Drinking water aromaticity and treatability is predicted by dissolved organic matter fluorescence

Water Research 220 (2022) 118592

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Drinking water aromaticity and treatability is predicted by dissolved organic matter fluorescence

Marc Philibert [a], Simin Luo [a], Lavel Moussanas [a], Qingqing Yuan [a], Emmanuelle Filloux [a], Flavia Zraick [a], Kathleen R. Murphy [b,*]

[a] SUEZ - CIRSEE, 38, rue du Président-Wilson, 78230, Le Pecq, France
[b] Chalmers University of Technology, Department of Architecture and Civil Engineering, SE-412 96, Gothenburg, Sweden

ARTICLE INFO

Keywords:
Dissolved organic matter (DOM)
Parallel factor analysis (PARAFAC)
Drinking water
Treatability
CDOM, SUVA

ABSTRACT

Samples from fifty-five surface water resources and twenty-five drinking water treatment plants in Europe, Africa, Asia, and USA were used to analyse the fluorescence composition of global surface waters and predict aromaticity and treatability from fluorescence excitation emission matrices. Nine underlying fluorescence components were identified in the dataset using parallel factor analysis (PARAFAC) and differences in aromaticity and treatability could be predicted from ratios between components H_{ii} ($\lambda_{ex}/\lambda_{em}$= 395/521), H_{iii} ($\lambda_{ex}/\lambda_{em}$= 330/404), P_i, ($\lambda_{ex}/\lambda_{em}$=290/365) and P_{ii} ($\lambda_{ex}/\lambda_{em}$= 275/302). Component H_{ii} tracked humic acids of primarily plant origin, H_{iii} tracked weathered/oxidised humics and the "building block" fraction measured by LC-OCD, while P_i and P_{ii} tracked amino acids in the "low molecular weight neutrals" LC-OCD fraction. Ratios between PARAFAC components predicted DOC removal at lab scale for French rivers in standardized tests involving coagulation, powdered activated carbon (PAC), chlorination, ion exchange (IEX), and ozonation, alone and in combination. The ratio H_{ii}/H_{iii}, for convenience named "PARIX" standing for "PARAFAC index", predicted SUVA according to a simple relationship: SUVA = 4.0 x PARIX (RMSEp=0.55) $Lmg^{-1}m^{-1}$. These results expand the utility of fluorescence spectroscopy in water treatment applications, by demonstrating the existence of previously unknown relationships between fluorescence composition, aromaticity and treatability that appear to hold across diverse surface waters at various stages of drinking water treatment.

1. Introduction

Dissolved natural organic matter (DOM) occurs in all surface waters where it originates from the decomposition and leaching of vegetation and in-situ primary productivity and undergoes continuous biotic and abiotic processing (Aiken, 2014; Moran et al., 2016; Moran et al., 1991). Natural waters consequently contain thousands of different DOM molecules with varying physicochemical characteristics. Removing DOM is a major objective of drinking water treatment, since too much DOM protects pathogens, causes aesthetic issues at the tap such as unpleasant colour, taste and odour, and leads to harmful by-products being formed during disinfection (Leenheer and Croué, 2003; Leenheer et al., 2001). Additionally, DOM undermines treatment performance and increases costs by increasing coagulant and oxidant demand and fouling membranes (Kennedy et al., 2005). Since various organic matter fractions respond differently to physical and chemical treatment (Croué et al., 1993; Edzwald, 1993; Moona et al., 2021), DOM character should directly inform the selection and operation of best available treatment options.

An important step toward optimising drinking water treatment is to predict the performance of specific treatment barriers (i.e. DOC removal efficiency) from the chemical composition of the incoming raw water. Waters with similar chemical compositions are expected to respond similarly to a given treatment process. For example, specific UV_{254}

Abbreviations: DOM, Dissolved organic matter; CDOM, Chromophoric dissolved organic matter; SUVA, Specific UV absorbance (UV254/DOC); DOC, Dissolved organic carbon; FIX, Fluorescence Index; BIX, Biological Index; HIX, Humification Index; PARAFAC, Parallel factor analysis; UV, Ultraviolet; UV254, Ultraviolet light excited at 254 nm; LC-OCD, Liquid chromotography with organic carbon detection; PARIX, PARAFAC index; IEX, Ion Exchange Resin; PAC, Powdered activated carbon; LMW, Low-molecular weight; HS, Humic substances; NMR, Nuclear magnetic resonance; EEM, Excitation emission matrix; FDOM, Fluorescent dissolved organic matter; RMSE, Root mean square error; RMSEp, Root mean square errorof prediction.

* Corresponding author.
E-mail address: murphyk@chalmers.se (K.R. Murphy).

https://doi.org/10.1016/j.watres.2022.118592
Received 4 March 2022; Received in revised form 19 April 2022; Accepted 10 May 2022
Available online 12 May 2022

absorption relative to DOC concentration (SUVA) in water samples is directly proportional to percent aromaticity measured by C^{13} NMR (Weishaar et al., 2003). SUVA strongly predicts coagulation performance across diverse water resources (Archer and Singer, 2006; Edwards, 1997) and also predicts other parameters of interest including disinfection byproduct formation potential (Edzwald et al., 1985; Hua et al., 2015) and potential for membrane fouling (Amy, 2008).

A disadvantage of SUVA is that to measure it according to standard US-EPA method 415.3 requires two different instruments, i.e. a UV spectrophotometer and a TOC analyzer (Potter and Wimsatt, 2005). TOC analyzers are rare at treatment plants due to expense and operational challenges that make them susceptible to a range of analytical interferences (Mopper and Qian, 2006). Additionally, commercial TOC analyses for drinking water are expensive relative to other monitoring parameters and often have relatively high detection limits (>1 mg/L). UV spectrophotometers also vary greatly in sensitivity (Mopper and Qian, 2006), and the overall result is that SUVA measured at different labs can vary significantly even under carefully controlled conditions (Potter and Wimsatt, 2005). This makes it difficult for water utilities to monitor changes in SUVA, especially when concentrations of DOC and/or CDOM are low. Finally, while useful for predicting a range of treatment variables, water treatment processes like oxidation (Van Geluwe et al., 2011) and adsorption (Newcombe et al., 2002) are not well predicted by SUVA, due to their dependence on low-molecular-weight DOM fractions that are invisible to UV spectrophotometers. Analysis by liquid chromotography with organic carbon detection (LC-OCD) can provide both a more accurate SUVA quantification and can distinguish and quantify several non-absorbing low-molecular-weight DOM fractions, but is too costly for routine monitoring. Consequently, there is a high value in identifying affordable, single-instrument approaches to measuring water aromaticity and tracking low-molecular-weight DOM fractions in drinking water.

Fluorescence spectroscopy is increasingly used to study water treatment due to high sensitivity and affordability, coupled with two-directional wavelength scanning capabilities that inform about water character as well as concentration (Ahmad and Reynolds, 1999; Aiken, 2014; Bhattacharya and Osburn, 2020). Despite accounting for a small fraction of the total dissolved carbon pool in aquatic samples, fluorescent DOM tracks major changes in DOM molecular composition (Kellerman et al., 2015; Stubbins et al., 2014; Wünsch et al., 2018a). Raw fluorescence measurements capture the sum of fluorescence emitted at each excitation and emission wavelength by all dissolved chemicals present, thus combining signals from many different fluorophores. Accordingly, fluorescence spectroscopy is frequently coupled with parallel factor analysis (PARAFAC) to mathematically isolate and quantify independently-varying DOM fractions (Bro, 1997; Murphy et al., 2014a; Stedmon and Markager, 2005b).

A number of prior studies have used PARAFAC to analyse fluorescence datasets in the context of water treatment. Shutova and colleagues (Shutova et al., 2014) reported linear correlations between between ratios of humic-like PARAFAC components vs percent DOC removal in treatments involving coagulation. However, in that study as in many others (e.g. Murphy et al., 2011; Vines and Terry, 2020), PARAFAC models of drinking water resources were site-specific with only modest statistical similarities obtained when comparing models from different treatment plants. This reflects a serious limitation of the PARAFAC approach, namely that selecting a different model tends to produce systematically different fluorescence intensities and ratios. This has greatly limited the generality of conclusions about DOM behaviour and treatability drawn from PARAFAC analyses (Ishii and Boyer, 2012).

There are two main prerequisites for defining useful global models that predict water treatability from fluorescence measurements. The first is that the spectral properties of underlying fluorophore groups are highly similar between geographically diverse water resources. Fortunately, although a diverse array of PARAFAC components appear in published literature (Wünsch et al., 2019), there are strong indications that a handful of well-described PARAFAC components of apparently ubiquitous distribution are responsible for most observed steady state fluorescence (Kowalczuk et al., 2009; Murphy et al., 2018; Wünsch et al., 2019). Several of these have been shown to respond similarly to similar environmental stressors (Ishii and Boyer, 2012; Murphy et al., 2018; Wünsch et al., 2017), suggesting that it may be possible to use PARAFAC components to predict certain aspects of water chemistry across aquatic ecosystems.

A second requirement is that at least some of these PARAFAC components track chemical characteristics with a deciding influence upon water treatability, whether directly or indirectly. Aromaticity, often measured as SUVA, is a key property for predicting treatability in conventional water treatment as already described. Previously, a "fluorescence index" (FIX) that tracks the ratio between emissions at a specific pair of wavelengths in the visible range, was shown to correlate inversely to both aromaticity and SUVA (McKnight et al., 2001; Weishaar et al., 2003). Other fluorescence indicies that are investigated frequently in the context of drinking water treatability include the "freshness" or "biological" index (BIX) and the "humification index" (HIX), which attempt to track the ratio of labile to recalcitrant fractions (Ohno, 2002; Parlanti et al., 2000). Fluorescence indices have the advantage that they are simple to obtain and can indicate environmentally-relevant variations in water quality and reactivity (Bhattacharya and Osburn, 2020; Korak et al., 2015). However, since they are derived from total fluorescence they are susceptible to uncalibrated spectral interferences e.g. from contaminants. It therefore seems likely that greater sensitivity for detecting environmentally-relevant changes in DOM composition might be achievable from tracking ratios between ubiquitous PARAFAC components.

This study assessed the utility of fluorescence spectroscopy combined with PARAFAC for predicting surface water aromaticity and treatability. The aim was to first identify the spectral properties of fluorophore groups common to geographically diverse drinking water samples, and second to identify global relationships between fluorescence, aromaticity and molecular composition (LC-OCD fractions) in order to predict DOC removal in water samples subject to a standard set of drinking water treatments.

2. Materials and methods

2.1. Sampling campaigns

2.1.1. Global campaign

Filtered (0.45 μm) water samples (N=190) were collected from water resources and treatment plants in Europe (France, Sweden, Spain, Italy, n=148), USA (n=19), Asia (China, n=6) and Africa (Cameroon, n=3) (Table S1). Samples were collected in acid-cleaned and combusted amber glass bottles when possible though some were collected in sample-washed HDPE bottles to prevent breakage during international shipping. Each sample represented a single whole water grab sample collected at a particular location on a particular day. Samples from water resources were collected from the surface of rivers or lakes or the intakes of treatment plants. Sampled treatment plants employed a diverse range of treatments ranging from conventional treatment to treatments involving membrane filtration (nanofiltration, reverse osmosis), adsorption onto activated carbon (GAC, PAC), and advanced oxidation (ozone), and samples were collected from various stages in each treatment process.

Samples were air shipped at 4°C to Paris within 48 h of sampling, for fluorescence, absorbance and DOC analysis at Suez research laboratories. Aliquots from a subset of samples (N=63) were additionally shipped in small batches from Paris to Karlsruhe, Germany for analysis by LC-OCD. Overall, the time elapsed between sampling and measurement ranged from 1-7 days for spectroscopic and DOC analyses and 5-10 days for LC-OCD analyses. The only exceptions were the samples from China which were analysed for LC-OCD at a local university laboratory.

2.1.2. Local campaign

Unfiltered 20 L water samples ("natural Challenge Samples") were collected in clean HDPE drums from five French rivers (Table 1) at the intakes to water treatment plants. Each sample (N = 5) was shipped overnight to the Suez laboratory where it was subdivided and used in an experiment consisting of nine standard treatment tests. Each test measured the changes in DOM quantity and character resulting from one type of treatment, as detailed in Section 2.4 below.

Challenge samples from the local campaigns were analysed using absorbance and fluorescence spectroscopy, and DOC analysis. Aliquots from all samples arising from the Bretagne challenge tests were additionally shipped to Germany for LC-OCD analysis.

2.1.3. Synthetic NOM samples

Synthetic challenge samples were prepared by combining three natural DOM isolates (reference materials from the International Humic Substances Society) with various mineral substances to create a suite (N = 20) of samples representing properties that are commonly encountered in raw drinking surface water (Tables S2 and S3). The three DOM isolates were Suwannee River (SRNOM), Nordic Reservoir and Mississippi River dissolved at concentrations representative of average surface water resources (~3 mgL^{-1}). For each isolate, the mineral background was adjusted using $CaCl_2$ to simulate hardness, $NaHCO_3$ to simulate alkalinity and Kaolinite added to simulate turbidity. Aliquots from all samples arising from the SRNOM + 100 mgL^{-1} $CaCO_3$ challenge experiment were shipped to Germany for LC-OCD analysis.

2.2. Analytical methods

DOC analysis was performed in the Suez Lab using a benchtop analyzer (Aurora 1030W, OI instruments) that uses persulfate wet oxidation to quantify organics in water samples. UV absorbance at 254 nm, representing chromophoric dissolved organic matter (CDOM), was measured using a Nanocolor UV/Vis II (Macherey Nagel) and was used in combination with DOC to calculate specific UV absorbance (SUVA = UV_{254}/DOC).

A spectrophotometer (Aqualog, Horiba-Jobin Yvon) was used to measure fluorescence excitation emission matrices (EEMs) on the filtered water samples. EEMs were made from scans across the selected range of excitation wavelengths (excitation 210-620 nm with 5-nm bandpass), with emission spectra simultaneously recorded by CCD detector (220-800 nm at 4-nm increments). Fluorescence data exported from the AquaLog were fully corrected for instrumental and sample-related biases. First, spectral corrections accounting for wavelength-related biases were applied using factory-generated excitation and emission correction factors. Second, fluorescence quenching caused by sample self-shading (inner filter effect) was corrected in the instrument, using correction factors derived from the fluorometers's simultaneously collected absorbance spectrum (Kothawala et al., 2013).

LC-OCD analyses were performed by DOC-Labor in Karlsruhe, Germany, according to published procedures (Huber et al., 2011). This technique uses size exclusion chromatography to quantify five DOC subfractions, operationally classified as low-molecular weight (LMW) neutrals (alcohols, aldehydes, ketones, amino acids, sugars, <350 g mol^{-1}), LMW acids (aliphatic organic acids, <350 g mol^{-1}), building blocks (BB) representing deaggregated chains of polyphenolics and polyaromatic acids formed during the breakdown of humic substances, 300-500 g mol^{-1}), humic substances (HS, 500-1000 g mol^{-1}) comprising a mixture of humic acids containing carboxyl and phenolate groups, and biopolymers (>1000 g mol^{-1}). LC-OCD analysis also measures DOC concentration and SUVA (total UV absorbance / total DOC).

2.3. Statistical methods

Parallel Factor Analysis was used to mathematically decompose the global fluorescence EEM dataset according to well-established methodologies (Bro, 1997). PARAFAC identifies the unique underlying spectra of a fixed number of independent components that sum to reproduce the measured EEMs with least error and determines the relative concentrations (F_{max}) for each component. Modelling was implemented using the drEEM toolbox in MATLAB (Murphy et al., 2013). Extensive modelling iterations confirmed that a nine-component model provided the best fit to the dataset. However, only post-membrane samples from a pilot plant contained the ninth component; and this component had spectral properties indicating a whitening agent (Section 3.3). After excluding outlier samples affected by whitening agents, only eight components remained in the model. The 8-component model was validated both by split-half analysis (S2T1), and by separately modelling short-wavelength (protein-like) and longer wavelength (humic-like) regions of each EEM and fusing the resulting spectra. After the model was finalized, the fluorescence EEMs from challenge samples and outliers were projected on the model, producing F_{max} for every sample in the study.

F_{max} ratios were determined by dividing F_{max} for one PARAFAC component by F_{max} for another. Regression equations relating F_{max} ratios with other measurements were determined in MATLAB (2020a) using either a standard linear model for small sample sizes ($N \leq 5$) or a robust linear model determined by iteratively reweighted least squares regression ($N \geq 10$) (Holland and Welsch, 1977). To determine prediction errors ($RMSE_p$) for regression analyses, regression models were tested on a prediction dataset consisting of randomly selected samples that had not been used to develop the regression model. In the case of regressions involving PARAFAC components, the prediction dataset consisted of the five challenge samples together with 19 other samples that were collected after the model had been finalized. For these samples, PARAFAC scores were determined by projecting the EEMs on the existing PARAFAC model, then regression fits were determined by projecting those scores on the existing regression equations. In the case of fluorescence ratios, the prediction samples were selected randomly by algorithm, then projected on the regression equation derived from the remaining samples.

2.4. Treatability experiments

The natural and synthetic challenge samples were submitted to nine different treatments at bench-scale (Table 2). First, samples were equilibrated with lab atmosphere and temperature then divided into subsamples that were individually submitted to one test. Coagulation, enhanced coagulation, powdered activated carbon (PAC), ion exchange (IEX), and chlorination were each tested individually. Ozone (O_3) was tested individually and in combination with coagulation, IEX and PAC.

Table 1
Properties of natural river water samples in the challenge dataset

River	Sampling date	Alkalinity (mgL^{-1} $CaCO_3$)	Hardness (mgL^{-1} $CaCO_3$)	Turbidity (NTU)	DOC mgL^{-1}	UV_{254} (m^{-1})	SUVA ($Lmg^{-1}m^{-1}$)
Seine	3/2018	270	270	8	2.5	6.1	2.4
Sarthe	4/2018	150	120	7	5.7	17.5	3.1
Loire	4/2018	70	50	54	5.6	18.4	3.3
Bretagne	6/2018	80	30	5	2.9	11.9	4.2
Nancy	6/2018	60	50	10	3.6	12.5	3.5

Table 2
Operational conditions in lab-scale treatability experiments

Treatment	Code	Chemicals	Doses	Method details
Coagulation	Coag	$FeCl_3$ commercial solution (40%)	40 mgL^{-1}	3 min flash mix (200 rpm) followed by 17 min of slow mixing (40 rpm) and at least 10 min of settling
Enhanced coagulation	Adv. Coag		200 mgL^{-1}, pH adjusted to 5.8	
PAC	PAC	SA Super (NORIT)	40 mgL^{-1}	
Ion Exchange	IEX	Amberlite IRA458Cl	5 mL of activated resin in 1L, 200 BV equivalent	15-min mixing at 150 rpm followed by 15min of settling
Ozonation	O3	Ozone	3 mgL^{-1}	5-min contact time, manual shaking, verification of no residual ozone
O_3 + coagulation	O3 + coag.	Ozone and $FeCl_3$	3 mgL^{-1} O_3, 40 mgL^{-1} $FeCl_3$	5-min O_3 contact time with shaking followed by coagulant injection with 3 min flash mix (200 rpm) followed by 17min of slow mixing (40 rpm) and at least 10min of settling
O_3 + PAC	O_3 + PAC	Ozone and SA Super	3 mgL^{-1} O_3, 40 mgL^{-1} PAC	
O_3 + IEX	O_3 + IEX	Ozone and IRA458Cl	3 mgL^{-1} O_3, 5 mL of resin	5-min O_3 contact time with shaking followed by resin injection with 15 min mixing at 150 rpm followed by 15min of settling
Chlorination	Chlor.	Chlorine	3 mgL^{-1}	5-min contact time, quenched with NH_4Cl

These tests were completed using a jar test set-up with 1L cylindrical beakers. Ozone was tested using a specially designed container allowing for gas injection through a septum.

The contact times and chemical doses corresponded to conventionally applied operational conditions for lab-scale simulation of surface water treatment plants. Ferric chloride coagulant, SA Super powder activated carbon and Amberlite IRA458Cl ion-exchange resin were selected due to their specific capacity for organic matter removal. The same doses were administered to each synthetic and natural water sample. Since the synthetic samples each had similar DOC concentrations, they each received similar effective doses, making removal a function of only DOM source and mineral matrix. DOC concentrations in the natural water samples varied between 2.5-5.7 mgL^{-1}, making removal a function of both DOM source and effective dose. Two natural samples had nearly identical DOC concentrations (Sarthe, Loire) and received similar effective doses.

3. Results and discussion

3.1. Fluorescence composition

Fig. 1 depicts the spectral properties of all nine PARAFAC components identified in the global EEM database and Table 3 lists their similarity with published spectra (Tucker congruence coefficient, TCC) according to the OpenFluor database (Murphy et al., 2014b) which at the time of writing contained 237 published PARAFAC models. Raw spectral data for the PARAFAC components in this study will be made available upon publication in OpenFluor ("Suez_9C").

Five identified components (H_i, H_{ii}, H_{iii}, H_{iv}, W_i) each had emission maxima at wavelengths longer than 400 nm as is characteristic for fluorescence derived from humic- and fulvic-like organic matter. All have been identified previously and recent studies highlight their apparently ubiquitous distribution (Table 3). Three published PARAFAC models of drinking water samples are particularly strongly congruent. A 5-component model of Swedish river water "Kungalv_5" (Moona et al., 2021) is practically a perfect match for five components in the current model (Fig. S1). Additionally, five of six PARAFAC components in the

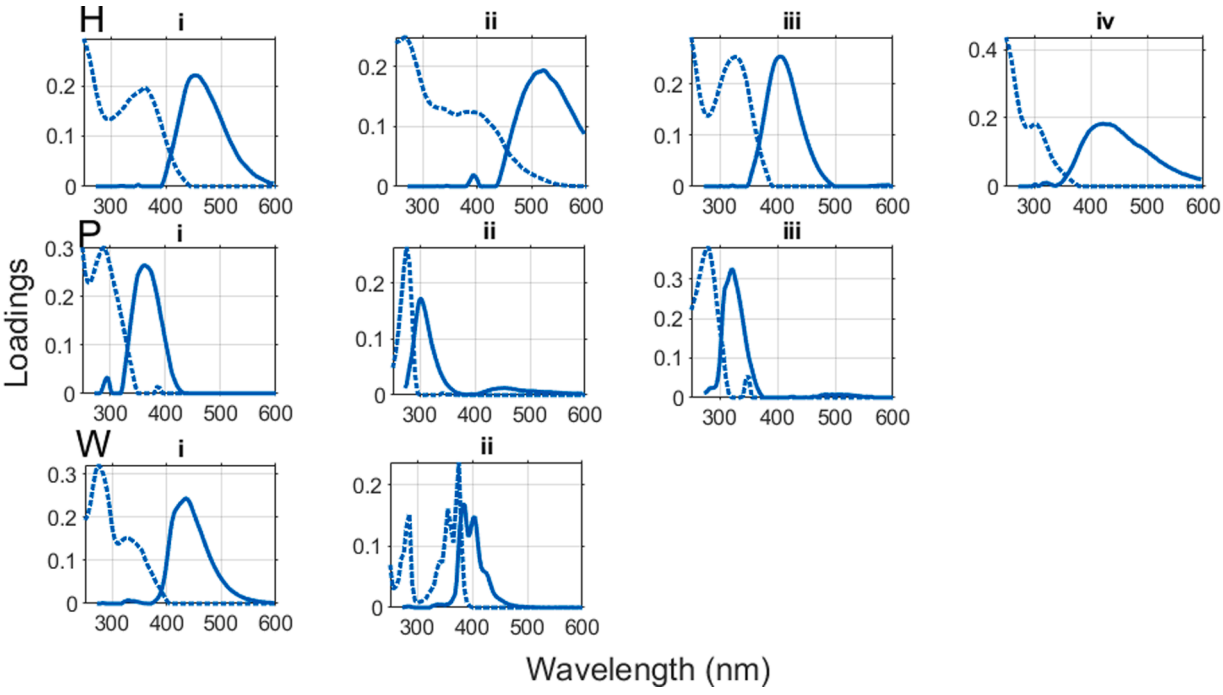

Fig. 1. Underlying components in the nine component PARAFAC model include ubiquitous humic-like fractions ($H_{i\text{-}iv}$, top row), protein-like fractions ($P_{i\text{-}iii}$) and wastewater signals that may be related textile products and optical brighteners ($W_{i\text{-}ii}$) .

Table 3
Description of fluorescent fractions according to PARAFAC analysis

Name	peak position ($\lambda_{ex}^{2'}/\lambda_{em}$)	No. matches in OpenFluor (TCC_{ex} >0.95/ TCC_{em} >0.98)	Possible sources / previous attributions	Ref. (top matches)
H_i	360/455	33	Humics: ubiquitous F_{450}	Garcia et al. 2018; Cardenas et al. 2017; Kowalczuk et al., 2009
H_{ii}	395/521	36	Conjugated humics: ubiquitous F_{520}	Moona et al. 2021; Wünsch et al. 2017; Murphy et al. 2018
H_{iii}	330/404	16	Microbially-linked humics: ubiquitous F_{400}	Asmala et al. 2018; Kulkarni et al. 2017; Yamashita et al. 2010
H_{iv}	300/430	49	Terrestrial humics: ubiquitous $F_{42_{0.}{}_{4}30}$	Wünsch et al. 2018b; Moona et al. 2021
W_i	275/434	1	Pulp/paper industry contaminant	Cawley et al. 2012
P_i	290/365	4	Amino acids-tryptophan	Murphy et al. 2013; Zito et al. 2019; Kowalczuk et al., 2009
P_{ii}	275/302	6	Amino acids-tyrosine	Wünsch et al. 2015
P_{iii}	275/323	2	Petroleum derivatives	Zito et al. 2019; Yang et al. 2019; Huang et al. 2022
W_{ii}	375/385	1	Optical brightener	Cohen et al. 2014

SAB6 model of seawater from the South Atlantic Bight (Kowalczuk et al., 2009) are very highly congruent (correlation coefficients exceeding 0.97 when comparing excitation or emission spectra). Finally, three components in drinking water models from Australia and Sweden ("Model D" and "DW_Sweden_Gavle", respectively) are similarly highly congruent (Heibati et al., 2017; Shutova et al., 2014).

Component H_i coincides with the "fulvic-like" "C" peak position (Coble et al., 1990) with spectrally similar components common to virtually all PARAFAC models of DOM fluorescence (Table 3). In samples from diverse freshwater systems, H_i is typically the first or second most important/abundant component in the model accounting for a large proportion of total variability. H_i tends to covary with H_{ii} in surveys from river to sea, with the latter being associated with higher aromaticity and greater susceptibility to flocculation processes in natural systems and water treatment (Kothawala et al., 2014; Yamashita et al., 2010). In PARAFAC models with only three or four components, H_i and H_{ii} are often combined indicating stable relative abundances across many datasets (Wünsch et al., 2019). However, H_{ii} has more humic character and represents are more aromatic and higher-molecular weight fraction than H_i (Cuss and Guéguen, 2015).

Component H_{iii} is reported to represent soluble microbial products other than proteins (Wells et al., 2022) and is typically abundant in wastewater DOM and at agriculturally-impacted and eutrophic sites (Stedmon and Markager, 2005a; Yamashita et al., 2010). Its emission peak overlaps with the position of the microbial "M" peak (Coble, 1996; Stedmon and Markager, 2005b; Wünsch et al., 2019), although "M" also captures emission wavelengths associated with H_i and H_{iv}. Although produced by microbial activity H_{iii} is itself resistant to biological degradation (Ishii and Boyer, 2012; Moona et al., 2021) and tends to accumulate in sediments (Krasner et al., 1996). H_{iii} is additionally resistant to coagulation/flocculation (Kothawala et al., 2014) and has been implicated in irreversible membrane fouling (Ly et al., 2018). Component H_{iv} has a broad peak with slightly longer emission maximum than H_{iii} and is only typically found in models having four or more humic-like components, sometimes with a spectral shape indicating a combination of protein-like and humic-like fractions (Kowalczuk et al., 2010; Murphy et al., 2018; Wells et al., 2022). Unlike H_{iii}, H_{iv} is readily biodegraded (Cohen et al., 2014; Moona et al., 2021). The large spectral overlap between H_{iii} and H_{iv} despite different sources and reactivities accounts for sometimes contradictory reports regarding the behaviour of components responsible for peak M fluorescence.

Components P_i and P_{ii} match protein-like fluorescence similar to pure tryptophan (P_i) and pure tyrosine (P_{ii}). P_i emits at slightly longer wavelengths than is usually observed for pure tryptophan (λ_{em} ~ 330-350) and may represent multiple unresolved protein-like fluorophores. Signals similar to P_{iii} have been reported from Chinese surface waters (Yang et al. 2019; Huang et al., 2022) and from petroleum-derived fluorescence (Zito et al., 2019) although this does not preclude the possibility of a biological origin.

Only two components identified in this study are previously unreported from drinking water resources. One of these, W_i, is a near-perfect spectral match for a component identified previously as pulp mill effluent (Cawley et al., 2012); it was abundant only in the challenge sample from river Nancy (Fig. S2) where several pulp mills are located upstream of the sampling location. The other (W_{ii}) matches a spectrum isolated from wastewater (Cohen et al., 2014) and has a multi-peaked emission spectrum characteristic of optical brighteners that are used widely in textile and paper industries (Fig. S1). In this study, W_{ii} was the ninth component and observed only in some samples of membrane permeates and not in any challenge test samples. The isolation of these components from a diverse drinking water dataset highlights the sensitivity of fluorescence spectroscopy in combination with PARAFAC for detecting the potential organic contamination of drinking water samples.

3.2. *Surface water aromaticity*

SUVA in the dataset ranged between 1.0 and 5.5 $Lmg^{-1}m^{-1}$ and thus spanned a range from low to very high aromaticity. Fig. 2 shows correlations between established and newly derived predictors of surface water aromaticity. In the first two subplots, raw water samples accounted for most measurements in the upper right portion of each subplot (SUVA 1.7 - 5.5, mean 3.1 ± 0.8 $Lmg^{-1}m^{-1}$) and partially or fully treated water samples accounted for the lower left portion (SUVA 1.0 – 3.7, mean 1.8 ± 0.6 $Lmg^{-1}m^{-1}$). However, a single linear regression line fits all samples equally well, with no indication that treatment altered the underlying relationship between SUVA and fluorescence.

A better predictor of SUVA discovered in this study was the ratio between F_{max} for the longest emitting PARAFAC component (H_{ii}, λ_{em}=520 nm) versus the shortest emitting component within the "humic" region (H_{iii}, λ_{em}=404 nm), hereafter referred to H_{ii}/H_{iii} or "PARIX" standing for "PARAFAC Index". Note that H_{iii} fluorescence overlaps spectrally with component H_{iv} (λ_{em}=420 nm), although these two signals have distinct sources and behave differently during water treatment (see further below).

In Fig. 2, circular markers indicate the calibration dataset, i.e., 166 samples from Cameroon, China, France, Italy and USA that were included when building the PARAFAC model, of which 53 had matching LC-OCD measurements. Red stars indicate the prediction dataset, consisting of samples that were excluded from PARAFAC modelling. The prediction dataset consists of 24 samples from Spain, USA and France (including the five natural Challenge Samples) of which 10 samples had LC-OCD measurements. For prediction samples, PARAFAC scores were estimated by projecting the sample EEMs on the model. With the one exception described below, no data points were omitted from any graph, so the depicted correlations are robust across continents and ecosystems with visible outliers that show the range of departure to be expected due

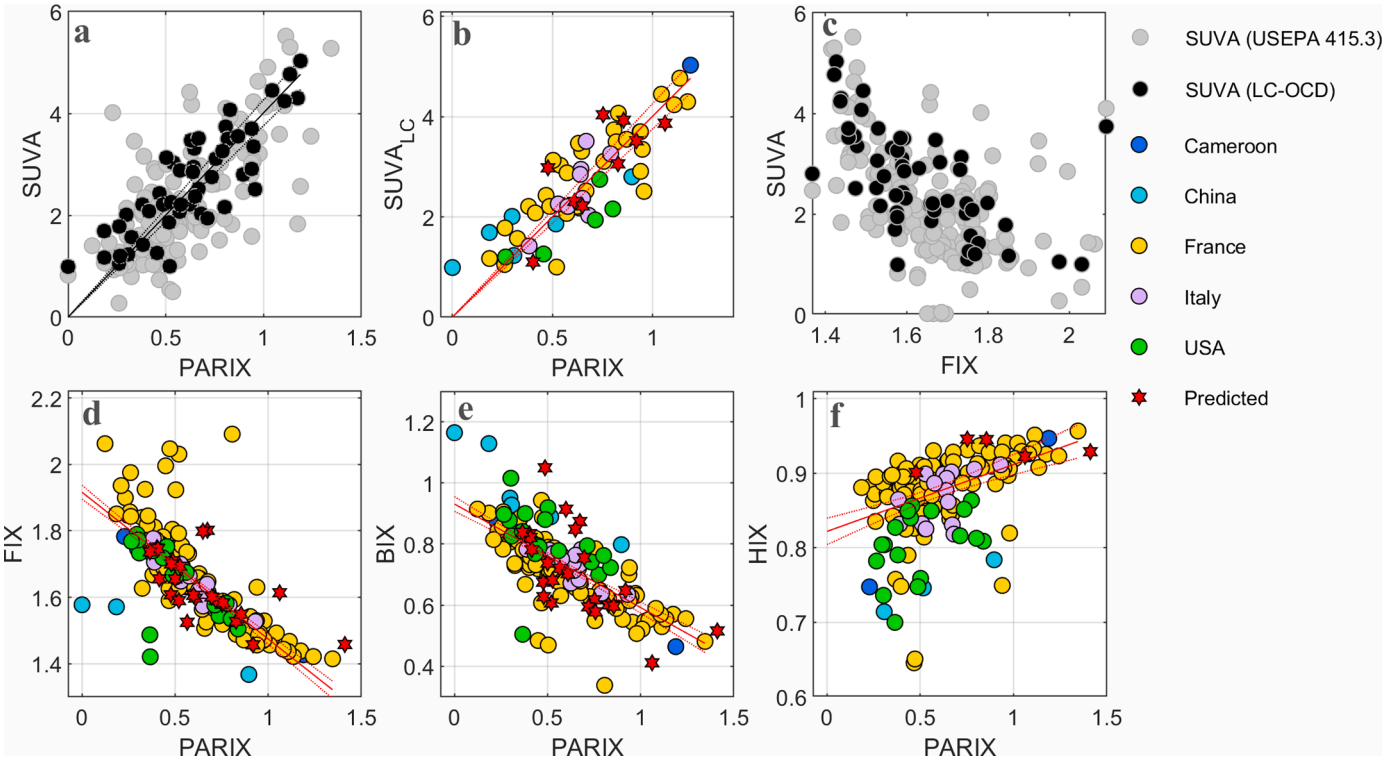

Fig. 2. Relationships between SUVA, H_{ii}/H_{iii} (PARIX) and traditional fluorescence indices (FIX, BIX, HIX) in raw and treated surface waters from six countries. Units of SUVA are $Lmg^{-1}m^{-1}$; all other depicted parameters are unitless. Regression equations and fit statistics are in Table 4.

to combined sampling, measurement, and modelling errors. The single exception was the HIX subplot, where six extreme outliers from France and China were removed for visual clarity (HIX in the range of 0.15-0.6 and PARIX below 0.5).

Fig. 2a correlates PARIX with SUVA measurements made according to the traditional (USEPA) method and by LC-OCD. A much stronger linear correlation is seen with the LC-OCD dataset than with the traditional SUVA dataset, indicating that much of the scatter among grey points in Fig. 2a is attributable to measurement error in the traditional SUVA dataset. For this reason, and because it is typical to favour the LC-OCD dataset in cases of disagreement (Hutchins et al., 2017), the LC-OCD dataset was used to derive equations to predict SUVA from fluorescence. Fig. 2a-b and Table 4 show that SUVA in the calibration dataset could be estimated with good accuracy by simply multiplying PARIX by a constant equal to four (RMSEc = 0.64 $Lmg^{-1}m^{-1}$). This applied to surface water samples in various stages of water treatment and was the case even for the prediction dataset, i.e., when this ratio was estimated by projecting new EEMs upon the pre-existing PARAFAC model (RMSEp = 0.55 $Lmg^{-1}m^{-1}$). Note that the magnitude of the constant in a regression equation involving PARAFAC components depends on their spectral ranges and shapes. Thus, this constant may vary between datasets if fitted with slightly different PARAFAC components. Conversely, a general result e.g., a linear relationship between SUVA and the ratio of components representing H_{ii} and H_{iii}, would be expected to hold for two different PARAFAC models.

It is relevant to compare the predictive power of PARIX with the predictive power of FIX and other well established fluorescence indices. Fig. 2c-d shows that FIX is non-linearly correlated to SUVA and PARIX, which mirrors early observations of nonlinearity between FIX and aromaticity when measured by NMR (McKnight et al., 2001). Other indices that primarily track the ratio between tryptophan-like and humic-like fluorescence (Freshness Index or BIX, and HIX) (Ohno, 2002; Parlanti et al., 2000) also show some non-linearity plus significant scatter when correlated with PARIX (Fig. 2c, e, f).

Table 4 lists robust regression equations and fit statistics for predicting SUVA from fluorescence ratios and PARIX. Since these relationships were developed from diverse raw and treated surface waters from three continents, they may hold widely for freshwaters in other geographical locations, except for scaling factors that will apply to PARIX estimated using different PARAFAC models. Overall, the established indices each had weaker correlations and lower accuracy for predicting SUVA compared to PARIX (Table 4). Furthermore, PARIX was only marginally worse at predicting SUVA measured by LC-OCD than was SUVA measured according to the standard method (Table 4). This is presumably due the higher measurement precision of fluorescence measurements compared to absorbance and DOC measurements.

Table 4
Equations for predicting SUVA by robust linear regression on one variable. Note that regression constants may vary between models using different spectra to model H_{ii}, H_{iii} and P_i due to different scaling factors. $SUVA_{LC}$ in the dataset ranged from 0.99-5.0 and PARIX from 0-1.3.

Predicted: *y*	Predictor: *x*	Equation	N_{cal}, N_{pred}	R_c^2	$RMSE_c$	$RMSE_p$
$SUVA_{LC}$	$SUVA_{trad}$	1.12*x*	47, 16	0.87	0.45	0.40
$SUVA_{LC}$	PARIX (H_{ii}/H_{iii})	4.0*x*	47, 16	0.78	0.64	0.55
$SUVA_{LC}$	H_{ii}/P_i	1.4+1.3*x*	53, 10	0.70	0.60	0.47
$SUVA_{LC}$	FIX	11.4-5.40*x*	47, 16	0.55	0.75	0.62
$SUVA_{LC}$	BIX	6.5-5.25*x*	47, 16	0.51	0.75	0.57
$SUVA_{LC}$	HIX	2.8-0.90*x*	47, 16	0.16	1.0	0.76

3.3. Treatability

Fig. 3 compares treatment efficiencies determined in bench-scale experiments on natural and synthetic challenge samples. Treatment efficiencies are shown as percent removals of bulk DOM (DOC), its total chromophoric fraction (CDOM) and its total fluorescent fraction (FDOM). Differences in treatment efficiencies between the synthetic

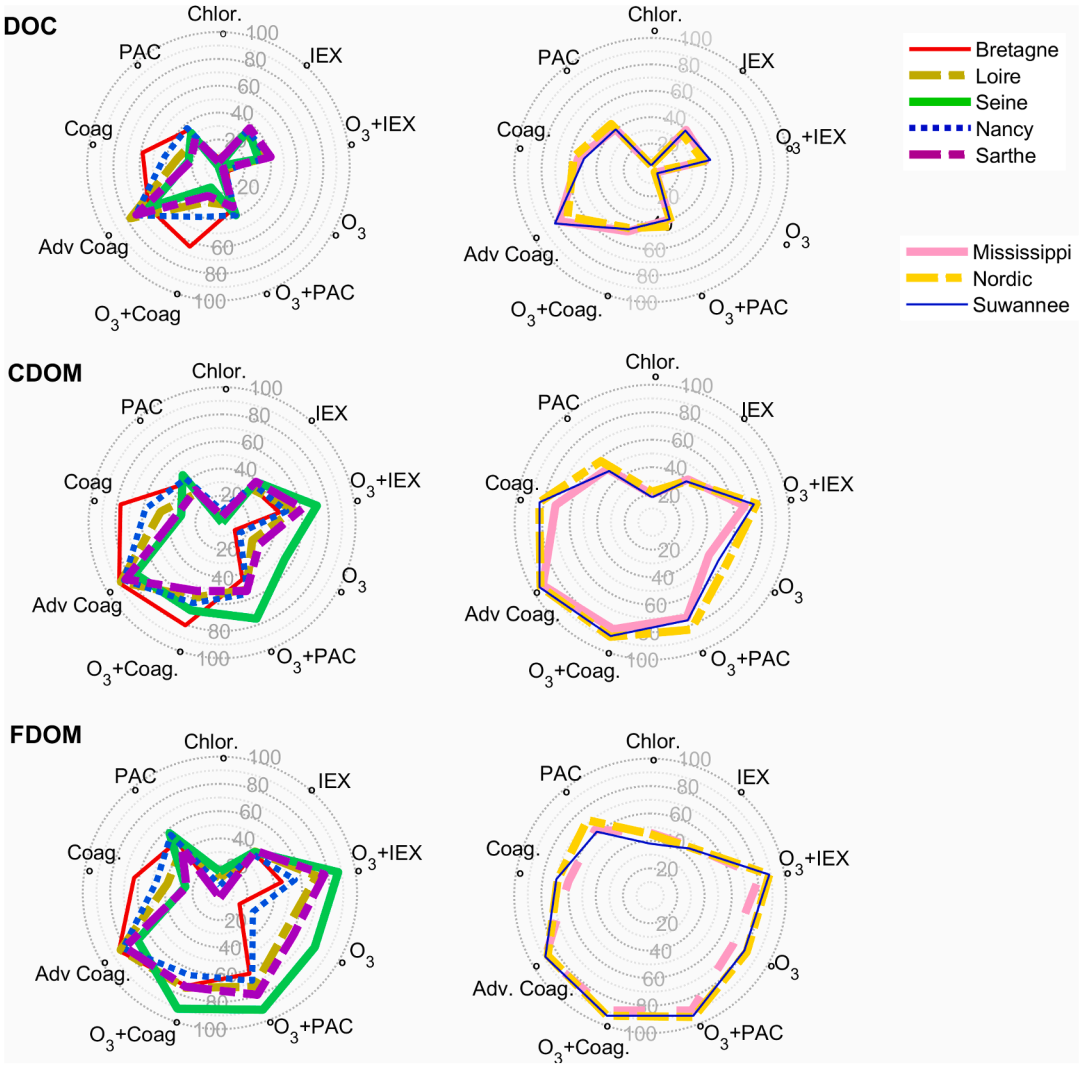

Fig. 3. Treatment efficiency (% removal) for DOC, CDOM and FDOM in natural (left) and synthetic challenge samples (right). Each sample was subjected to nine different treatments as defined in Table 2. The natural samples were sourced from five French rivers: Bretagne , Loire , Seine , Nancy , and Sarthe. The synthetic samples were DOM isolates from Suwanee River, Nordic Reservoir, and Mississippi River.

samples were negligible, even when deliberately varying their background mineral content, thus Fig. 3 shows average removals for the complete set of trials performed with each isolate. Fig. 3 shows several expected trends including much better removal of CDOM than DOC and slightly better removal of FDOM than CDOM. Exceptions included IEX and PAC, which had similar removal efficiencies for DOC as for CDOM.

The isolates showed little variation in removal efficiencies when subjected to a particular treatment, whereas the natural samples showed a wider range consistent with their variable DOM characters. Removal efficiencies for the isolates generally tracked the highest removal efficiencies measured for natural samples, i.e. similar to the Bretagne sample in treatments involving coagulation, and similar to the Seine sample in treatments involving PAC and ozonation. Biopolymers were largely excluded in the process of extracting DOM isolates (Table S4), and this may have resulted in water that is relatively easy to treat by a wide range of processes.

Applied on its own, ozonation did not significantly remove DOC from river samples. This is consistent with results from prior studies demonstrating that larger molecular weight fractions (humics and biopolymers) are not removed by ozonation so much as converted to lower molecular weight fractions (Sarathy and Mohseni, 2007). Additionally, pre-ozonation did not improve DOC removal when followed by PAC, IEX or coagulation, but did improve the removal of CDOM and FDOM, particularly for the Seine sample. This does not seem to reflect a variation in ozone dose since both Seine and Bretagne samples had a similar DOC concentration, but for Bretagne only, pre-ozonation had little effect.

DOC removal by conventional coagulation was high especially among samples with high SUVA (Bretagne, Nancy), as has been reported in many previous studies (Archer and Singer, 2006; Edwards, 1997). The performance of PAC and IEX for removing DOC was rather similar across all natural and synthetic samples (35.5 $\pm$ 6 % and $\pm$ 3 % respectively), suggesting that all these had similar-sized adsorbable (around 27 - 45%) and "charged" (31 - 40%) fractions. PAC removed slightly greater proportions of chromophoric and fluorescent material than did IEX (43% and 56% compared to 38% and 42%).

Fig. 4 compares the average performance of each treatment barrier when removing fluorescent DOM fractions from natural and synthetic waters. Consistent with earlier results (sections 3.1 and 3.2), performance on synthetic challenge samples was always better than, or equal to, performance on the natural samples. However, comparing natural and synthetic samples, each fraction's relative treatability by different treatment barriers was strikingly similar.

Overall, humic-like components H_i, H_{ii}, H_{iv}, and W_i were well removed by coagulation, but H_{iii} was removed relatively poorly. Published reports on the behaviour of components like H_{iii} are often

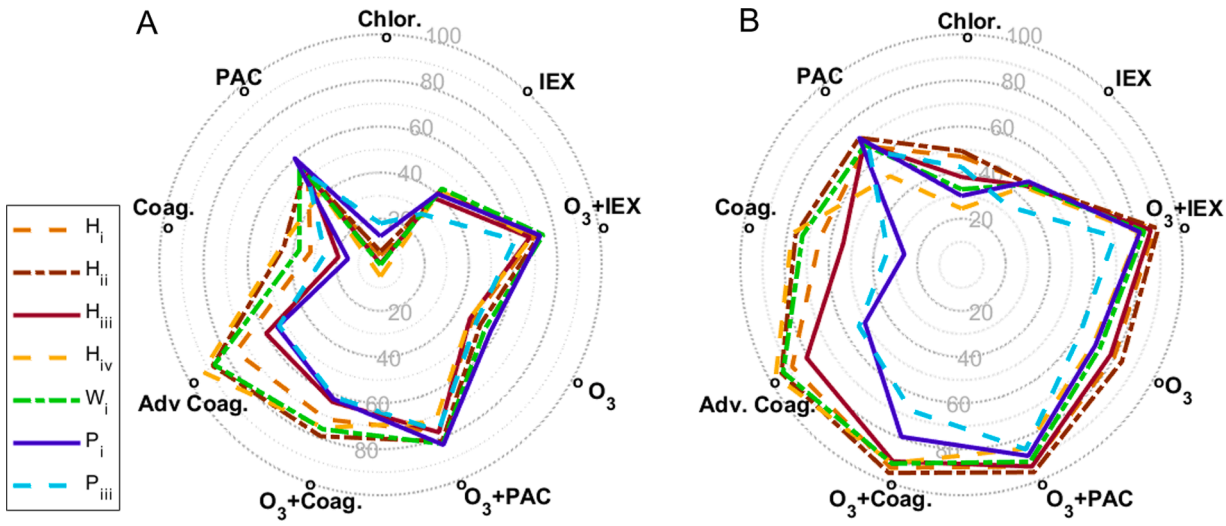

Fig. 4. Treatment barrier performance for removing fluorescent DOM fractions as defined in Table 3. Lines show average removal efficiency for five natural (A) and three synthetic (B) challenge samples when subjected to nine different treatment barriers (see Table 2).

contradictory because many PARAFAC models have H_{iii} and H_i combined in a single component, although in models where they are separated, the two vary independently (Chen et al., 2010; Ishii and Boyer, 2012; Palma et al., 2021). This is probably the reason why H_{iii} is sometimes described as being of terrestrial (allochthonous) origin, and sometimes as 'microbially-derived' (autochthonous). In this study, the only treatments that effectively removed H_{iii} involved ozonation. Similarly, protein-like components P_i and P_{ii} were always difficult to treat (the low abundance of P_{iii} precluded studying its removal efficiency). PAC or ozone in combination with other barriers were most effective at removing protein-like fluorescence.

3.4. Predicting treatability from FDOM ratios

Fig. 5 shows treatability results for the five French rivers. Since there were nine different treatments to apply, only a small number of different raw waters could be tested, however SUVA in these samples spanned most of the normal range observed in freshwater systems (approx. 1.6 – 6) (Kellerman, 2015; Massicotte et al., 2017; Weishaar et al., 2003) and in total Fig. 5 shows data from 40 separate pilot tests.

Treatability, assessed as percent DOC removal, was generally well predicted from the fluorescence composition of raw river, when represented as ratios between H_{iii} and either H_{ii}, P_i or P_{ii}. Especially, since PARIX was highly correlated with SUVA, it was also an excellent predictor of water treatability by any process in which aromaticity plays a deciding role, especially coagulation and ozonation and combinations thereof. In contrast, there was typically no correlation between initial DOC concentration and DOC removal which indicates that differences in effective dose was not a major factor determining treatability in these experiments. This is seen by the Seine and Bretagne samples occupying the extremes of most graphs despite these samples having similar initial DOC concentration. There were two treatments for which DOC removal was uncorrelated with fluorescence composition but was somewhat correlated with initial DOC concentration; these were enhanced coagulation ($R^2 = 0.72$) and ozonation combined with ion exchange ($R^2 = 0.47$).

Fig. 5 shows that high PARIX resulted in strong coagulation and ozonation performance, while low PARIX resulted in better performance by ion-exchange. Since LC-OCD analysis confirmed that low PARIX indicated of a higher proportion of LMW molecules (see below), water with lower PARIX might be expected to adsorb more strongly onto activated carbon (Shimabuku et al., 2017)); however, this was not observed for the tested river waters. Instead for PAC, treatability was strongly correlated to P_i/H_{iii} and uncorrelated to PARIX. Previous work indicates that adsorption onto activated carbon involves a complex interplay of factors related to molecular size, polarity and aromaticity as well as the number and size of available adsorption sites, thus can vary over time and according to the degree of saturation of the carbon substrate (Lee and Hur, 2016; Moona et al., 2018). This leads to apparently contradictory results whereby adsorption may be favoured by either high aromaticity or low molecular size, depending on substate properties and water chemistry. In this study, the relationship between P_i/H_{iii} and DOC removal by PAC indicates a competition for adsorption sites within the lower molecular weight fraction, with H_{iii} excluding small amino acids and non-fluorescent molecules from PAC adsorption sites.

3.5. Fluorescence surrogates of LC-OCD fractions

Correlations between fluorescence and LC-OCD measurements were explored after first normalising LC-OCD concentrations to humic substance (HS) concentration, and normalising fluorescence parameters to H_{ii} intensity. This avoids apparently strong correlations driven by concentration artefacts, i.e., higher DOM concentrations tended to produce higher signals for both fluorescence and LC-OCD parameters. Additionally, eight samples with HS concentration below 600 µg/L were deleted due to low signal to noise, leaving 60 samples in the dataset.

The aromaticity index was strongly correlated to nominal molecular weight, as would be expected from its relationship with SUVA (Fig. 6). The inverse of PARIX was strongly positively correlated to (HS-normalised) concentrations of building blocks. The regression equation goes through the origin (BB/HS = 0.19/PARIX, RMSE=0.073, R^2=0.86) indicating a parsimonious model with good predictive capacity. H_{iv}/H_{ii} was also positively correlated to normalised building block concentrations (BB/HS = 0.20* H_{iv}/H_{ii}, RMSE=0.144, R^2=0.73). The building block fraction in LC-OCD analysis is attributed to weathering and oxidation products of humic substances (Huber 2011). This aligns with interpretation that H_{iii} is due to aquagenic refractory organic matter with relatively high N:C ratio, whereas H_{iv} represents labile degradation products, including photochemically-produced DOM (Murphy et al., 2018). The building block fraction is of lower molecular weight than other humic substances and experiments have shown it to be poorly removed by coagulation/flocculation (Huber et al., 2011), which is consistent with the treatability tests (Fig. 5).

Several weaker correlations were seen between (H_{ii} -normalised) fluorescence intensities and (HS-normalised) concentrations of LMW neutrals. The inverse of PARIX was moderately correlated (LMWn/

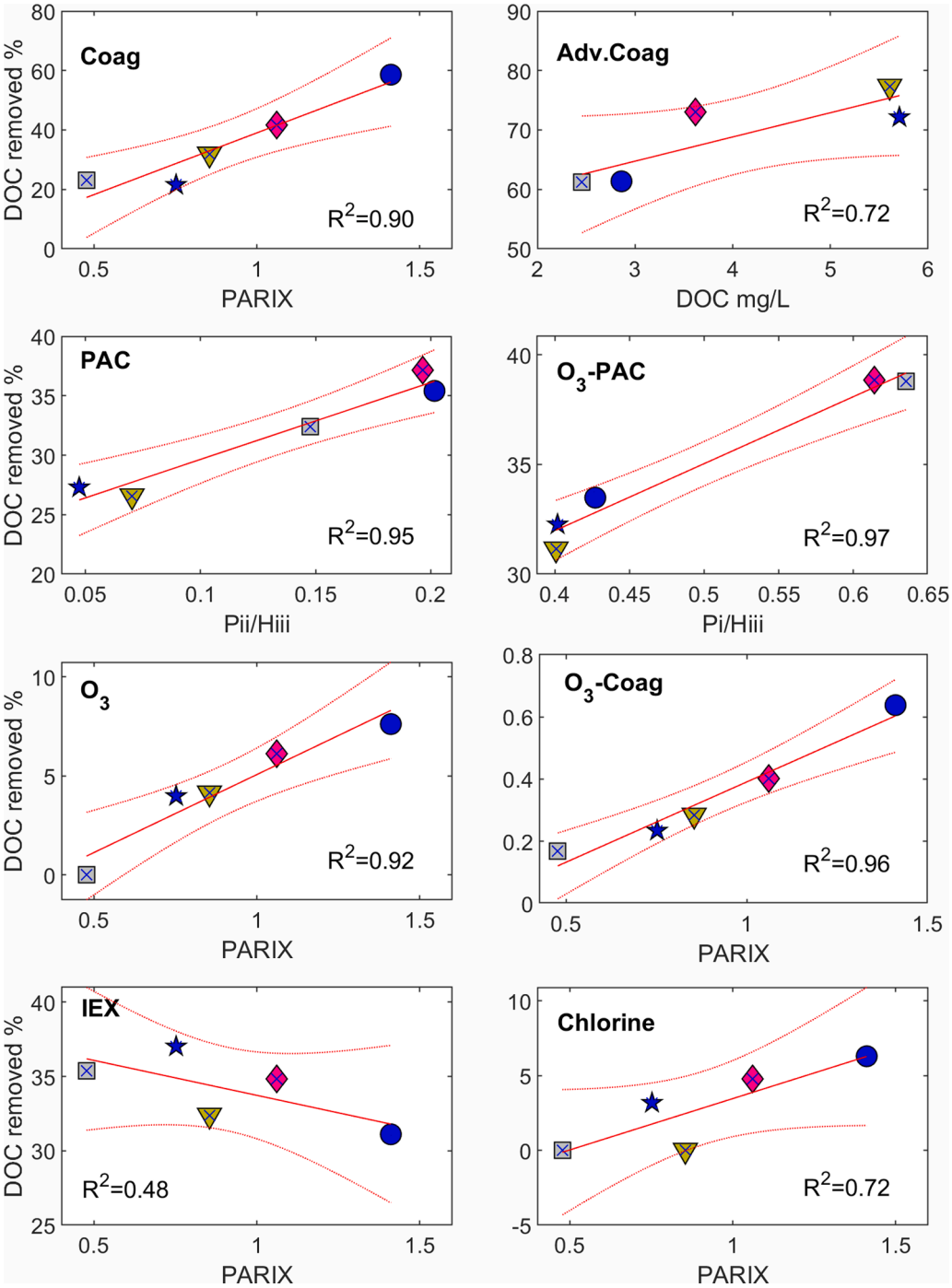

Fig. 5. Relationship between surface water composition and its treatability (DOC removal) in challenge experiments. The symbols depict experiments on samples from five french rivers: Bretagne (circle), Nancy (diamond), Seine (square), Loire (triangle) and Sarthe (star).

HS=0.04+0.14/PARIX, R^2=0.67, RMSE=0.091) as was P_i/H_{iii} (LMWn/HS=0.16+0.07*Pi/H_{ii}, R^2=0.63, RMSE=0.080). Tryptophan-like P_{ii} was only weakly correlated with LMW neutrals ($R^2 < 0.5$) due to considerable scatter. Positive correlations with tyrosine-like and tryptophan-like fluorescence are expected since amino acids belong to the LMW neutral fraction (Huber 2011). However, the correlation with H_{iii} may be indirect rather than causal, i.e., a result of degradation processes tending to produce both amino acids and building blocks at the same time.

There were no correlations observed between any fluorescence component and either biopolymers or LMW acids. An earlier study of drinking water treatment reported a link between biopolymer concentrations and protein-like fluorescence (Baghoth et al., 2011)); however, in that analysis the LC-OCD measurements were not normalised to DOC which is needed to account for bulk DOC removal between treatment steps, and hence nearly all investigated correlations were highly statistically significant. Hutchins and co-workers (Hutchins et al., 2017) reported that tryptophan-like fluorescence and biopolymers co-varied along a stream-water continuum, after first normalising tryptophan-like fluorescence to total fluorescence and biopolymers to DOC concentration. Although they did not draw attention to additional correlations, their principal components analysis additionally indicated the existence of correlations that closely match this study, i.e., between components similar to H_{iii} and H_i and both the building block and LMW

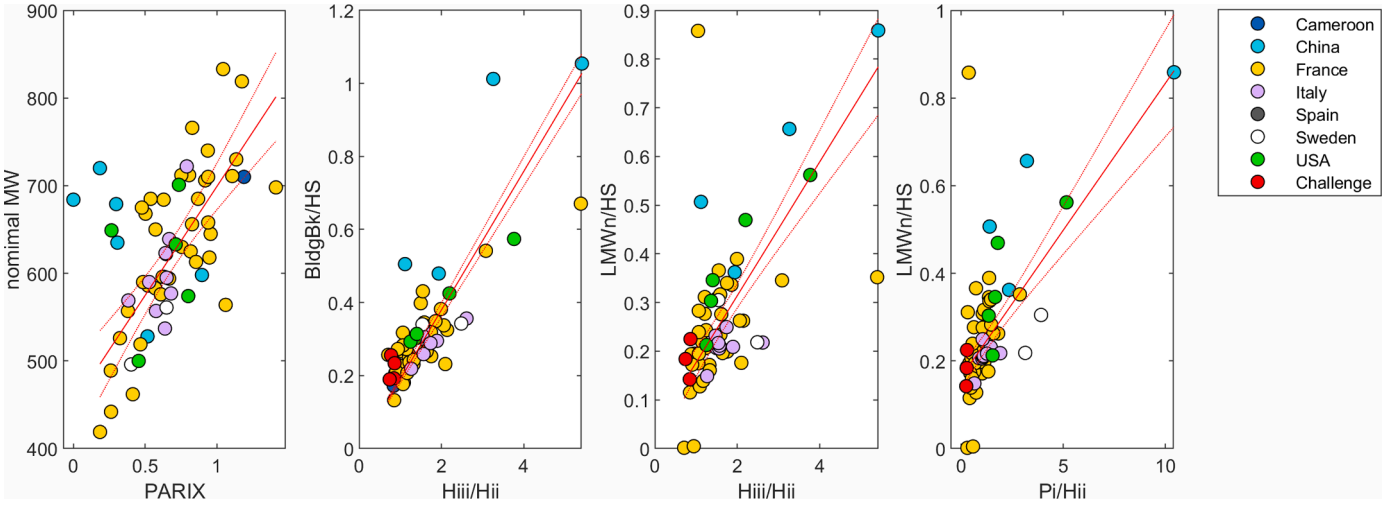

Fig. 6. Fluorescence ratios compared to LC-OCD composition in the global dataset. Red lines indicate robust regression models. Note that samples from China (light blue) were measured using a different LC-OCD analyzer compared to all other samples.

neutral fractions.

These relationships help to explain the differences in treatability observed for the Seine and Bretagne challenge samples. Despite similar initial DOC, the Seine sample was most susceptible to PAC treatment whereas the Bretagne sample was most susceptible to coagulation (Fig. 5). In their PARAFAC decompositions (Fig. S3) both were dominated by Component H_i; however, the Seine sample had higher P_i/H_{ii} and P_{ii}/H_{ii} in addition to lower PARIX and P_{ii}/H_{iii} (1.3, 0.3, 0.5 and 0.15 respectively) indicating low susceptibility to coagulation, due to a high proportion of LMW neutrals and building blocks, but only moderate susceptibility to PAC adsorption, due to relatively high H_{iii} compared to P_{ii}. In contrast, the Bretagne sample had lower P_i/H_{ii} and P_{ii}/H_{ii} and higher PARIX and P_{ii}/H_{iii} (0.6, 0.15, 1.4 and 0.2 respectively), indicating high aromaticity and susceptibility to both coagulation and PAC, but less susceptibility to ion exchange.

3.6. The molecular basis of PARIX

Recent studies have attempted to deduce the molecular origin of DOM fluorescence by linking fluorescence spectroscopy with ultrahigh-resolution mass spectrometry. The usual approach is to obtain a collection of samples, subsampled and measured by each of the two analytical techniques, and search for post-hoc correlations between the two datasets (Herzsprung et al., 2012). Although the specific molecular structures responsible for specific fluorescence components can't be identified this way, this approach has been used many times to indicate the relative proportions of carbon, hydrogen and oxygen in the ionizable molecular formulas that covary with PARAFAC components. In two prior studies that both identified components similar to H_{ii} and H_{iii} and mapped post-hoc correlations in van Krevlan space, each was associated with a broad range of molecular formulae properties with no apparent differences between the two (Kellerman et al., 2015; Wünsch et al., 2018a). However, Wünsch and colleagues also tested a different approach whereby the fluorescence and mass spectrometry datasets were simultaneously decomposed using a coupled tensor factorization model. In that analysis, H_{ii} and H_{iii} mapped in distinct regions of van Krevlen space, with H_{iii} (H/C=1.2-1.4, O/C=0.5-0.6) more saturated and less aromatic than H_{ii} (H/C=0.9-1.3, O/C>0.6), which is consistent with the interpretation of PARIX in this study. The same analysis indicated that H_i encompassed nearly the full range of molecular formulae correlated with H_{iii}, as well as additional formulae of more aliphatic character. This is consistent with FIX and PARIX each representing the ratio between a condensed, aromatic humic fraction compared to a relatively aliphatic humic fraction. Thus, for FIX the aliphatic fraction is H_i while for PARIX, the aliphatic fraction is H_{iii}.

3.7. Applications in water treatment

Fluorescence ratios will be useful for monitoring DOC removal in water treatment so long as they sensitively track small changes in water quality. This study builds upon earlier studies indicating that a large number of treatment options cause measurable changes in fluorescence ratios (e.g. Fig. 3), by demonstrating that some changes are predictable in size and direction across diverse surface waters. Biological filters are a widely-used treatment option that was not tested in this study, but earlier studies offer insights that can be used as the basis of predictions. Biodegradation predominantly removes hydrophilic DOM so is generally associated with an increase in SUVA consistent with the relative enrichment of aliphatic carbon fractions (Hansen et al., 2016; Moran et al., 2000). A recent study showed that biodegradation efficiently removed H_{iv} and P_{ii} but not H_i, H_{ii} or H_{iii} (Moona et al., 2021). This explains why FIX for surface water samples is relatively unchanged by biodegradation (Fellman et al., 2008; Hansen et al., 2016) and suggests that the same will be true of PARIX.

PARIX and SUVA appear to have rather similar capacities for predicting drinking water treatability, but PARIX has the advantage of being measurable on a single, affordable instrument with high precision, which is especially useful when DOC concentrations are low and difficult to accurately quantify. As with all optical indices, there will be limits to usefulness of PARIX for tracking changes in water composition or predicting reactivity and treatability (Hansen et al., 2016; Helms et al., 2008; Korak et al., 2014; Korak et al., 2015; Weishaar et al., 2003). PARIX is likely to overestimate SUVA and susceptibility to coagulation in samples with relatively low aromaticity if this was caused by the production of small DOM fractions that fluoresce at wavelengths shorter than 400 nm or not at all, and that were produced more efficiently than H_{iii}. Conversely, PARIX will underestimate SUVA in samples with relatively high aromaticity if this was due to the efficient removal of LMW fractions not covarying with H_{iii}. Ratios involving H_{iv} or protein-like components are likely to be better predictors of bioavailable DOC than PARIX (Fellman et al., 2008; Moona et al., 2021).

Despite apparently widespread relevance for predicting the composition of surface waters at varying stages of treatment, PARIX has not been tested for other sample types, e.g. groundwater, wastewater, and seawater. In iron-rich lakes, light absorption by iron causes overestimated SUVA values but would likely have the opposite effect on PARIX due to greater quenching of long-wavelength fluorescence (Poulin et al., 2014), while in wastewaters, dyes and other fluorescent

contaminants often are present (Carstea, 2012; Cohen et al., 2014; Goffin et al., 2018), some of which (like W_{ii}) may interfere with the measurement of H_{ii} or H_{iii}. Future work should determine the utility of PARIX and other PARAFAC ratios for tracking physical and biogeochemical processes across the land/sea continuum and in different types of engineered environment.

4. Conclusions

This study found that PARAFAC fractions with specific spectral and chemical properties are ubiquitous in surface waters and tend to behave predictably during drinking water treatment. A new fluorescence proxy "PARIX" was proposed that tracks the ratio of condensed aromatic humic acids containing abundant carboxyl and phenolate groups (H_{ii}) to aquagenic refractory organic matter (H_{iii}) including weathered and deaggregated polyphenolics and polyaromatic acids with relatively high nitrogen content. In addition to predicting SUVA, PARIX is correlated to indices used to predict DOM freshness and extent of humification.

More data are also needed to confirm the treatability results in this study and determine whether PARIX and other ratios (especially P_i/H_{ii} and P_{ii}/H_{iii}) are consistent predictors of DOC removal in treatments similar to those examined in this study, and how they relate to other types of treatments that were not studied.

Indirectly, this study provides further support for the PARAFAC approach to interpreting steady state DOM fluorescence, which relies upon the assumption that different DOM fractions produce linearly additive fluorescence signals as a consequence of mathematical and chemical independence (Bro, 1997; McKay, 2020).

Synopsis

Fluorescence composition predicts SUVA and DOC removal during drinking water treatment.

Author contributions

MP and KM are equal first authors on this study, with MP responsible for study design, logistics, planning and coordination, experimental design, and implementing experiments, while KM was responsible for statistical analyses, data visualisation and article writing. Supporting roles were: Study design: FZ; analytical measurements: LM & QY; database: MP & SL; data visualisation: MP; article writing: MP; internal scientific review: EF. All authors approved the submitted article.

Declaration of competing interest

The authors declare that they have no known competingfinancial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.118592.

References

Ahmad, S.R., Reynolds, D.M., 1999. Monitoring of water quality using fluorescence technique: Prospect of on-line process control. Wat Res 33 (9), 2069–2074.

Aiken, G.R. (2014) Aquatic organic matter fluorescence. Coble, P., Baker, A., Lead, J., Reynolds, D. and Spencer, R. (eds), pp. 35-74, Cambridge University Press, New York.

Amy, G., 2008. Fundamental understanding of organic matter fouling of membranes. Desalination 231 (1), 44–51.

Archer, A.D., Singer, P.C., 2006. An evaluation of the relationship between SUVA and NOM coagulation using the ICR database. Journal AWWA 98 (7), 110–123.

Asmala, E., Haraguchi, L., Markager, S., Massicotte, P., Riemann, B., Staehr, P.A., Carstensen, J., 2018. Eutrophication leads to accumulation of recalcitrant autochthonous organic matter in coastal environment. Global Biogeochemical Cycles 32, 1673–1687. https://doi.org/10.1029/2017GB005848.

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. Wat Res 45 (2), 797–809.

Bhattacharya, R., Osburn, C.L., 2020. Spatial patterns in dissolved organic matter composition controlled by watershed characteristics in a coastal river network: The Neuse River Basin, USA. Wat Res 169, 115248.

Bro, R., 1997. PARAFAC: Tutorial and applications. Chemom Intell Lab 38, 149–171.

Cardenas, CS, Gerea, M, Garcia, PE, Perez, GL, Dieguez, MC, Rapacioli, R, Reissig, M, Queimali, C, 2017. Interplay between climate and hydrogeomorphic features and their effect on the seasonal variation of dissolved organic matter in shallow temperate lakes of the Southern Andes (Patagonia, Argentina): a field study based on optical properties. Ecohydrology 10, e1872. https://doi.org/10.1002/eco.1872.

Carstea, E.M., Balkis, N., 2012. Water Pollution. InTech, Rijeka, Croatia, pp. 47–68.

Cawley, K.M., Butler, K.D., Aiken, G.R., Larsen, L.G., Huntington, T.G., McKnight, D.M., 2012. Identifying fluorescent pulp mill effluent in the Gulf of Maine and its watershed. Mar. Poll. Bull. 64 (8), 1678–1687.

Chen, M.L., Price, R.M., Yamashita, Y., Jaffe, R., 2010. Comparative study of dissolved organic matter from groundwater and surface water in the Florida coastal Everglades using multi-dimensional spectrofluorometry combined with multivariate statistics. Appl. Geochem. 25 (6), 872–880.

Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Mar. Chem. 51, 325–346.

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

Cohen, E., Levy, G.J., Borisover, M., 2014. Fluorescent components of organic matter in wastewater: Efficacy and selectivity of the water treatment. Wat Res 55, 323–334.

Croué, J.P., Lefebvre, E., Martin, B., Legube, B., 1993. Removal of Dissolved Hydrophobic and Hydrophilic Organic Substances during Coagulation/Flocculation of Surface Waters. Wat Sci Technol 27 (11), 143–152.

Cuss, C.W., Guéguen, C., 2015. Relationships between molecular weight and fluorescence properties for size-fractionated dissolved organic matter from fresh and aged sources. Wat Res 68, 487–497.

Edwards, M., 1997. Predicting DOC removal during enhanced coagulation. Journal AWWA 89 (5), 78–89.

Edzwald, J.K., 1993. Coagulation in drinking water treatment: Particles, organics and coagulants. Wat Sci Technol 27 (11), 21–35.

Edzwald, J.K., Becker, W.C., Wattier, K.L., 1985. Surrogate Parameters for Monitoring Organic Matter and THM Precursors. Journal AWWA 77 (4), 122–132.

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.

Garcia, R.D., Diéguez, M.C., Gerea, M., Garcia, P.E., Reissig, M., 2018. Characterisation and reactivity continuum of dissolved organic matter in forested headwater catchments of Andean Patagonia. Freshwater Biology 63 (9), 1049–1062. https://doi.org/10.1111/fwb.13114.

Goffin, A., Guérin, S., Rocher, V., Varrault, G., 2018. Towards a better control of the wastewater treatment process: excitation-emission matrix fluorescence spectroscopy of dissolved organic matter as a predictive tool of soluble BOD5 in influents of six Parisian wastewater treatment plants. Environmental Science and Pollution Research 25 (9), 8765–8776.

Hansen, A.M., Kraus, T.E.C., Pellerin, B.A., Fleck, J.A., Downing, B.D., Bergamaschi, B.A., 2016. Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. Limnol. Oceanogr. 61 (3), 1015–1032.

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. Wat Res 125 (Supplement C), 1–10.

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnol. Oceanogr. 53 (3), 955–969.

Herzsprung, P., von Tümpling, W., Hertkorn, N., Harir, M., Büttner, O., Bravidor, J., Friese, K., Schmitt-Kopplin, P., 2012. Variations of DOM Quality in Inflows of a Drinking Water Reservoir: Linking of van Krevelen Diagrams with EEMF Spectra by Rank Correlation. Environ. Sci. Technol. 46 (10), 5511–5518.

Holland, P.W., Welsch, R.E., 1977. Robust regression using iteratively reweighted least-squares. Communications in Statistics - Theory and Methods 6 (9), 813–827.

Hua, G., Reckhow, D.A., Abusallout, I., 2015. Correlation between SUVA and DBP formation during chlorination and chloramination of NOM fractions from different sources. Chemosphere 130, 82–89.

Huang, X., Yan, C., Nie, M., Chen, J., Ding, M., 2022. Effect of colloidal fluorescence properties on the complexation of chloramphenicol and carbamazepine to the natural aquatic colloids. Chemosphere 286, 131604.

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). Wat Res 45 (2), 879–885.

Hutchins, R.H.S., Aukes, P., Schiff, S.L., Dittmar, T., Prairie, Y.T., del Giorgio, P.A., 2017. The Optical, Chemical, and Molecular Dissolved Organic Matter Succession Along a Boreal Soil-Stream-River Continuum. J. Geophys. Res.-Biogeosci. 122 (11), 2892–2908.

Ishii, S.K.L., Boyer, T.H., 2012. Behavior of reoccurring PARAFAC components in fluorescent dissolved organic matter in natural and engineered systems: A critical review. Environ. Sci. Technol. 46 (4), 2006–2017.

Kellerman, A.M., 2015. Molecular dissolved organic matter composition in lakes across Sweden as relative intensities of FT-ICR-MS peaks. PANGAEA.
Kellerman, A.M., Kothawala, D.N., Dittmar, T., Tranvik, L.J., 2015. Persistence of dissolved organic matter in lakes related to its molecular characteristics. Nature Geosci 8 (6), 454–457.
Kennedy, M.D., Chun, H.K., Quintanilla Yangali, V.A., Heijman, B.G.J., Schippers, J.C., 2005. Natural organic matter (NOM) fouling of ultrafiltration membranes: fractionation of NOM in surface water and characterisation by LC-OCD. Desalination 178 (1), 73–83.
Korak, J.A., Dotson, A.D., Summers, R.S., Rosario-Ortiz, F.L., 2014. Critical analysis of commonly used fluorescence metrics to characterize dissolved organic matter. Wat Res 49, 327–338.
Korak, J.A., Rosario-Ortiz, F.L., Scott Summers, R., 2015. Evaluation of optical surrogates for the characterization of DOM removal by coagulation. Environ. Sci. Water Res. Technol. 1 (4), 493–506.
Kothawala, D., Murphy, K., Stedmon, C., Weyhenmeyer, G., Tranvik, L., 2013. Inner filter correction of dissolved organic matter fluorescence. Limnol Oceanogr Meth 11, 616–630.
Kothawala, D.N., Stedmon, C.A., Müller, R.A., Weyhenmeyer, G.A., Köhler, S.J., Tranvik, L.J., 2014. Controls of dissolved organic matter quality: evidence from a large-scale boreal lake survey. Global Change Biol. 20 (4), 1101–1114.
Kowalczuk, P., Cooper, W.J., Durako, M.J., Kahn, A.E., Gonsior, M., Young, H., 2010. Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Relationships between fluorescence and its components, absorption coefficients and organic carbon concentrations. Mar. Chem. 118 (1–2), 22–36.
Kowalczuk, P., Durako, M.J., Young, H., Kahn, A.E., Cooper, W.J., Gonsior, M., 2009. Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Interannual variability. Mar. Chem. 113 (3–4), 182–196.
Krasner, S.W., Croué, J.-P., Buffle, J., Perdue, E.M., 1996. Three approaches for characterizing NOM. Journal AWWA 88 (6), 66–79.
Kulkarni, H.V., Mladenov, N., Johannesson, K.H., Datta, S., 2017. Contrasting dissolved organic matter quality in groundwater in Holocene and Pleistocene aquifers and implications for influencing arsenic mobility. Applied Geochemistry 77, 194–205. https://doi.org/10.1016/j.apgeochem.2016.06.002.
Lee, B.-M., Hur, J., 2016. Adsorption Behavior of Extracellular Polymeric Substances on Graphene Materials Explored by Fluorescence Spectroscopy and Two-Dimensional Fourier Transform Infrared Correlation Spectroscopy. Environ. Sci. Technol. 50 (14), 7364–7372.
Leenheer, J.A., Croué, J.-P., 2003. Peer Reviewed: Characterizing Aquatic Dissolved Organic Matter. Environ. Sci. Technol. 37 (1), 18A–26A.
Leenheer, J.A., Rostad, C.E., Barber, L.B., Schroeder, R.A., Anders, R., Davisson, M.L., 2001. Mature and chlorine reactivity of organic constituents from reclaimed water in groundwater, Los Angeles County, California. Environ. Sci. Technol. 35 (19), 3869–3876.
Ly, Q.V., Kim, H.-C., Hur, J., 2018. Tracking fluorescent dissolved organic matter in hybrid ultrafiltration systems with TiO2/UV oxidation via EEM-PARAFAC. J. Membr. Sci. 549, 275–282.
Massicotte, P., Asmala, E., Stedmon, C., Markager, S., 2017. Global distribution of dissolved organic matter along the aquatic continuum: Across rivers, lakes and oceans. Sci Tot Env 609, 180–191.
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.
McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. Limnol. Oceanogr. 46 (1), 38–48.
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.
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. Environ. Sci. Water Res. Technol. 4 (4), 529–538.
Mopper, K., Qian, J., Meyers, R.A., Miller, M.P., 2006. Encyclopedia of Analytical Chemistry.
Moran, M., Sheldon, W.M.J., Zepp, R.G., 2000. Carbon loss and optical property changes during long-term photochemical and biological degradation of estuarine dissolved organic matter. Limnol. Oceanogr. 45 (6), 1254–1264.
Moran, M.A., Kujawinski, E.B., Stubbins, A., Fatland, R., Aluwihare, L.I., Buchan, A., Crump, B.C., Dorrestein, P.C., Dyhrman, S.T., Hess, N.J., Howe, B., Longnecker, K., Medeiros, P.M., Niggemann, J., Obernosterer, I., Repeta, D.J., Waldbauer, J.R., 2016. Deciphering ocean carbon in a changing world. Proc. Natl Acad. Sci. 113 (12), 3143–3151.
Moran, M.A., Pomeroy, L.R., Sheppard, E.S., Atkinson, L.P., Hodson, R.E., 1991. Distribution of terrestrially derived dissolved organic-matter on the southeastern United-States continental-shelf. Limnol. Oceanogr. 36 (6), 1134–1149.
Murphy, K.R., Bro, R., Stedmon, C.A., Coble, P., Baker, A., Lead, J., Reynolds, D., Spencer, R., 2014a. Aquatic organic matter fluorescence. Cambridge University Press, New York, pp. 339–376.
Murphy, K.R., Hambly, A., Singh, S., Henderson, R.K., Baker, A., Stuetz, R., Khan, S.J., 2011. Organic matter fluorescence in municipal water recycling schemes: toward a unified PARAFAC model. Environ. Sci. Technol. 45 (7), 2909–2916.
Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. Anal Meth 5 (23), 6557–6566.
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 Meth 6 (3), 658–661.
Murphy, K.R., Timko, S.A., Gonsior, M., Powers, L., Wünsch, U., Stedmon, C.A., 2018. Photochemistry illuminates ubiquitous organic matter fluorescence spectra. Environ. Sci. Technol. 52 (19), 11243–11250.
Newcombe, G., Morrison, J., Hepplewhite, C., Knappe, D.R.U., 2002. Simultaneous adsorption of MIB and NOM onto activated carbon: II. Competitive effects. Carbon 40 (12), 2147–2156.
Ohno, T., 2002. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. Environ. Sci. Technol. 36, 742–746.
Palma, D., Parlanti, E., Sourzac, M., Voldoire, O., Beauger, A., Sleiman, M., Richard, C., 2021. Fluorescence analysis allows to predict the oxidative capacity of humic quinones in dissolved organic matter: implication for pollutant degradation. Environ. Chem. Lett. 19 (2), 1857–1863.
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. Org. Geochem. 31 (12), 1765–1781.
Potter, B.B., Wimsatt, J.C., 2005. Method 415.3 - Measurement of Total Organic Carbon, Dissolved Organic Carbon and Specific uv Absorbance at 254 nm in Source Water and Drinking Water. U.S. Environmental Protection Agency, Washington, DC.
Poulin, B.A., Ryan, J.N., Aiken, G.R., 2014. Effects of iron on optical properties of dissolved organic matter. Environ. Sci. Technol. 48 (17), 10098–10106.
Sarathy, S.R., Mohseni, M., 2007. The impact of UV/H2O2 advanced oxidation on molecular size distribution of chromophoric natural organic matter. Environ. Sci. Technol. 41 (24), 8315–8320.
Shimabuku, K.K., Kennedy, A.M., Mulhern, R.E., Summers, R.S., 2017. Evaluating Activated Carbon Adsorption of Dissolved Organic Matter and Micropollutants Using Fluorescence Spectroscopy. Environ. Sci. Technol. 51 (5), 2676–2684.
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. Wat Res 54, 159–169.
Stedmon, C.A., Markager, S., 2005a. Resolving the variability of dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. Limnol. Oceanogr. 50, 686–697.
Stedmon, C.A., Markager, S., 2005b. Tracing the production and degradation of autochthonous fractions of dissolved organic matter using fluorescence analysis. Limnol. Oceanogr. 50 (5), 1415–1426.
Stubbins, A., Lapierre, J.F., Berggren, M., Prairie, Y.T., Dittmar, T., del Giorgio, P.A., 2014. What's in an EEM? Molecular Signatures Associated with Dissolved Organic Fluorescence in Boreal Canada. Environ. Sci. Technol. 48 (18), 10598–10606.
Van Geluwe, S., Braeken, L., Van der Bruggen, B., 2011. Ozone oxidation for the alleviation of membrane fouling by natural organic matter: A review. Wat Res 45 (12), 3551–3570.
Vines, M., Terry, L.G., 2020. Evaluation of the biodegradability of fluorescent dissolved organic matter via biological filtration. AWWA Water Science 2 (5), e1201.
Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Fram, M.S., Fujii, R., Mopper, K., 2003. Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. Environ. Sci. Technol. 37 (20), 4702–4708.
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. Sci Tot Env 820, 153070.
Wünsch, U.J., Acar, E., Koch, B.P., Murphy, K.R., Schmitt-Kopplin, P., Stedmon, C.A., 2018a. The Molecular Fingerprint of Fluorescent Natural Organic Matter Offers Insight into Biogeochemical Sources and Diagenetic State. Anal. Chem. 90 (24), 14188–14197.
Wünsch, U.J., Bro, R., Stedmon, C.A., Wenig, P., Murphy, K.R., 2019. Emerging patterns in the global distribution of dissolved organic matter fluorescence. Anal Meth 11 (7).
Wünsch, U.J., Geuer, J.K., Lechtenfeld, O.J., Koch, B.P., Murphy, K.R., Stedmon, C.A., 2018b. Quantifying the impact of solid-phase extraction on chromophoric dissolved organic matter composition. Mar Chem 207, 33–41.
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. Environ. Sci. Technol. 51 (20), 8.
Yamashita, Y., Scinto, L.J., Maie, N., Jaffé, R., 2010. Dissolved Organic Matter Characteristics Across a Subtropical Wetland's Landscape: Application of Optical Properties in the Assessment of Environmental Dynamics. Ecosystems 13 (7), 1006–1019.
Yang, L., Cheng, Q., Zhuang, W.-E., Wang, H., Chen, W., 2019. Seasonal changes in the chemical composition and reactivity of dissolved organic matter at the land-ocean interface of a subtropical river. Environmental Science and Pollution Research 26 (24), 24595–24608.
Zito, P., Podgorski, D.C., Johnson, J., Chen, H., Rodgers, R.P., Guillemette, F., Kellerman, A.M., Spencer, R.G.M., Tarr, M.A., 2019. Molecular-Level Composition and Acute Toxicity of Photosolubilized Petrogenic Carbon. Environ. Sci. Technol. 53 (14), 8235–8243.