.1016/j.scitotenv.2015.05.051

Huber, J. (2022). Fluoreszenzspektroskopische real-time Analyse – Entwicklung einer automatisierten Methode zur Detektion von anormalen Veränderungen in der Trinkwassermatrix mittels PARAFAC [Master's Thesis]. Technische Universität Hamburg.

Huber, S. A., Balz, A., Abert, M., & Pronk, W. (2011). Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography – organic carbon detection – organic nitrogen detection (LC-OCD-OND). Water Research, 45(2), 879–885. https://doi.org/10.1016/j.watres.2010.09.023

Huck, P. M. (1990). Measurement of Biodegradable Organic Matter and Bacterial Growth Potential in Drinking Water. Journal AWWA, 82(7), 78–86. https://doi.org/10.1002/j.1551-8833.1990.tb06995.x

Hulst, H. C. (1981). Light Scattering by Small Particles. John Wiley & Sons, Inc.

IHSS. (2024, February 10). 13C NMR Estimates of Carbon Distribution. https://humic-substances.org/13c-nmr-estimates-of-carbon-distribution-in-ihss-samples/ (2024, February 10)

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.)] (pp. 35–115). Intergovernmental Panel on Climate Change (IPCC). https://www.ipcc.ch/report/ar6/syr/

Ishii, S. K. L., & Boyer, T. H. (2012). Behavior of Reoccurring PARAFAC Components in Fluorescent Dissolved Organic Matter in Natural and Engineered Systems: A Critical Review. Environmental Science & Technology, 46(4), 2006–2017. https://doi.org/10.1021/es2043504

Islam, Md. A., Morton, D. W., Johnson, B. B., & Angove, M. J. (2020). Adsorption of humic and fulvic acids onto a range of adsorbents in aqueous systems, and their effect on the adsorption of other species: A review. Separation and Purification Technology, 247, 116949. https://doi.org/10.1016/j.seppur.2020.116949

Jie, L., Zhipeng, Z., Cuihong, C., Peilin, C., & Yingying, W. (2017). In-situ features of LNA and HNA bacteria in branch ends of drinking water distribution systems. Journal of Water Supply: Research and Technology-Aqua, 66(5), 300–307. https://doi.org/10.2166/aqua.2017.108

Jutaporn, P., Armstrong, M. D., & Coronell, O. (2020). Assessment of C-DBP and N-DBP formation potential and its reduction by MIEX® DOC and MIEX® GOLD resins using fluorescence spectroscopy and parallel factor analysis. Water Research, 172, 115460. https://doi.org/10.1016/j.watres.2019.115460

Kämmler, J. (2023). Ozonation-biofiltration for the production of drinking water from organic-rich groundwaters [Doctoral Thesis, Technische Universität Hamburg]. https://doi.org/10.15480/882.5208

Keesstra, S., Geissen, V., Mosse, K., Piiranen, S., Scudiero, E., Leistra, M., & van Schaik, L. (2012). Soil as a filter for groundwater quality. Current Opinion in Environmental Sustainability, 4(5), 507–516. https://doi.org/10.1016/j.cosust.2012.10.007

Khamis, K., Sorensen, J. P. R., Bradley, C., Hannah, D. M., Lapworth, D. J., & Stevens, R. (2015). In situ tryptophan-like fluorometers: Assessing turbidity and temperature effects for freshwater applications. Environmental Science: Processes & Impacts, 17(4), 740–752. https://doi.org/10.1039/C5EM00030K

Khamis, K., Bradley, C., Stevens, R., & Hannah, D. M. (2017). Continuous field estimation of dissolved organic carbon concentration and biochemical oxygen demand using dual-wavelength fluorescence, turbidity and temperature. Hydrological Processes, 31(3), 540–555. https://doi.org/10.1002/hyp.11040

Kida, M., Watanabe, I., Kinjo, K., Kondo, M., Yoshitake, S., Tomotsune, M., Iimura, Y., Umnouysin, S., Suchewaboripont, V., Poungparn, S., Ohtsuka, T., & Fujitake, N. (2021). Organic carbon stock and composition in 3.5-m core mangrove soils (Trat, Thailand). Science of The Total Environment, 801, 149682–149682. https://doi.org/10.1016/J.SCITOTENV.2021.149682

Kinsey, J. D., Corradino, G., Ziervogel, K., Schnetzer, A., & Osburn, C. L. (2018). Formation of Chromophoric Dissolved Organic Matter by Bacterial Degradation of Phytoplankton-Derived Aggregates. Frontiers in Marine Science, 4. https://doi.org/10.3389/fmars.2017.00430

Kördel, W., Dassenakis, M., Lintelmann, J., & Padberg, S. (1997). The importance of natural organic material for environmental processes in waters and soils (Technical Report). Pure and Applied Chemistry, 69(7), 1571–1600. https://doi.org/10.1351/pac199769071571

Korshin, G., Chow, C. W. K., Fabris, R., & Drikas, M. (2009). Absorbance spectroscopy-based examination of effects of coagulation on the reactivity of fractions of natural organic matter with varying apparent molecular weights. Water Research, 43(6), 1541–1548. https://doi.org/10.1016/j.watres.2008.12.041

Korth, A., & Donarth, O. (2015). Spülverfahren für Trinkwasserleitungen problemspezifisch auswählen. DVGW energie I wasser-praxis(1/2015), 36–41.

Kötzsch, S., Alisch, S., & Egli, T. (2012). Durchflusszytometrische Analyse von Wasserproben. Schweizerisches Bundesamt für Gesundheit (BAG).

Kötzsch, S., & Sinreich, M. (2014). Zellzahlen zum Grundwasser. Bestimmung mittels Durchflusszytometrie. Aqua & Gas, 94(3), 14–21.

Kubista, M., Sjöback, R., Eriksson, S., & Albinsson, B. (1994). Experimental correction for the inner-filter effect in fluorescence spectra. Analyst, 119(3), 417–419. https://doi.org/10.1039/AN9941900417

Kunkel, R., Hannappel, S., Voigt, H.-J., & Wendland, F. (2002). Die natürliche Grundwasserbeschaffenheit ausgewählter hydrostratigrafischer Einheiten in Deutschland [Endbericht eines FuE-Vorhabens im Rahmen des Länderfinanzierungsprogramms "Wasser und Boden" der Länderarbeitsgemeinschaft Wasser].

Kunkel, R. (Ed.). (2004). Die natürliche, ubiquitär überprägte Grundwasser-beschaffenheit in Deutschland. Forschungszentrum Jülich, Zentralbibliothek.

Lakowicz, J. R. (2006). Principles of fluorescence spectroscopy (3rd ed.). Springer. https://doi.org/10.1007/978-0-387-46312-4

Lambert, T., Teodoru, C. R., Nyoni, F. C., Bouillon, S., Darchambeau, F., Massicotte, P., & Borges, A. V. (2016). Along-stream transport and transformation of dissolved organic matter in a large tropical river. Biogeosciences, 13(9), 2727–2741. https://doi.org/10.5194/bg-13-2727-2016

Lapierre, J.-F., & del Giorgio, P. A. (2014). Partial coupling and differential regulation of biologically and photochemically labile dissolved organic carbon across boreal aquatic networks. Biogeosciences, 11(20), 5969–5985. https://doi.org/10.5194/bg-11-5969-2014

LeChevallier, M. W., Welch, N. J., & Smith, D. B. (1996). Full-scale studies of factors related to coliform regrowth in drinking water. Applied and Environmental Microbiology, 62(7), 2201–2211. https://doi.org/10.1128/aem.62.7.2201-2211.1996

LeChevallier, M. W., Gullick, R. W., Karim, M. R., Friedman, M., & Funk, J. E. (2003). The potential for health risks from intrusion of contaminants into the distribution system from pressure transients. Journal of Water and Health, 1(1), 3–14. https://doi.org/10.2166/wh.2003.0002

Leenheer, J. A., & Croué, J.-P. (2003). Characterizing Aquatic Dissolved Organic Matter. Environmental Science & Technology, 37(1), 18A-26A. https://doi.org/10.1021/es032333c

Lenhardt Acković, L., Zeković, I., Dramićanin, T., Bro, R., & Dramićanin, M. D. (2018). Modeling Food Fluorescence with PARAFAC (pp. 161–197). Springer, Cham. https://doi.org/10.1007/978-3-030-01569-5_8

Leresche, F., Vialykh, E. A., & Rosario-Ortiz, F. L. (2022). Computational Calculation of Dissolved Organic Matter Absorption Spectra. Environmental Science & Technology, 56(1), 491–500. https://doi.org/10.1021/acs.est.1c06252

Leveque, B., Burnet, J.-B., Dorner, S., & Bichai, F. (2021). Impact of climate change on the vulnerability of drinking water intakes in a northern region. Sustainable Cities and Society, 66, 102656. https://doi.org/10.1016/j.scs.2020.102656

Lewis, G., Taylor, I. W., Nienow, A. W., & Hewitt, C. J. (2004). The application of multi-parameter flow cytometry to the study of recombinant Escherichia coli batch fermentation processes. Journal of Industrial Microbiology and Biotechnology, 31(7), 311–322. https://doi.org/10.1007/s10295-004-0151-8

Li, Y. (2017). Comprehensive Characterization of Dissolved Organic Matter by Using Chemical Fractionation and High Resolution Organic Structural Spectroscopy [Doctoral Thesis]. Technische Universität München.

Lin, H., & Guo, L. (2020). Variations in Colloidal DOM Composition with Molecular Weight within Individual Water Samples as Characterized by Flow Field-Flow Fractionation and EEM-PARAFAC Analysis. Environmental Science and Technology, 54(3), 1657–1667. https://doi.org/10.1021/ACS.EST.9B07123/SUPPL_FILE/ES9B07123_SI_001.PDF

Liu, G., Van der Mark, E. J., Verberk, J. Q. J. C., & Van Dijk, J. C. (2013). Flow Cytometry Total Cell Counts: A Field Study Assessing Microbiological Water Quality and Growth in Unchlorinated Drinking Water Distribution Systems. BioMed Research International, 2013, e595872. https://doi.org/10.1155/2013/595872

Liu, W., Wu, H., Wang, Z., Ong, S. L., Hu, J. Y., & Ng, W. J. (2002). Investigation of assimilable organic carbon (AOC) and bacterial regrowth in drinking water distribution system. Water Research, 36(4), 891–898. https://doi.org/10.1016/S0043-1354(01)00296-2

López-Gálvez, J., Schiessl, K., Besmer, M. D., Bruckmann, C., Harms, H., & Müller, S. (2023). Development of an Automated Online Flow Cytometry Method to Quantify Cell Density and Fingerprint Bacterial Communities. Cells, 12(12), Article 12. https://doi.org/10.3390/cells12121559

Macey, M. G. (2007). Principles of Flow Cytometry. In M. G. Macey (Ed.), Flow Cytometry: Principles and Applications (pp. 1–15). Humana Press. https://doi.org/10.1007/978-1-59745-451-3_1

Madigan, M. T., Martinko, J. M., Stahl, D. A., & Clark, D. P. (2015). Brock Mikrobiologie kompakt (13th ed.). Pearson Deutschland.

Maie, N., Scully, N. M., Pisani, O., & Jaffé, R. (2007). Composition of a protein-like fluorophore of dissolved organic matter in coastal wetland and estuarine ecosystems. Water Research, 41(3), 563–570. https://doi.org/10.1016/j.watres.2006.11.006

Maiyo, J. K., Dasika, S., & Jafvert, C. T. (2023). Slow Sand Filters for the 21st Century: A Review. International Journal of Environmental Research and Public Health, 20(2), Article 2. https://doi.org/10.3390/ijerph20021019

Mantel, T. J. (2022). Elektrisch angesteuerte Gold-Polymer-Gold-Ultrafiltrationsmembranen zur Aufbereitung NOM-haltiger Wässer [Doctoral Thesis, Technische Universität Hamburg]. https://doi.org/10.15480/882.4287

Manti, A., Boi, P., Falcioni, T., Canonico, B., Ventura, A., Sisti, D., Pianetti, A., Balsamo, M., & Papa, S. (2008). Bacterial cell monitoring in wastewater treatment plants by flow cytometry. Water Environment Research: A Research Publication of the Water Environment Federation, 80(4), 346–354. https://doi.org/10.2175/106143007x221418

Masciopinto, C., De Giglio, O., Scrascia, M., Fortunato, F., La Rosa, G., Suffredini, E., Pazzani, C., Prato, R., & Montagna, M. T. (2019). Human health risk assessment for the occurrence of enteric viruses in drinking water from wells: Role of flood runoff injections. Science of The Total Environment, 666, 559–571. https://doi.org/10.1016/j.scitotenv.2019.02.107

Matilainen, A., Gjessing, E. T., Lahtinen, T., Hed, L., Bhatnagar, A., & Sillanpää, M. (2011). An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. Chemosphere, 83(11), 1431–1442. https://doi.org/10.1016/j.chemosphere.2011.01.018

Mazivila, S. J., Bortolato, S. A., & Olivieri, A. C. (2018). MVC3_GUI: A MATLAB graphical user interface for third-order multivariate calibration. An upgrade including new multi-way models. Chemometrics and Intelligent Laboratory Systems, 173, 21–29. https://doi.org/10.1016/j.chemolab.2017.12.012

McKay, G., Korak, Julie. A., & Rosario-Ortiz, F. L. (2018). Temperature Dependence of Dissolved Organic Matter Fluorescence. Environmental Science & Technology, 52(16), 9022–9032. https://doi.org/10.1021/acs.est.8b00643

McKay, G. (2020). Emerging investigator series: Critical review of photophysical models for the optical and photochemical properties of dissolved organic matter. Environmental Science: Processes & Impacts, 22(5), 1139–1165. https://doi.org/10.1039/D0EM00056F

Micó, P., García-Ballesteros, S., Mora, M., Vicente, R., Amat, A. M., & Arques, A. (2019). EEMlab: A graphical user-friendly interface for fluorimetry experiments based on the drEEM toolbox. Chemometrics and Intelligent Laboratory Systems, 188, 6–13. https://doi.org/10.1016/j.chemolab.2019.03.001

Mobed, J. J., Hemmingsen, S. L., Autry, J. L., & McGown, L. B. (1996). Fluorescence Characterization of IHSS Humic Substances: Total Luminescence Spectra with Absorbance Correction. Environmental Science & Technology, 30(10), 3061–3065. https://doi.org/10.1021/es960132l

Moon, S., Kwak, W., Lee, S., Kim, W., Oh, J., & Youn, S.-K. (2013). Epidemiological Characteristics of the First Water-Borne Outbreak of Cryptosporidiosis in Seoul, Korea. Journal of Korean Medical Science, 28(7), 983–989. https://doi.org/10.3346/jkms.2013.28.7.983

Moona, N., Murphy, K. R., Bondelind, M., Bergstedt, O., & Pettersson, T. J. R. (2018). Partial renewal of granular activated carbon biofilters for improved drinking water treatment. Environmental Science: Water Research & Technology, 4(4), 529–538. https://doi.org/10.1039/C7EW00413C

Moona, N., Holmes, A., Wünsch, U. J., Pettersson, T. J. R., & Murphy, K. R. (2021). Full-Scale Manipulation of the Empty Bed Contact Time to Optimize Dissolved Organic Matter Removal by Drinking Water Biofilters. ACS ES&T Water, 1(5), 1117–1126. https://doi.org/10.1021/acsestwater.0c00105

Moreira, N. A., & Bondelind, M. (2017). Safe drinking water and waterborne outbreaks. Journal of Water and Health, 15(1), 83–96. https://doi.org/10.2166/wh.2016.103

Müller, S., Harms, H., & Bley, T. (2010). Origin and analysis of microbial population heterogeneity in bioprocesses. Current Opinion in Biotechnology, 21(1), 100–113. https://doi.org/10.1016/j.copbio.2010.01.002

Murphy, K. R., Stedmon, C. A., Waite, T. D., & Ruiz, G. M. (2008). Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. Marine Chemistry, 108(1), 40–58. https://doi.org/10.1016/j.marchem.2007.10.003

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. Environmental Science and Technology, 44(24), 9405–9412. https://doi.org/10.1021/ES102362T/SUPPL_FILE/ES102362T_SI_001.PDF

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. Environmental Science and Technology, 45(7), 2909–2916. https://doi.org/10.1021/ES103015E/SUPPL_FILE/ES103015E_SI_001.PDF

Murphy, K. R., Stedmon, C. A., Graeber, D., & Bro, R. (2013). Fluorescence spectroscopy and multi-way techniques. PARAFAC. Analytical Methods, 5(23), 6557–6566. https://doi.org/10.1039/c3ay41160e

Murphy, K. R., Bro, R., & Stedmon, C. A. (2014a). Chemometric Analysis of Organic Matter Fluorescence. In A. Baker, D. M. Reynolds, J. Lead, P. G. Coble, & R. G. M. Spencer (Eds.), Aquatic Organic Matter Fluorescence (pp. 339–375). Cambridge University Press. https://doi.org/10.1017/CBO9781139045452.016

Murphy, K. R., Stedmon, C. A., Wenig, P., & Bro, R. (2014b). OpenFluor– an online spectral library of auto-fluorescence by organic compounds in the environment. Anal. Methods, 6(3), 658–661. https://doi.org/10.1039/C3AY41935E

Nescerecka, A., Hammes, F., & Juhna, T. (2016). A pipeline for developing and testing staining protocols for flow cytometry, demonstrated with SYBR Green I and propidium iodide viability staining. Journal of Microbiological Methods, 131, 172–180. https://doi.org/10.1016/j.mimet.2016.10.022

Niloy, N. M., Haque, Md. M., & Tareq, S. M. (2021). Characterization of dissolved organic matter at urban and industrial rainwater of Bangladesh by fluorescence spectroscopy and EEM-PARAFAC modeling. Environmental Challenges, 5, 100250. https://doi.org/10.1016/j.envc.2021.100250

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

Ohno, T. (2002). Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. Environmental Science and Technology, 36(4), 742–746. https://doi.org/10.1021/ES0155276

Osburn, C. L., Handsel, L. T., Peierls, B. L., & Paerl, H. W. (2016). Predicting Sources of Dissolved Organic Nitrogen to an Estuary from an Agro-Urban Coastal Watershed. Environmental Science & Technology, 50(16), 8473–8484. https://doi.org/10.1021/acs.est.6b00053

Otieno, F. A. O., Olumuyiwa, I. O., & Ochieng, G. M. (2012). Groundwater: Characteristics, qualities, pollutions and treatments : an overview. International Journal of Water Resources and Environmental Engineering, Vol 4(6), 162–170. https://doi.org/10.5897/IJWREE12.038

Paerl, R. W., Claudio, I. M., Shields, M. R., Bianchi, T. S., & Osburn, C. L. (2020). Dityrosine formation via reactive oxygen consumption yields increasingly recalcitrant humic-like fluorescent organic matter in the ocean. Limnology and Oceanography Letters, 5(5), 337–345. https://doi.org/10.1002/LOL2.10154

Park, J. W., Kim, H. C., Meyer, A. S., Kim, S., & Maeng, S. K. (2016). Influences of NOM composition and bacteriological characteristics on biological stability in a full-scale drinking water treatment plant. Chemosphere, 160, 189–198. https://doi.org/10.1016/J.CHEMOSPHERE.2016.06.079

Parker, C. A., & Barnes, W. J. (1957). Some experiments with spectrofluorimeters and filter fluorimeters. Analyst, 82(978), 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(12), 1765–1781. https://doi.org/10.1016/S0146-6380(00)00124-8

Peleato, N. M., McKie, M., Taylor-Edmonds, L., Andrews, S. A., Legge, R. L., & Andrews, R. C. (2016). Fluorescence spectroscopy for monitoring reduction of natural organic matter and halogenated furanone precursors by biofiltration. Chemosphere, 153, 155–161. https://doi.org/10.1016/j.chemosphere.2016.03.018

Peleato, N. M., Sidhu, B. S., Legge, R. L., & Andrews, R. C. (2017). Investigation of ozone and peroxone impacts on natural organic matter character and biofiltration performance using fluorescence spectroscopy. Chemosphere, 172, 225–233. https://doi.org/10.1016/j.chemosphere.2016.12.118

Philibert, M., Luo, S., Moussanas, L., Yuan, Q., Filloux, E., Zraick, F., & Murphy, K. R. (2022). Drinking water aromaticity and treatability is predicted by dissolved organic matter fluorescence. Water Research, 220, 118592. https://doi.org/10.1016/j.watres.2022.118592

Pick, F. C., Fish, K. E., Biggs, C. A., Moses, J. P., Moore, G., & Boxall, J. B. (2019). Application of enhanced assimilable organic carbon method across operational drinking water systems. PLOS ONE, 14(12), e0225477. https://doi.org/10.1371/journal.pone.0225477

Pick, F. C., Fish, K. E., & Boxall, J. B. (2021). Assimilable organic carbon cycling within drinking water distribution systems. Water Research, 198, 117147. https://doi.org/10.1016/j.watres.2021.117147

Prest, E. I., Hammes, F., Kötzsch, S., van Loosdrecht, M. C. M., & Vrouwenvelder, J. S. (2013). Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. Water Research, 47(19), 7131–7142. https://doi.org/10.1016/j.watres.2013.07.051

Prest, E. I., El-Chakhtoura, J., Hammes, F., Saikaly, P. E., van Loosdrecht, M. C. M., & Vrouwenvelder, J. S. (2014). Combining flow cytometry and 16S rRNA gene pyrosequencing: A promising approach for drinking water monitoring and characterization. Water Research, 63, 179–189. https://doi.org/10.1016/j.watres.2014.06.020

Prest, E. I., Hammes, F., van Loosdrecht, M. C. M., & Vrouwenvelder, J. S. (2016). Biological Stability of Drinking Water: Controlling Factors, Methods, and Challenges. Frontiers in Microbiology, 7, 45. https://doi.org/10.3389/fmicb.2016.00045

Pucher, M., Wünsch, U., Weigelhofer, G., Murphy, K., Hein, T., & Graeber, D. (2019). staRdom: Versatile Software for Analyzing Spectroscopic Data of Dissolved Organic Matter in R. Water, 11(11), Article 11. https://doi.org/10.3390/w11112366

Ramírez-Castillo, F. Y., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F. J., Harel, J., & Guerrero-Barrera, A. L. (2015). Waterborne Pathogens: Detection Methods and Challenges. Pathogens, 4(2), Article 2. https://doi.org/10.3390/pathogens4020307

Ratner, B. (2009). The correlation coefficient: Its values range between +1/−1, or do they? Journal of Targeting, Measurement and Analysis for Marketing, 17(2), 139–142. https://doi.org/10.1057/jt.2009.5

Reineke, W., & Schlömann, M. (2023). Environmental Microbiology. Springer. https://doi.org/10.1007/978-3-662-66547-3

Retelletti Brogi, S., Balestra, C., Casotti, R., Cossarini, G., Galletti, Y., Gonnelli, M., Vestri, S., & Santinelli, C. (2020). Time resolved data unveils the complex DOM dynamics in a Mediterranean river. Science of The Total Environment, 733, 139212. https://doi.org/10.1016/j.scitotenv.2020.139212

Rutlidge, H., Andersen, M. S., Baker, A., Chinu, K. J., Cuthbert, M. O., Jex, C. N., Marjo, C. E., Markowska, M., & Rau, G. C. (2015). Organic characterisation of cave drip water by LC-OCD and fluorescence analysis. Geochimica et Cosmochimica Acta, 166, 15–28. https://doi.org/10.1016/j.gca.2015.05.042

Sadler, M. C., Senouillet, J., Kuenzi, S., Grasso, L., & Watson, D. C. (2020). Computational Surveillance of Microbial Water Quality With Online Flow Cytometry. Frontiers in Water, 2, 586969. https://doi.org/10.3389/frwa.2020.586969

Safford, H. R., & Bischel, H. N. (2019). Flow cytometry applications in water treatment, distribution, and reuse: A review. Water Research, 151, 110–133. https://doi.org/10.1016/j.watres.2018.12.016

Santos, M., Oliveira, H., Pereira, J. L., Pereira, M. J., Gonçalves, F. J. M., & Vidal, T. (2019). Flow cytometry analysis of low/high DNA content (LNA/HNA) bacteria as bioindicator of water quality evaluation. Ecological Indicators, 103, 774–781. https://doi.org/10.1016/j.ecolind.2019.03.033

Schleich, C., Chan, S., Pullerits, K., Besmer, M. D., Paul, C. J., Rådström, P., & Keucken, A. (2019). Mapping dynamics of bacterial communities in a full-scale drinking water distribution system using flow cytometry. Water (Switzerland), 11(10). https://doi.org/10.3390/w11102137

Schnitzer, M., Kodama, H., & Ripmeester, J. A. (1991). Determination of the Aromaticity of Humic Substances by X-Ray Diffraction Analysis. Soil Science Society of America Journal, 55(3), 745–750. https://doi.org/10.2136/sssaj1991.03615995005500030018x

Schönher, C., Proksch, P., Kerschbaumer, D., Fiedler, C. J., Schmidt, B.-J., Keskinöz, C., Aguilar Gonzalez, A. E., Mayr, E., Perfler, R., & Zunabovic-Pichler, M. (2021). "Every cell counts"—Experiences with flow cytometry for Austrian drinking water supply. Österreichische Wasser- Und Abfallwirtschaft 2021, 1–11. https://doi.org/10.1007/S00506-021-00802-Z

Schreiner-McGraw, A. P., & Ajami, H. (2021). Delayed response of groundwater to multi-year meteorological droughts in the absence of anthropogenic management. Journal of Hydrology, 603, 126917. https://doi.org/10.1016/j.jhydrol.2021.126917

Schulz, M. (2020). Entfernung natürlicher organischer Stoffe durch die Verfahrenskombination Flockung-Ultrafiltration bei der Aufbereitung reduzierter Grundwässer [Doctoral Thesis, Technische Universität Hamburg]. https://doi.org/10.15480/882.2687

Schuster, J., Huber, J., Stumme, J., Grieb, A., & Ernst, M. (2022a). Combining real-time fluorescence spectroscopy and flow cytometry to reveal new insights in DOC and cell characterization of drinking water. Frontiers in Environmental Chemistry, 3, 931067. https://doi.org/10.3389/fenvc.2022.931067

Schuster, J., Kadinski, L., Cao, H., Abhijith, G. R., Grieb, A., Li, P., Ostfeld, A., & Ernst, M. (2022b). Real-Time Monitoring and Controlling of Water Quality in Water Distribution Networks Based on Flow Cytometry and Fluorescence Spectroscopy. World Environmental and Water Resources Congress 2022, 1155–1167. https://doi.org/10.1061/9780784484258.108

Schuster, J. (2024, April 9). GitHub—Drinking-water-quality-online-monitoring [GitHub Upload]. https://github.com/waterjonas/drinking-water-quality-online-monitoring (2024, April 09)

Sciscenko, I., Arques, A., Micó, P., Mora, M., & García-Ballesteros, S. (2022). Emerging applications of EEM-PARAFAC for water treatment: A concise review. Chemical Engineering Journal Advances, 10, 100286. https://doi.org/10.1016/j.ceja.2022.100286

Senesi, N. (1990). Molecular and quantitative aspects of the chemistry of fulvic acid and its interactions with metal ions and organic chemicals: Part II. The fluorescence spectroscopy approach. Analytica Chimica Acta, 232, 77–106. https://doi.org/10.1016/S0003-2670(00)81226-X

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. Journal of Bacteriology, 189(23), 8746–8749. https://doi.org/10.1128/JB.01368-07

Shutova, Y., Baker, A., Bridgeman, J., & Henderson, R. K. (2014). Spectroscopic characterisation of dissolved organic matter changes in drinking water treatment: From PARAFAC analysis to online monitoring wavelengths. Water Research, 54, 159–169. https://doi.org/10.1016/j.watres.2014.01.053

Shutova, Y., Baker, A., Bridgeman, J., & Henderson, R. K. (2016). On-line monitoring of organic matter concentrations and character in drinking water treatment systems using fluorescence spectroscopy. Environmental Science: Water Research & Technology, 2(4), 749–760. https://doi.org/10.1039/C6EW00048G

Sillanpää, M., & Park, Y. (Eds.). (2023). Natural Organic Matter in Water. Butterworth-Heinemann. https://doi.org/10.1016/B978-0-12-824274-2.20001-X

Silva, G. M. e, Campos, D. F., Brasil, J. A. T., Tremblay, M., Mendiondo, E. M., & Ghiglieno, F. (2022). Advances in Technological Research for Online and In Situ Water Quality Monitoring—A Review. Sustainability, 14(9), Article 9. https://doi.org/10.3390/su14095059

Simpson, D. R. (2008). Biofilm processes in biologically active carbon water purification. Water Research, 42(12), 2839–2848. https://doi.org/10.1016/j.watres.2008.02.025

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

Statistisches Bundesamt. (2019). Öffentliche Wasserversorgung und öffentliche Abwasserentsorgung—Öffentliche Wasserversorgung (Reihe 2.1) [Fachserie 19]. Statistisches Bundesamt (Destatis).

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(3–4), 239–254. https://doi.org/10.1016/S0304-4203(03)00072-0

Stedmon, C. A., & Markager, S. (2005). Tracing the production and degradation of autochthonous fractions of dissolved organic matter by fluorescence analysis. Limnology and Oceanography, 50(5), 1415–1426. https://doi.org/10.4319/lo.2005.50.5.1415

Stedmon, C. A., Markager, S., Tranvik, L., Kronberg, L., Slätis, T., & Martinsen, W. (2007). Photochemical production of ammonium and transformation of dissolved organic matter in the Baltic Sea. Marine Chemistry, 104(3), 227–240. https://doi.org/10.1016/j.marchem.2006.11.005

Stedmon, C. A., & Bro, R. (2008). Characterizing dissolved organic matter fluorescence with parallel factor analysis: A tutorial: Fluorescence-PARAFAC analysis of DOM. Limnology and Oceanography: Methods, 6(11), 572–579. https://doi.org/10.4319/lom.2008.6.572

Stedmon, C. A., Seredyńska-Sobecka, B., Boe-Hansen, R., Le Tallec, N., Waul, C. K., & Arvin, E. (2011). A potential approach for monitoring drinking water quality from groundwater systems using organic matter fluorescence as an early warning for contamination events. Water Research, 45(18), 6030–6038. https://doi.org/10.1016/j.watres.2011.08.066

Stokes, G. G. (1852). On the change of refrangibility of light. Philosophical Transactions of the Royal Society of London, 142, 463–562. https://doi.org/10.1098/rstl.1852.0022

Stubbins, A., Lapierre, J.-F., Berggren, M., Prairie, Y. T., Dittmar, T., & del Giorgio, P. A. (2014). What's in an EEM? Molecular Signatures Associated with Dissolved Organic Fluorescence in Boreal Canada. Environmental Science & Technology, 48(18), 10598–10606. https://doi.org/10.1021/es502086e

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., … Treidel, H. (2013). Ground water and climate change. Nature Climate Change, 3(4), Article 4. https://doi.org/10.1038/nclimate1744

Teichmann, C. (2014). Multidimensionale Floureszenzspektroskopie zur Analyse des Interaktionsverhaltens von Huminstoffen mit Metallen [Bachelor's Thesis]. Hochschule für angewandte Wissenschaften Hamburg.

Thayanukul, P., Kurisu, F., Kasuga, I., & Furumai, H. (2013). Evaluation of microbial regrowth potential by assimilable organic carbon in various reclaimed water and distribution systems. Water Research, 47(1), 225–232. https://doi.org/10.1016/j.watres.2012.09.051

Thurman, E. M., Wershaw, R. L., Malcolm, R. L., & Pinckney, D. J. (1982). Molecular size of aquatic humic substances. Organic Geochemistry, 4(1), 27–35. https://doi.org/10.1016/0146-6380(82)90005-5

Thurman, E. M. (1985). Organic Geochemistry of Natural Waters. Springer Netherlands. https://doi.org/10.1007/978-94-009-5095-5

Tietjen, T., Vähätalo, A. V., & Wetzel, R. G. (2005). Effects of clay mineral turbidity on dissolved organic carbon and bacterial production. Aquatic Sciences, 67(1), 51–60. https://doi.org/10.1007/s00027-004-0753-2

Tok, S., Haan, K. de, Tseng, D., Usanmaz, C. F., Koydemir, H. C., & Ozcan, A. (2019). Early detection of E. coli and total coliform using an automated, colorimetric and fluorometric fiber optics-based device. Lab on a Chip, 19(17), 2925–2935. https://doi.org/10.1039/C9LC00652D

Tornevi, A. (2015). Precipitation, Raw Water Quality, Drinking Water Treatment and Gastrointestinal Illness [Doctoral Thesis]. Umeå University.

Torres, A., Bond, T. C., Lehmann, C. M. B., Subramanian, R., & Hadley, O. L. (2014). Measuring Organic Carbon and Black Carbon in Rainwater: Evaluation of Methods. Aerosol Science and Technology, 48(3), 239–250. https://doi.org/10.1080/02786826.2013.868596

TrinkwV. (2023). Verordnung über die Qualität von Wasser für den menschlichen Gebrauch. Trinkwasserverordnung vom 20. Juni 2023 (BGBl. 2023 I Nr. 159). Bundesgesetzblatt, 2023.

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley Pub. Co. http://www.gbv.de/dms/bowker/toc/9780201076165.pdf (2024, February 23)

Tuptuk, N., Hazell, P., Watson, J., & Hailes, S. (2021). A Systematic Review of the State of Cyber-Security in Water Systems. Water, 13(1), Article 1. https://doi.org/10.3390/w13010081

Van der Kooij, D. (1990). Assimilable Organic Carbon (AOC) in Drinking Water. In G. A. McFeters (Ed.), Drinking Water Microbiology: Progress and Recent Developments (pp. 57–87). Springer. https://doi.org/10.1007/978-1-4612-4464-6_3

Van der Kooij, D. (1992). Assimilable Organic Carbon as an Indicator of Bacterial Regrowth. Journal AWWA. https://doi.org/10.1002/j.1551-8833.1992.tb07305.x

Van der Kooij, D., & Van der Wielen, P. W. J. J. (2013). Microbial Growth in Drinking-Water Supplies: Problems, Causes, Control and Research Needs. IWA Publishing. https://doi.org/10.2166/9781780400419

Van der Pol, E., Coumans, F. A. W., Grootemaat, A. E., Gardiner, C., Sargent, I. L., Harrison, P., Sturk, A., van Leeuwen, T. G., & Nieuwland, R. (2014). Particle size distribution of exosomes and microvesicles determined by transmission electron microscopy, flow cytometry, nanoparticle tracking analysis, and resistive pulse sensing. Journal of Thrombosis and Haemostasis, 12(7), 1182–1192. https://doi.org/10.1111/jth.12602

Van Nevel, S., Kötzsch, S., Weilenmann, H.-U., Boon, N., & Hammes, F. (2013). Routine bacterial analysis with automated flow cytometry. Journal of Microbiological Methods, 94(2), 73–76. https://doi.org/10.1016/j.mimet.2013.05.007

Van Nevel, S. (2014). Growth and flow cytometric monitoring of bacteria in drinking water [Doctoral Thesis]. Ghent University.

Van Nevel, S., Buysschaert, B., De Roy, K., De Gusseme, B., Clement, L., & Boon, N. (2017a). Flow cytometry for immediate follow-up of drinking water networks after maintenance. Water Research, 111, 66–73. https://doi.org/10.1016/j.watres.2016.12.040

Van Nevel, S., Koetzsch, S., Proctor, C. R., Besmer, M. D., Prest, E. I., Vrouwenvelder, J. S., Knezev, A., Boon, N., & Hammes, F. (2017b). Flow cytometric bacterial cell counts challenge conventional heterotrophic plate counts for routine microbiological drinking water monitoring. Water Research, 113, 191–206. https://doi.org/10.1016/j.watres.2017.01.065

Vila-Costa, M., Gasol, J. M., Sharma, S., & Moran, M. A. (2012). Community analysis of high- and low-nucleic acid-containing bacteria in NW Mediterranean coastal waters using 16S rDNA

pyrosequencing. Environmental Microbiology, 14(6), 1390–1402. https://doi.org/10.1111/j.1462-2920.2012.02720.x

Volk, C., Wood, L., Johnson, B., Robinson, J., Wei Zhu, H., & Kaplan, L. (2002). Monitoring dissolved organic carbon in surface and drinking waters. Journal of Environmental Monitoring, 4(1), 43–47. https://doi.org/10.1039/B107768F

Vrouwenvelder, J. S., Beyer, F., Dahmani, K., Hasan, N., Galjaard, G., Kruithof, J. C., & Van Loosdrecht, M. C. M. (2010). Phosphate limitation to control biofouling. Water Research, 44(11), 3454–3466. https://doi.org/10.1016/j.watres.2010.03.026

Vucinic, L., O'Connell, D., Dubber, D., Coxon, C., & Gill, L. (2023). Multiple fluorescence approaches to identify rapid changes in microbial indicators at karst springs. Journal of Contaminant Hydrology, 104129. https://doi.org/10.1016/j.jconhyd.2022.104129

Wagner, M. (2014). Charakterisierung und Quantifizierung natürlicher organischer Wasserinhaltsstoffe mittels Fluoreszenzspektroskopie [Doctoral Thesis]. Technische Universität Dresden.

Walker, S. A., Amon, R. M. W., & Stedmon, C. A. (2013). Variations in high-latitude riverine fluorescent dissolved organic matter: A comparison of large Arctic rivers. Journal of Geophysical Research: Biogeosciences, 118(4), 1689–1702. https://doi.org/10.1002/2013JG002320

Wang, S., Li, Y., Xiao, K., & Huang, X. (2022). Fluorescence excitation-emission matrix as a novel indicator of assimilable organic carbon in wastewater: Implication from a coal chemical wastewater study. Science of The Total Environment, 804, 150144–150144. https://doi.org/10.1016/j.scitotenv.2021.150144

Wang, Y., Hammes, F., Boon, N., Chami, M., & Egli, T. (2009). Isolation and characterization of low nucleic acid (LNA)-content bacteria. The ISME Journal, 3(8), 889–902. https://doi.org/10.1038/ismej.2009.46

Wang, Y., Hammes, F., De Roy, K., Verstraete, W., & Boon, N. (2010). Past, present and future applications of flow cytometry in aquatic microbiology. Trends in Biotechnology, 28(8), 416–424. https://doi.org/10.1016/j.tibtech.2010.04.006

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

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

Wegner, C. H., & Hubbuch, J. (2022). Calibration-free PAT: Locating selective crystallization or precipitation sweet spot in screenings with multi-way PARAFAC models. Frontiers in Bioengineering and Biotechnology, 10. https://doi.org/10.3389/fbioe.2022.1051129

Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R., & Mopper, K. (2003). Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. Environmental Science & Technology, 37(20), 4702–4708. https://doi.org/10.1021/es030360x

Wells, M. J. M., Hooper, J., Mullins, G. A., & Bell, K. Y. (2022). Development of a fluorescence EEM-PARAFAC model for potable water reuse monitoring: Implications for inter-component protein–fulvic–humic interactions. Science of The Total Environment, 820, 153070. https://doi.org/10.1016/j.scitotenv.2022.153070

White, S., Quinn, J., Enzor, J., Staats, J., Mosier, S. M., Almarode, J., Denny, T. N., Weinhold, K. J., Ferrari, G., & Chan, C. (2021). FlowKit: A Python Toolkit for Integrated Manual and Automated Cytometry Analysis Workflows. Frontiers in Immunology, 12, 4652–4652. https://doi.org/10.3389/FIMMU.2021.768541/BIBTEX

WHO. (2014). Water safety in distribution systems. World Health Organization.

WHO. (2017). Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum. World Health Organization.

WHO. (2022). Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda. World Health Organization.

Worch, E. (1997). Wasser und Wasserinhaltsstoffe. Teubner.

Wullenweber, J. (2022). Implementierung und Validierung von real-time Sensorik an einer Pilotanlage zur Bestimmung der Trinkwasserqualität [Master's Thesis]. Technische Universität Hamburg.

Wünsch, U. J., Murphy, K. R., & Stedmon, C. A. (2015). Fluorescence quantum yields of natural organic matter and organic compounds: Implications for the fluorescence-based interpretation of organic matter composition. Frontiers in Marine Science, 2(NOV), 98–98. https://doi.org/10.3389/FMARS.2015.00098/BIBTEX

Wünsch, U. J., Murphy, K. R., & Stedmon, C. A. (2017). The One-Sample PARAFAC Approach Reveals Molecular Size Distributions of Fluorescent Components in Dissolved Organic Matter. Environmental Science & Technology, 51(20), 11900–11908. https://doi.org/10.1021/acs.est.7b03260

Wünsch, U. J., & Murphy, K. (2021). A simple method to isolate fluorescence spectra from small dissolved organic matter datasets. Water Research, 190, 116730. https://doi.org/10.1016/j.watres.2020.116730

Wünsch, U. J. (2024, January 11). Tutorials—The official webpage of the drEEM toolbox for MATLAB. https://dreem.openfluor.org/tutorials/ (2024, January 11)

Yamashita, Y., Panton, A., Mahaffey, C., & Jaffé, R. (2011). Assessing the spatial and temporal variability of dissolved organic matter in Liverpool Bay using excitation-emission matrix fluorescence and parallel factor analysis. Ocean Dynamics, 61(5), 569–579. https://doi.org/10.1007/S10236-010-0365-4/FIGURES/7

Yu, G.-H., He, P.-J., & Shao, L.-M. (2010). Novel insights into sludge dewaterability by fluorescence excitation–emission matrix combined with parallel factor analysis. Water Research, 44(3), 797–806. https://doi.org/10.1016/j.watres.2009.10.021

Yu, H., Song, Y., Tu, X., Du, E., Liu, R., & Peng, J. (2013). Assessing removal efficiency of dissolved organic matter in wastewater treatment using fluorescence excitation emission matrices with parallel factor analysis and second derivative synchronous fluorescence. Bioresource Technology, 144, 595–601. https://doi.org/10.1016/j.biortech.2013.07.025

Yuan, M. M., Guo, X., Wu, L., Zhang, Y., Xiao, N., Ning, D., Shi, Z., Zhou, X., Wu, L., Yang, Y., Tiedje, J. M., & Zhou, J. (2021). Climate warming enhances microbial network complexity and stability. Nature Climate Change, 11(4), Article 4. https://doi.org/10.1038/s41558-021-00989-9

Zhang, X., Jiang, X., Yang, Q., Wang, X., Zhang, Y., Zhao, J., Qu, K., & Zhao, C. (2018). Online Monitoring of Bacterial Growth with an Electrical Sensor. Analytical Chemistry, 90(10), 6006–6011. https://doi.org/10.1021/acs.analchem.8b01214

Zhao, H., Hou, D., Huang, P., & Zhang, G. (2014). Water Quality Event Detection in Drinking Water Network. Water, Air, & Soil Pollution, 225(11), 2183. https://doi.org/10.1007/s11270-014-2183-7

Zhou, Y., Martin, P., & Müller, M. (2019). Composition and cycling of dissolved organic matter from tropical peatlands of coastal Sarawak, Borneo, revealed by fluorescence spectroscopy and parallel factor analysis. Biogeosciences, 16(13), 2733–2749. https://doi.org/10.5194/bg-16-2733-2019

Zunabovic-Pichler, M., Mayr, E., Schönher, C., Fiedler, C. J., Kerschbaumer, D., Proksch, P., & Perfler, R. (2018). Durchflusszytometrie in der Wasserversorgung (p. 157). BMLRT.

Appendix I. Supplemental Theoretical Background, Materials and Methods

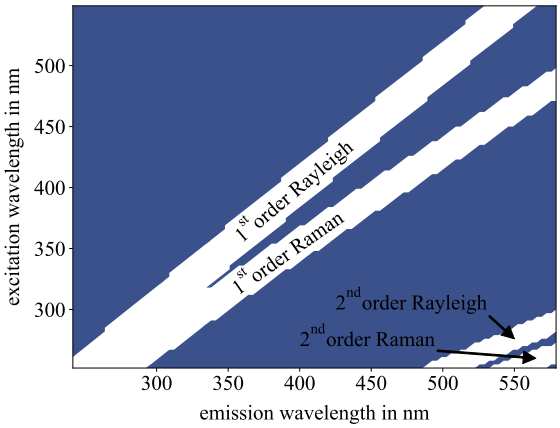

Figure I-1: An example of Rayleigh and Raman scattering bands positions in an EEM.

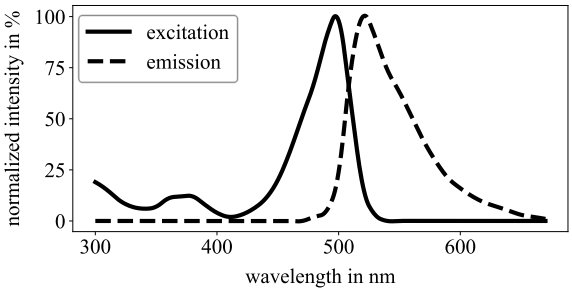

Figure I-2: Normalized excitation and emission spectra of SYBR Green I fluorescence dye of Sigma-Aldrich.

Table I-I: List of chemicals and chemical solutions used as standards for PARAFAC fluorescence spectroscopy calibration.

substance	manufacturer	solvent	concentration in mg L^{-1}	corresponding TOC in mg L^{-1}
lactalbumin	Sigma-Aldrich	0.1 M KCl	0.1, 0.25, 0.5, 0.75, 1, 2, 2.5, 5, 10	n/a
lysozyme	Sigma-Aldrich	0.1 M KCl	0.1, 0.25, 0.5, 0.75, 1, 2, 2.5, 5, 10	n/a
bovin serum albumin (BSA)	Carl Roth	0.1 M KCl	0.1, 0.25, 0.5, 0.75, 1, 2, 2.5, 5, 10	n/a, n/a, n/a, n/a, 0.5, n/a, 0.6, 1.3, 3.6
tryptophan ($C_{11}H_{12}N_2O_2$)	Merck	ultrapure water	0.00025, 0.0005, 0.001, 0.002, 0.003, 0.004, 0.005, 0.0075, 0.01, 0.02, 0.05, 0.1	bdl
tyrosine ($C_9H_{11}NO_3$)	Carl Roth	ultrapure water	0.00025, 0.0005, 0.001, 0.002, 0.003, 0.004, 0.005, 0.0075, 0.01, 0.02, 0.05, 0.1	bdl
SRNOM	IHSS	ultrapure water	1, 2.5, 5, 7.5, 10	0.5, 1.1, 2.2, 3.4, 4,7
SRFA	IHSS	ultrapure water	1, 1.5, 2, 2.5, 3, 3.5, 4, 5	0.6, 0.9, 1.2, 1.4, 1.7, 2.0, 2.2, 2.7
SRHA	IHSS	ultrapure water	1, 1.5, 2, 2.5, 3, 3.5, 4, 5	0.5, 0.8, 1.0, 1.3, 1.5, 1.7, 1.9, 2.4
HSNOM	Sampling Lake Hohlohsee	ultrapure water	5 v%, 12 v%, 24 v%	0.9, 2.4, 4.7

Table I-II: Overview of the names and origins of the *base data* water samples which form the basic framework for the overall six-component PARAFAC model.

sample	origin	sample	origin	sample	origin
IB02	dw_{wds} (1 v%) + dw_{wds} (99 v%)	Misch02zu98_1	wwe (1 v%) + dw_{wds}	T10R10	dw_{wds} (50 v%) + rw (50 v%)
IB02_1	dw_{wds} (1 v%) + dw_{wds} (99 v%)	Misch02zu98_2	wwe (1 v%) + dw_{wds}	T10R10_1	dw_{wds} (50 v%) + rw (50 v%)
IB02_2	dw_{wds} (1 v%) + dw_{wds} (99 v%)	Misch02zu98_3	wwe (1 v%) + dw_{wds}	T10R10_2	dw_{wds} (50 v%) + rw (50 v%)
IB02_3	dw_{wds} (1 v%) + dw_{wds} (99 v%)	Misch2zu18	dw_{wds} (10 v%) + dw_{wds} (90 v%)	T10R10_3	dw_{wds} (50 v%) + rw (50 v%)
IB_04	dw_{wds} (2 v%) + dw_{wds} (98 v%)	Misch5zu15	dw_{wds} (25 v%) + dw_{wds} (75 v%)	T15R5	dw_{wds} (75 v%) + rw (25 v%)
M1isch1zu19	dw_{wds} (5 v%) + dw_{wds} (95 v%)	Misch10zu10	dw_{wds} (50 v%) + dw_{wds} (50 v%)	T18R2	dw_{wds} (90 v%) + rw (10 v%)
M1isch02zu198	dw_{wds} (5 v%) + dw_{wds} (95 v%)	Misch15zu5	dw_{wds} (75 v%) + dw_{wds} (25 v%)	T19R1	dw_{wds} (95 v%) + rw (5 v%)
M1isch02zu198_1	dw_{wds} (1 v%) + dw_{wds} (99 v%)	Misch25zu175	dw_{wds} (12.5 v%) + dw_{wds} (78.5 v%)	T19R1_1	dw_{wds} (95 v%) + rw (5 v%)
M1isch02zu198_2	dw_{wds} (1 v%) + dw_{wds} (99 v%)	sampleIJV09_001	dw_{wds}	T19R1_2	dw_{wds} (95 v%) + rw (5 v%)
M1isch02zu198_3	dw_{wds} (1 v%) + dw_{wds} (99 v%)	sampleIJV09_003	dw_{wds}	T19R1_3	dw_{wds} (95 v%) + rw (5 v%)
M1isch04zu196	dw_{wds} (2 v%) + dw_{wds} (98 v%)	sampleIJV10_001	dw_{wds}	T196R04	dw_{wds} (98 v%) + rw (2 v%)
Misch02zu98	wwe (1 v%) + dw_{wds} (99 v%)	sampleIJV10_003	dw_{wds}	T198R02	dw_{wds} (99 v%) + rw (1 v%)

Table I-III: List of chemicals and chemical solutions used in all water analysis experiments.

substance	grade	manufacturer	application
calibration Beads 3.0 µm, ready-to-use	n/a	Sysmex	calibration procedure flow cytometer
cleaning Solution	n/a	Sysmex	cleaning procedure flow cytometer
decontamination solution	n/a	Sysmex	cleaning procedure flow cytometer
dimethyl sulfoxide (DMSO) (C_2H_6OS)	anhydrous, ≥ 99.9 %	Sigma-Aldrich	organic solvent for fluorescent dye
Dipotassium phosphate (K_2HPO_4)	≥ 99.5 %	Carl Roth	LC-OCD
hydrochloric acid (HCl)	37 %	Carl Roth	pH adjustment
peptone ex casein	n/a	Carl Roth	LB medium
sodium chloride (NaCl)	≥ 99 %	Carl Roth	LB medium
sodium hypochlorite solution	n/a	Sysmex	cleaning procedure flow cytometer
sodium hypochlorite solution (NaClO)	12 % Cl	Carl Roth	disinfection solution flow cytometer and fluorescence spectrometer
sodium phosphate ($NaH_2PO4 \cdot 2H_2O$)	≥ 98 %	Merck	LC-OCD
sodium thiosulphate pentahydrate ($Na_2S_2O_3 \cdot 5H_2O$)	≥ 99.5 %	Carl Roth	quenching solution flow cytometer
SYBR® Green I nucleic acid gel stain	10^4 in DMSO	Merck	fluorescent dye
tris(hydroxymethyl)aminomethane (Tris) ($C_4H_{11}NO_3$)	99.9 %	Sigma-Aldrich	organic buffer for flow cytometric working solution
yeast extract	n/a	Carl Roth	LB medium

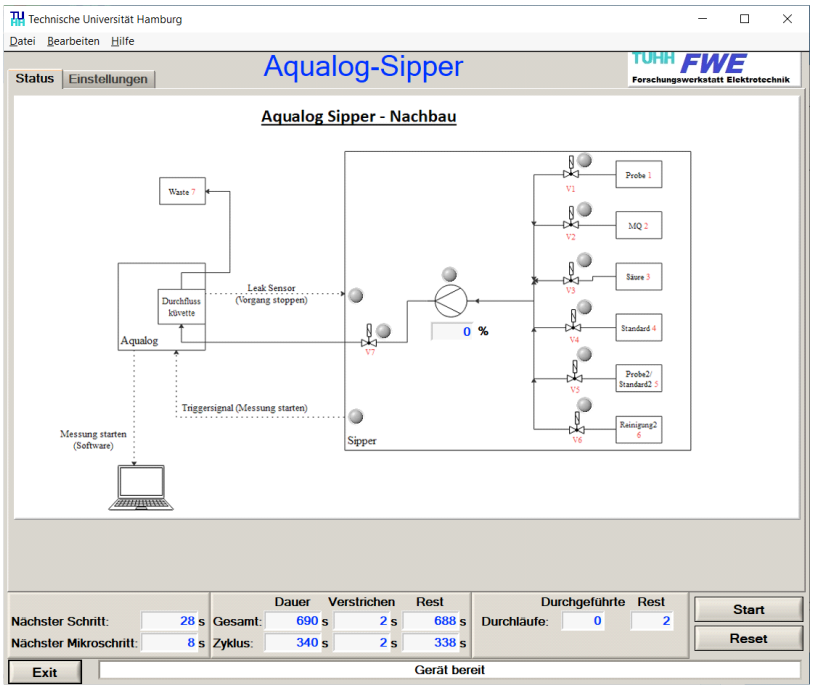

Figure I-3: Flow scheme of the Aqualog-Sipper.

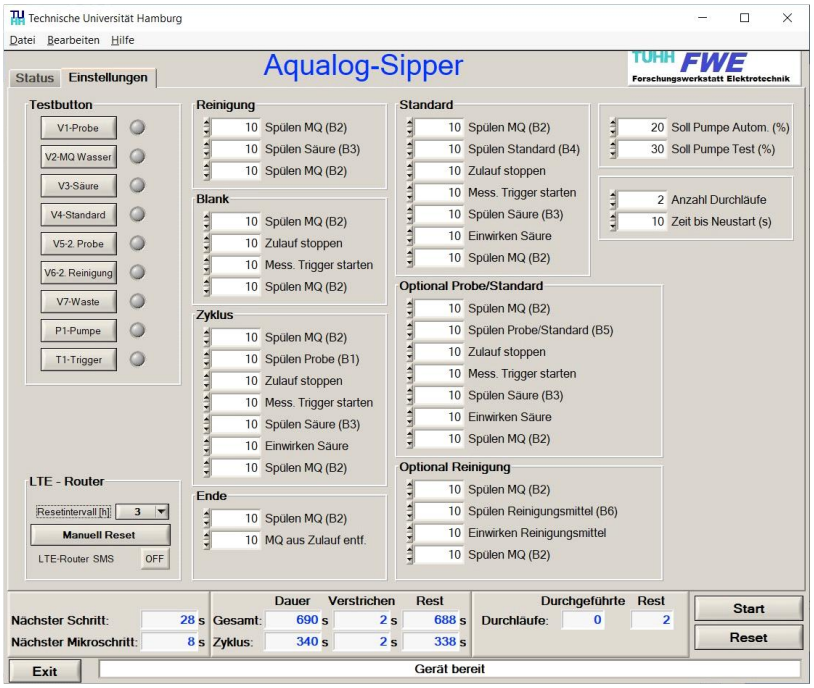

Figure I-4: Customizable settings of the Aqualog-Sipper.

```
%%          Install drEEM toolbox
    cd ("C:\Users\Jonas\MATLAB\drEEM-0.6.5")
    dreeminstall

    %%          Define size of the raw data set
    flusize = 'A2..EE126';        % Matrice with respective EEM data
    abssize = 'A4..K137';         % Matrice with respective absorbance data

    %%          Import raw data
    %   Import every sample (fluorescence)
    cd ("C:\Users\Jonas\MATLAB\240130_diss\2_Theory\actual_measurement")
    [X,Emmat,Exmat,fl,outdata]=readineems(3,'*Waterfall Plot Sample.dat',flusize,[0
1],0,0);
    DS = assembledataset(X,Exmat(1,:),Emmat(:,1),'RU','filelist',fl,[]);
    clearvars -except DS datalocation savelocation flusize abssize k

    %   Import corresponding blank (fluorescence)
    [X_b,Emmat_b,Exmat_b,fl_b,outdata_b]=readineems(3,'*Waterfall Plot
Blank.dat',flusize,[0 1],0,0);
    DSb = assembledataset(X_b,Exmat_b(1,:),Emmat_b(:,1),'RU','filelist',fl_b,[]);
    clearvars -except DS DSb datalocation savelocation flusize abssize k

    %   Import absorbance spectra
    [S_abs,W_abs,wave_abs,filelist_abs]=readinscans('Abs','dat_1_10',abssize,0,0,'* - Abs
Spectra Graphs');

    %   Define absorbance spectra
    A=[wave_abs;S_abs];
    Abs.A=A;
    Abs.filelist=cellstr(filelist_abs);
    Abs.wave=wave_abs;
    clearvars -except DS DSb Abs flusize abssize k

    %%          Check data alignment
    try
        disp([cellstr(num2str((1:DS.nSample)')) DS.filelist DSb.filelist])
    catch
        error('Missmatch: DS.nSample, DS.filelist, DSb.filelist')
    end
    try
        disp([DS.filelist DSb.filelist Abs.filelist])
    catch
        error('Missmatch: DS.filelist, DSb.filelist, Abs.filelist')
    end
    try
        disp([cellstr(num2str((1:DS.nSample)')) DS.filelist DSb.filelist Abs.filelist])
    catch
        error('Missmatch: DS.nSample, DS.filelist, DSb.filelist, Abs.filelist')
    end

    %%          Adaption of the EEM data according to absorbance data
    disp('Abs wavelengths: min and max')
    AbsRange=[min(Abs.wave) max(Abs.wave)];
    disp(AbsRange)
    disp('EEM wavelengths: min and max')
    EEMRange=[min([DS.Em' DS.Ex']) max([DS.Em' DS.Ex'])];
    disp(EEMRange)

    %   Removal of wavelenths <240nm and >600nm
    Xin=subdataset(DS,[],...
        logical(double(DS.Em<AbsRange(1))+ double(DS.Em>AbsRange(2))),...
        logical(double(DS.Ex<AbsRange(1))+ double(DS.Ex>AbsRange(2))));
    B=subdataset(DSb,[],...
        logical(double(DS.Em<AbsRange(1))+ double(DS.Em>AbsRange(2))),...
        logical(double(DS.Ex<AbsRange(1))+ double(DS.Ex>AbsRange(2))));

    %%          EEM correction - IFE and normalization of the EEMs with Raman peak
intensity
    RamanWavel=351;                     %350 nm in literature, here: 351 nm
    Sr=squeeze(B.X(:,:,DSb.Ex==351));
    W=[B.Em';Sr];                       %2D matrix of the 351nm Raman scan
    B.W=W;

    [XcRU, Arp, IFCmat, BcRU, XcQS, QS_RU]=fdomcorrect(Xin.X,Xin.Ex,Xin.Em,...
        [Xin.Em ones(size(Xin.Em))],[Xin.Ex ones(size(Xin.Ex))],...
        B.W,[351 381 426],Abs.A,B.X,B.W,[],[]);

```

```
%%          Consolidation of the data set
Xin.X=XcRU;                                  %X in Raman units
Xin.RamanArea=Arp;                           %Raman Peak Area
Xin.IFE=IFCmat;                              %IFE correction factor
Xin.Abs_wave=Abs.A(1,:);                     %absorbance wavelengths
Xin.Abs=Abs.A(2:end,:);                      %absorbance data
Xin.Ram_wave=B.Em';                          %Water Raman wavelengths
Xin.Emcor='internally corrected by AquaLog';    %correction by Aqualog, not drEEM
Xin.Excor='internally corrected by AquaLog';    %correction by Aqualog, not drEEM
Xin.RamOpt=[350 381 426];                    %Raman interaction range of emission spectra
Xin.RamSource='Water Raman scans extracted from blanks';
clearvars -except Xin k
checkdataset(Xin)

%%          Preprocessing of the data set
%   Scattertreatment - Removal of noisy and unreliable wavelengths (Em>580 and Ex <250 nm)
Xstart = subdataset(Xin,[],Xin.Em>580,Xin.Ex<250);
Xstart = subdataset(Xstart,[],Xstart.Em<240,Xstart.Ex>550);
clearvars -except Xstart k

%%          Rayleigh and Raman Scattering
% Xs = smootheem(Xstart, [Ray1 area], [Ram1 area], [Ray2 area], [Ram2 area], [0 - removal of the area, 1 - interpolation of the area], Frequency of energy loss due to non-elastic scattering, seconds EEM shows up)
Xs = smootheem(Xstart,[12 30],[7 2],[15 18],[],[0 1 0 0],[],3382,0);

%   Display Rayleigh and Raman Scattering
%Xs = smootheem(Xstart,[15 18],[15 18],[15 18],[15 18],[0 0 0 0],[],3382,0);

%   Check individually all EEMs
%eemreview(Xs);
%%          Error correction and data exploration
%   Automatic function removes errors for each EEM
dataout = rmspikes(Xs,'interpolate',true);
Xs = dataout;

%       Evaluation of the new data set
%   Check individually all EEMs
%eemreview(Xs)
%   Outlier test
Test1=outliertest(Xs,[],3:7);
%   Check if components correlate
%compcorrplot(Test1, 6);                                          %here: 6 components

% Normalization of the data set
Xpre = normeem(Xs);

%       Evaluation of the new normalized data set
%   Check individually all EEMs
%eemreview(Xpre);
%   Outlier test
%Test2=outliertest(Xpre,[],3:7);
%   Check if components correlate
%compcorrplot(Test2, 6);                    %here: 6 components
%   Check residues
%eemview({Test1, Test1},{'X','Model6','error_residuals'});      %here: 6 components
%compare2models(Test1,3,6,0.05)                                  %here: compare 3 and 6 components
%   Check spectral loadings
%spectralloadings(Test1,3:7);
%   Check sum of squared error
%specsse(Test1,3:7);

%%          Model Validation

%   Appropriate model calculation
%   randinitanal(raw data set, component number, starts, random initialization, constraints, nonnegativity, convgrit, number for convergence criterion)
modeln = randinitanal(Xpre,3:7,'starts',10,'constraints','nonnegativity','convgcrit',1e-6);

%       Evaluation of the new model
%   Check if components correlate
%compcorrplot(modeln, 6);                                          %here: 6 components
%   Check residues
%eemview({modeln, modeln},{'X','Model4','error residuals'});     %here: 6 components
```

```
145    %compare2models(modeln,3,6,0.05)                                    %here: compare 3 and 6
components
146    %   Check spectral loadings
147    %spectralloadings(modeln,3:7);
148    %   Check sum of squared error
149    %specsse(modeln,3:7);
150    %   Check core consistency
151    %coreandvar(modeln);
152
153    %       Further helpful functions for validate the model
154    %   Scores, Errors, leverage
155    %loadingsandleverages(modeln,6);
156    %comparespectra(modeln,4:6);
157    %   Fingerprint for x components
158    %fingerprint(modeln,6);                                          %here: 6 components
159
160    %%          Model validation - split half analysis
161
162    %   Generate an overall model
163    %final setting
164    %LSmodel_6 =
randinitanal(Xpre,6,'starts',10,'constraints','nonnegativity','convgcrit',1e-6);
%here: 6 components model
165    LSmodel_6 =
randinitanal(Xpre,6,'starts',30,'constraints','nonnegativity','convgcrit',1e-4);
166    spectralloadings(LSmodel_6,6)
167    %   Exporting without Split-half analysis
168    %modelexport(LSmodel_6,6,'model_data_6_30_e-4.xlsx');
169
170    %% Model validation - split-half analysis
171    DSsplit = splitds(modeln,[],4,'random',{[1 2],[3 4]});
172    splitmodels_6 = splitanalysis(DSsplit,6,10,'nonnegativity');
173    val_resultsn_6 = splitvalidation(splitmodels_6,6,[],[],LSmodel_6);
174    %save('val_resultsn_6')
175
176    %%          Result interpretation
177    %   Renormalization of the overall model
178    val_results = normeem(LSmodel_6,'reverse',6);
179    save('val_results');
180    %   Calculate and export fluorescence indices
181    %indices = pickpeaks(val_results);
182    %writetable(indices, 'indices_new.xlsx');
183    %   Export validated model (Excel)
184    Data = modelexport(val_results,6,'model_data_6.xlsx');
185    %   Export validated model (OpenFluor)
186    openfluor(val_results,6,'model_data_6_open_fluor.txt');
```

Figure I-5: MATLAB code of the validated overall six-components PARAFAC model.

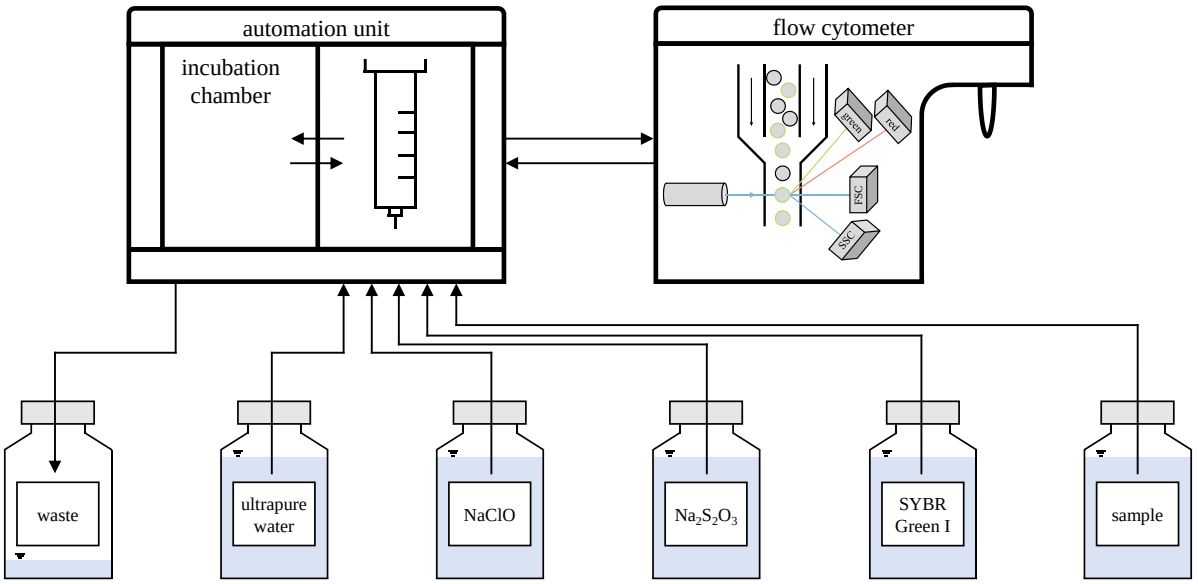

Figure I-6: Flow scheme of the flow cytometry automation unit OC-300.

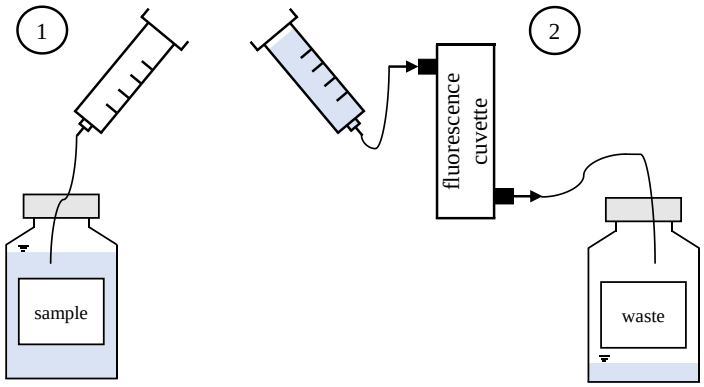

Figure I-7: How to transfer oxygen-free groundwater samples to fluorescence spectroscopy analysis. (1) Take the sample from a freshly opened glass bottle. (2) Transfer the sample to the flow-through fluorescence cuvette and start the measurement.

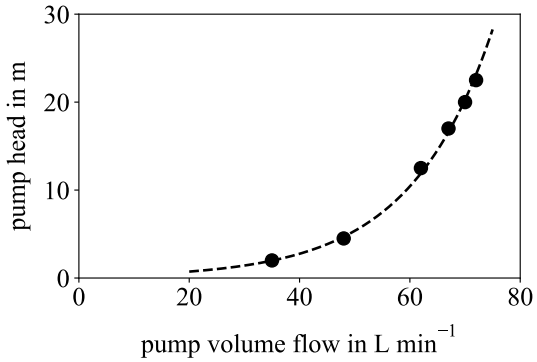

Figure I-8: Experimentally determined pump characteristic curve of the *closed* pilot plant *system*.

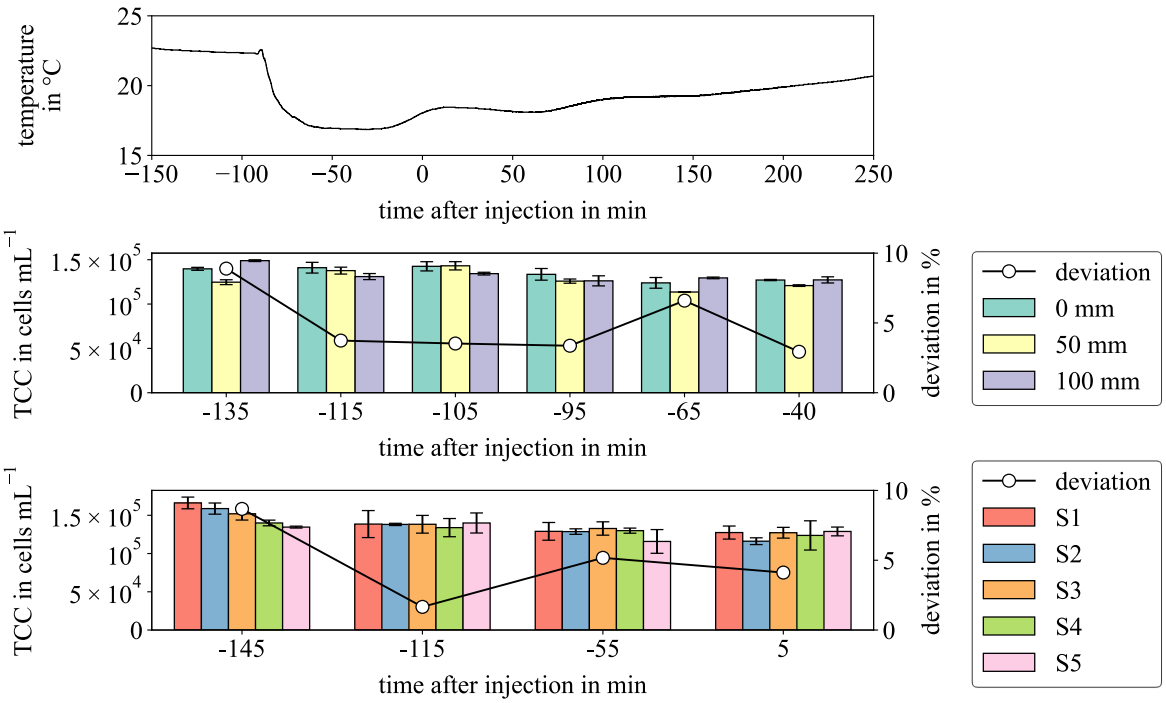

Figure I-9: Pilot plant preliminary test results. Temperature profile, sampling depth, and sampling point location.

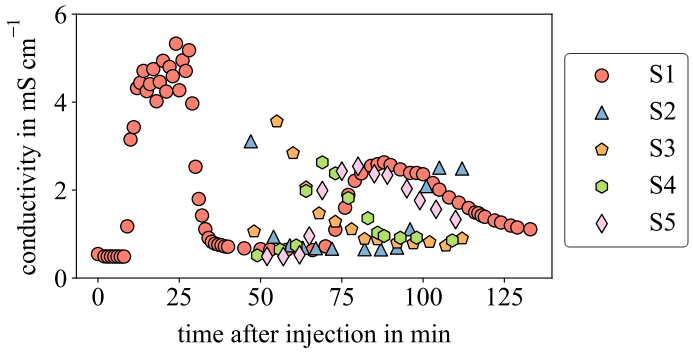

Figure I-10: Pilot plant preliminary test result. Injection of a NaCl solution (12 mS cm^{-1}) into the pilot plant system.

Appendix II. Supplemental Results

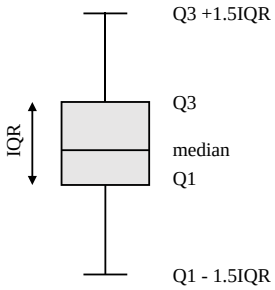

Figure II-1: The box covers 50 % of the values. The horizontal lines of the boxplot represent the first quartile (Q1), median and third quartile (Q3) with whiskers extending to 1.5 times the interquartile range (IQR) from Q1 and Q3. Whiskers end at the value from the data that is still within the limit. (Tukey, 1977).

Table II-1: Linear correlation (Pearson method) between all measured parameters of n = 23 samples from various locations of the water distribution system (dw_{wds}) and the waterworks exit (dw_{wwe}). *r* in bold font, and *p* in normal font, * $p < 0.005$, ** $p < 0.0001$. This table contains supporting material for Figure 4-6.

	TCC	HNA	LNA	% HNA	TOC	C1	C2	C3	C4	C5	C6	Conductivity	pH	AOC
TCC	**1** 0													
HNA	**0.97** **	**1** 0												
LNA	**0.98** **	**0.92** **	**1** 0											
%HNA	**0.24** 0.25	**0.45** 0.03	**0.09** 0.67	**1** 0										
TOC	**0.93** **	**0.94** **	**0.88** **	**0.45** 0.06	**1** 0									
C1	**0.91** **	**0.88** **	**0.89** **	**0.31** 0.15	**0.96** **	**1** 0								
C2	**0.59** *	**0.52** 0.01	**0.62** *	**0.12** 0.60	**0.71** *	**0.82** **	**1** 0							
C3	**0.85** **	**0.84** **	**0.83** **	**0.33** 0.13	**0.96** **	**0.97** **	**0.86** **	**1** 0						
C4	**0.85** **	**0.82** **	**0.78** **	**0.29** 0.17	**0.96** **	**0.96** **	**0.86** **	**0.99** **	**1** 0					
C5	**0.78** **	**0.72** **	**0.78** **	**0.12** 0.60	**0.84** **	**0.91** **	**0.81** **	**0.88** **	**0.92** **	**1** 0				
C6	**0.39** 0.07	**0.35** 0.10	**0.37** 0.08	**0.15** 0.50	**0.60** *	**0.61** *	**0.82** **	**0.72** *	**0.72** **	**0.61** *	**1** 0			
Conductivity	**0.21** 0.33	**0.18** 0.40	**0.19** 0.38	**-0.02** 0.95	**0.41** 0.05	**0.39** 0.07	**0.49** 0.02	**0.45** **0.03**	**0.48** 0.02	**0.48** 0.02	**0.82** **	**1** 0		
pH	**-0.29** 0.17	**-0.24** 0.27	**-0.33** 0.12	**-0.15** 0.51	**-0.29** 0.17	**-0.46** 0.03	**-0.52** 0.01	**-0.40** 0.06	**-0.40** 0.06	**-0.46** 0.03	**-0.23** 0.29	**0.05** 0.81	**1** 0	
AOC	**-0.02** 0.94	**0.06** 0.79	**-0.06** 0.77	**0.41** 0.05	**0.01** 0.96	**0.02** 0.93	**-0.08** 0.71	**-0.02** 0.92	**-0.04** 0.85	**-0.08** 0.72	**-0.12** 0.60	**-0.23** 0.30	**-0.34** 0.12	**1** 0

Table II-2: Overview of the seven-component PARAFAC model including the spectral peak wavelengths, substance or substance-like designations, main occurrence in natural and engineered waters and corresponding references for all components. The comparison with the literature was performed via the *OpenFluor* online library (Murphy et al., 2014b).

component	λ_{ex} peak in nm	λ_{em} peak in nm	substance	occurrence	references
H1	252/303	403	humic-like, microbial and terrestrial	wastewater river water seawater	[1] [2] [3]
H2	252/393	463	humic-like, terrestrial	drinking water wastewater	[4] [1]
H3	261/357	453	humic-like, terrestrial	lake water river water drinking water	[5] [5] [4]
H4	252	526	humic-like, terrestrial	mangrove soil	[6]
P1	280	337	tryptophan-like	seawater	[7]
P2	280	350	indole, tryptophan	artificial algae soil	[8] [9] [9]
P3	275	302	tyrosine-like	seawater	[10]

[1] Murphy et al. (2011), [2] Lin and Guo (2020), [3] Stedmon et al. (2007), [4] Shutova et al. (2014), [5] Wünsch et al. (2017), [6] Kida et al. (2021), [7] Yamashita et al. (2011), [8] Wünsch et al. (2015), [9] Derrien et al. (2019), [10] Paerl et al. (2020)

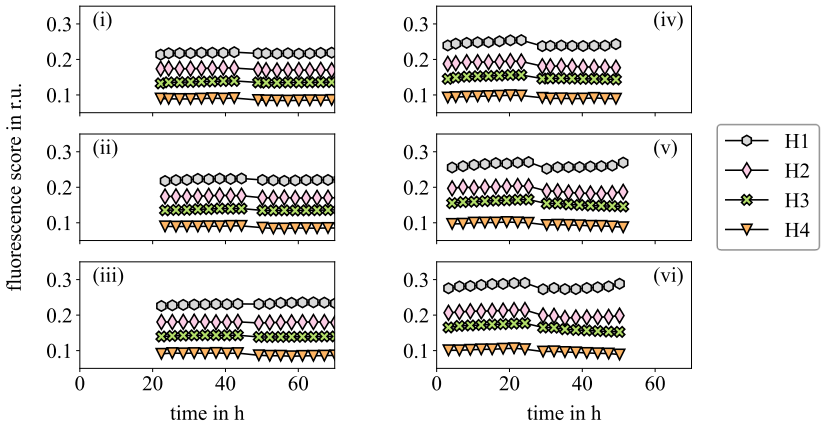

Figure II-2: Time course of the fluorescence score of the three humic-like components H1, H2, H3, H4 for six different (i) – (vi) online monitored experiments. The components obtain the following excitation and emission maxima: H1 λ_{ex} = 252/303 nm, λ_{em} = 403 nm, H2 λ_{ex} = 252/393 nm, λ_{em} = 463 nm, H3 λ_{ex} = 261/357 nm, λ_{em} = 453 nm, H4 λ_{ex} = 252 nm, λ_{em} = 526 nm. The conditions of the respective experiments are shown in Table 3-1 and Table 5-1.

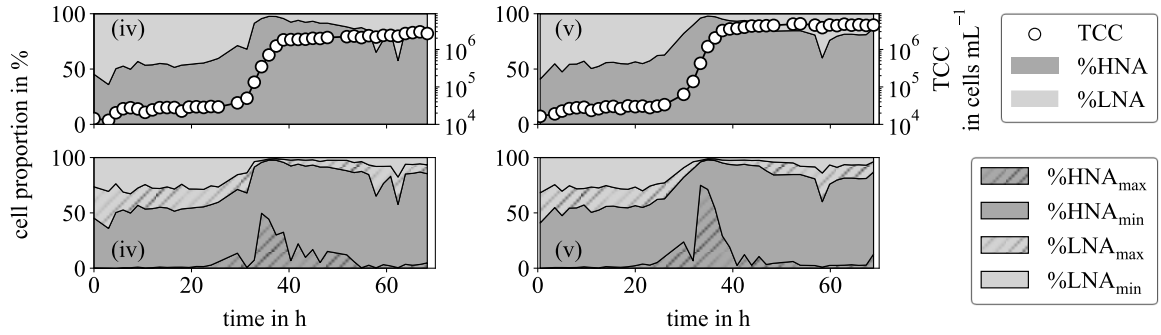

Figure II-3: Additional subdivision of the results from Figure 5-2 into HNA_{max}, HNA_{min}, LNA_{max}, and LNA_{min} of two different online monitored experiments (iv) and (v). The conditions of the respective experiments are shown in Table 3-1 and Table 5-1. Adapted and revised from Schuster et al. (2022a).

Table II-3: Relative calculation time (in %) for overall six-component PARAFAC modelling regarding the most time-consuming parameters number of random initializations and convergence criteria.

number of random initializations		5	10	20	30	50	100
		relative calculation time in %					
convergence criterion	1e-4	1	2	3	5	9	19
	1e-5	2	3	9	10	18	29
	1e-6	3	9	12	14	28	45
	1e-7	5	12	17	30	46	77
	1e-8	8	13	21	40	54	100

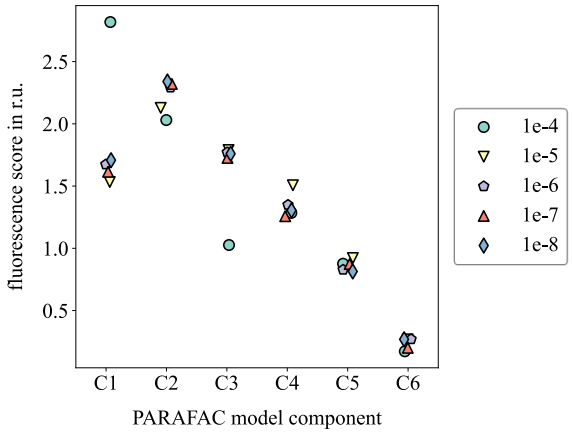

Figure II-4: Effects of the convergence criterion between 1e-4 and 1e-8 on the fluorescence score of each PARAFAC component for a stable number of random initializations of 10. This is the result of one sample from the *base data*.

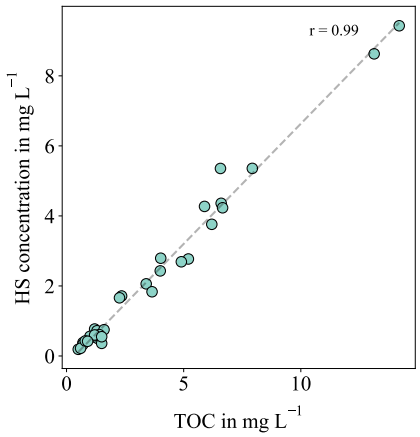

Figure II-5: Correlation between HS content analyzed by LC-OCD and TOC of groundwater samples (n = 34).

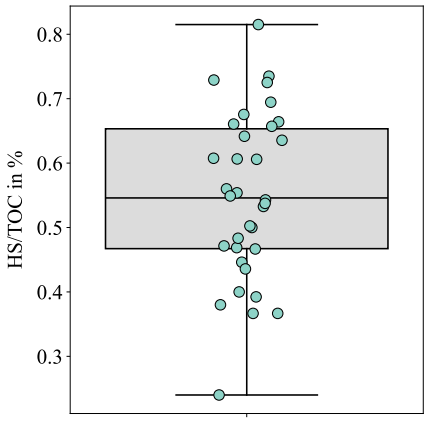

Figure II-6: Proportion of HS in TOC of groundwater samples (n = 34).

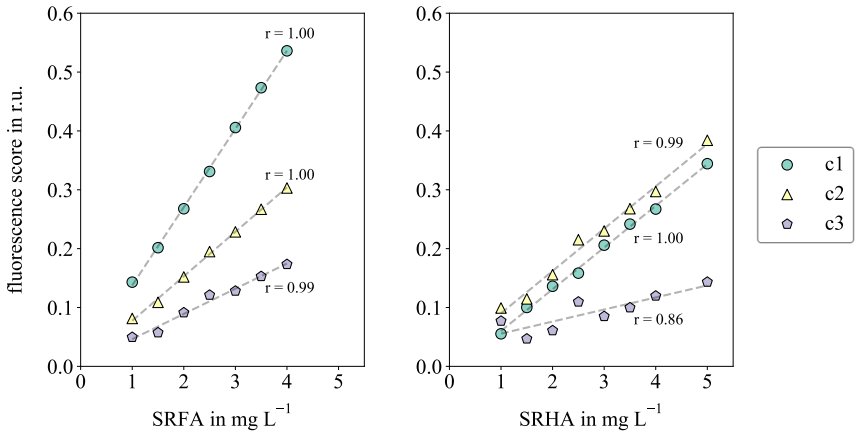

Figure II-7: Fluorescence spectroscopy intensities of SRHA and SRFA standards as calculated using the groundwater three-components PARAFAC model.

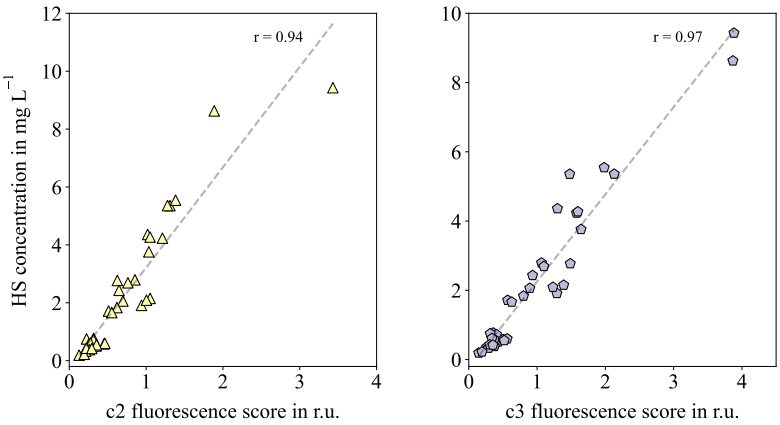

Figure II-8: Correlation between HS content analyzed by LC-OCD and fluorescence scores of c2, and c3 of the groundwater three-component PARAFAC model (n = 38).