Fluorescent Dissolved Organic Matter Constituents as Surrogates for Disinfection Byproduct Formation in Drinking Water Treatment: A Critical Review

Elena Fernandez-Pascual, ¥, †,‡ Boris Droz,* ¥, †,‡ Jean O'Dwyer, †,‡,§ Connie O'Driscoll, ¶

Emma H. Goslan, ⊥ Simon Harrison, †,‡ John Weatherill* †,‡,§

† School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland.

‡ Water and Environment Research Group, Environmental Research Institute, Lee Road, University College Cork, Cork, Ireland.

§ Irish Centre for Research in Applied Geosciences, University College Dublin, Dublin, Ireland

¶ Ryan Hanley Ltd., Castlebar, Co. Mayo, Ireland.

⊥ Cranfield Water Science Institute, Cranfield University, Cranfield Bedfordshire, UK.

¥ These authors contributed equally to the publication.
Disinfection byproduct (DBP) formation, prediction and minimization are a crucial issue for the domestic water supply industry, which needs to provide quality and safe drinking water to consumers. Fluorescence excitation–emission matrices-parallel factor analysis (EEM-PARAFAC), is used to characterize and quantify fluorescent dissolved organic matter (DOM) in aquatic systems. EEM-PARAFAC has been identified as a potential method to predict DBP formation in treated waters. However, the method ability for specific DBP classes or species prediction is uncertain. This critical review evaluates the published literature describing empirical relationships between DOM fluorophores identified by PARAFAC components and DBP formation obtained during water disinfection. From 42 selected peer-reviewed articles, 202 established linear relationships ($R^2 \geq 0.5$) with DBP classes or species were found.

Trihalomethanes (THMs) and haloacetic acids (HAAs), as regulated compounds, were extensively investigated and exhibit a strong relationship. Overall, carbonaceous-DBP classes exhibited strong relationships with humic/fulvic-like components. Conversely, a relationship between nitrogenous-DBP classes and PARAFAC components was less clear, but it was shown to be preferential to protein-like PARAFAC components in the case of algae/bacterial DOM sources. This review highlights the challenges of transposing site-specific or DOM source-specific empirical relationship between PARAFAC component and DBPs formation potential to a global model.
The use of chlorine and other disinfection methods such as ozone or chloramine treatment, for drinking water disinfection can lead to the formation of potentially carcinogenic disinfection byproducts (DBPs), through their reactions with dissolved organic matter (DOM) present in the raw water. Consequently, eleven DBPs, including trichloromethane and tribromomethane, are subject to regulation by authorities in the EU and US. However, more than 700 DBPs have been identified to date, the vast majority of which are unregulated and many of which are potentially carcinogenic. Moreover, a recent study estimated that between 32–81% of the total halogen component content constitutes unidentified DBPs.

DOM in freshwaters is comprised of many different organic carbon compounds, derived from autochthonous sources, such as algae/bacteria, and allochthonous sources, such as leaf/soil, material, whose hydrological export varies spatially and temporally within river basins around the world. Humic and fulvic acids from allochthonous sources, comprising high molecular weight and aromaticity humic substances are considered to be the primary DOM precursors for carbonaceous DBPs (C-DBPs). In contrast, autochthonous sources such as DOM derived from algae bloom and allochthonous nutrient enrichment from bacteria, as well as organic matter in wastewater, are thought to be a significant precursor sources for nitrogenous DBPs (N-DBPs), which may be potentially more carcinogenic to human health than C-DBPs. A global increase in DOM export from freshwater systems is forecast over the coming decades as a result of climate change and agricultural intensification. Increasing global DOM export and concentration in drinking water sources, particularly where these are derived from surface waters, will present significant challenges for safe and sustainable drinking water production.
In the last two decades, technological advances in fluorescence excitation–emission matrix (EEM) spectroscopy have offered a unique perspective on DOM characterization and quantification in freshwater environments. Fluorescence EEM spectroscopy is a low-cost, non-destructive, sensitive and selective technique that can provide critical information on the sources, optical nature, structures and molecular properties of complex DOM admixtures. The technique also offers considerable promise for characterization of DOM fluorophores involved in DBP formation. Various methods have been developed to extract quantitative and qualitative information from EEM spectra such as ‘peak picking’ and fluorescence regional integration (FRI). The peak picking method can quantify complex mixtures of fluorophores by integrating regional fluorescence maximal intensities around predefined wavelengths pairs (‘peaks’), to which a nomenclature has been assigned.

In addition, single-value decomposition methods, such as principal component analysis (PCA) or parallel factor analysis (PARAFAC) have been developed. These analytical approaches decompose EEM spectra into independent components representing groups of fluorophores with similar spectra. The performance of PARAFAC to extract individual components from EEM spectra has been found to be superior to PCA in the absence of unspecific contributions, e.g., noise or light scattering, therefore PARAFAC has been the choice for EEM data analysis. More recently, interest in the application of artificial neural networks (ANN), such as self-organizing maps to analyze EEM spectra has been growing. The ANN tend to support EEM-PARAFAC and open new perspectives to elucidate very complex fluorescence relationship using a large amount of samples.

Scope of the critical review. Several reviews have been published recently on the topic of fluorescence spectroscopy EEM-PARAFAC as a technique to characterize DOM. Starting from a
general overview of the technique and its applications, these articles cover the topic of the application of EEM-PARAFAC in drinking water and wastewater treatment plants, offer a critical analysis of commonly used fluorescence metrics, treat the potential pitfalls of oversimplification on interpreting EEM, define the application of the self-organizing map for data analysis, introduce new approach on similarity metrics and review the practical challenge for the online continuous monitoring. However, there has been no review of the application of PARAFAC components as surrogates for the formation potential of specific DBP classes. This review identifies the literature on existing empirical relationships between DOM fluorophores, identified by PARAFAC components, and their observed DBP formation potential. Focus is made on using these relationships to predict and monitor DBP formation at the drinking water treatment plant to optimize the water treatment process and protect health of consumers.

2. LITERATURE REVIEW

Out of the 375 collected articles, 42 articles were selected, hereafter referred to as selected articles, matching the scope and search criteria of our critical review (procedure fully detailed in the supporting information (SI), Figure S1 & Table S1). A report data framework (SI, Table S2) was created to harmonize the data extraction, e.g., the water sources, DOM content, disinfection method, fluorescence measurement procedure and DBPs formation potential method investigation, from the selected articles for the purpose of this review.

PARAFAC components found in the selected articles were classified and harmonized following the established nomenclature into five fluorophore regions (detailed in SI, Figure 1) and are summarized in Table 1 for the purpose of comparison. Statistically distinct emission values (pairwise Wilcoxon test, p-value <0.01) are reported between the fluorophore regions
except between humic- and fulvic-like components, which can be seen as one group containing

similar chemical structures with different degrees of oxidation, size of molecules and solubility

at low pH. Moreover, according to the mass carbon content, fulvic-compounds are estimated to

be more abundant than humic- and to exhibit higher fluorescence intensity.22

Table 1. PARAFAC Component Classified and Described in the 42 Selected.

<table>
<thead>
<tr>
<th>fluorophore regiona</th>
<th>n</th>
<th>range (min-max) λ maximum (nm)b</th>
<th>peak labela</th>
<th>identified potential environmental sourcesc</th>
</tr>
</thead>
<tbody>
<tr>
<td>humic-like (V)</td>
<td>299 (188)</td>
<td>&lt;250–375 (255–410)</td>
<td>&lt;300–520</td>
<td>terrestrial or river source containing polyhydroxylated aromatics such as those found in lignin, as well as phenols, hydroquinones and indoles. Generally insoluble fraction at neutral pH.</td>
</tr>
<tr>
<td>fulvic-like (III)</td>
<td>41 (33)</td>
<td>&lt;250–265 (238–360)</td>
<td>400–475</td>
<td>terrestrial and marine source of aromatic species, such quinones or other oxidized aromatics. Generally, smaller molecule, more polar and soluble than humic-like.</td>
</tr>
<tr>
<td>microbial humic-like (V)</td>
<td>55 (33)</td>
<td>&lt;250–346 (285–374)</td>
<td>&lt;300–459</td>
<td>originally identified in marine environments, associated with biodegradation of humic-like or with specific proteins or metabolic byproducts.</td>
</tr>
<tr>
<td>tyrosine-like (IV)</td>
<td>103 (49)</td>
<td>&lt;250–280 (270–285)</td>
<td>&lt;300–328</td>
<td>free or bound amino acids associated to microbial activity, autochthone source.</td>
</tr>
<tr>
<td>tryptophan-like (IV)</td>
<td>135 (58)</td>
<td>&lt;250–300 (255–300)</td>
<td>&lt;300–390</td>
<td>soluble protein or by-product associated to microbial activity, autochthone source.</td>
</tr>
</tbody>
</table>

aFollows the traditional assignment and peak label made elsewhere,22, 23, 41, 42 where Roman Numerals follow the spectra region given in Figure 1. Secondary excitation maxima are shown in parenthesis. n stands for the sum component described in the selected articles used to calculate the range of the first maxima and the secondary maxima in parenthesis. bRange of excitation and emission wavelengths were not considered below 240 and 300 nm respectively due to potential deteriorating signal to noise ratios. cPotential sources are described elsewhere.22, 23, 43-49
Figure 1. Classical emission–excitation matrix (EEM) of dissolved organic matter (Bunsheelin river, Ireland, the 4th July 2021). White dash lines delimit the five fluorophore regions defined in Table 1. Grey inserts refer to wavelengths at which fluorescence maximal intensities are “picked”.

DBP classes and their formation potentials have been previously classified following established nomenclature as follows: i) carbonaceous DBPs (C-DBPs), i.e., trihalomethanes (THMs), haloacetic acids (HAAs), haloketones (HKs), haloacetaldehydes (HALs), halogenated furanones (X-furanones), iodinated THMs (I-THMs) and ii) nitrogenous DBPs (N-DBPs), i.e., halonitromethanes (HNMs), haloacetonitriles (HANs), haloacetamides (HAMs), N-nitrosamines (NAs), cyanide (CNX). In total, 41 individual DBP species (4 THMs, 9 HAAs, 2 HKs, 1 HALs,
X-furanone, 6 I-THMs, 1 non-classified C-DBPs, 1 HNMs, 4 HANs, 2 HAMs, 9 NAs and 1 CNX) were subject to investigation in this review.

The 42 selected articles were published from 2009 to 2022 (Figure 2) and describe the global application of PARAFAC in DBP formation. Selected articles focus mainly on the conventional drinking water treatment plant optimization or upgrade (48% of the articles), tracking the spatiotemporal dynamics of DOM in surface water (14%), evaluating the problematics of biofilm algae (17%) or species-specific leaf leachate (9%) as a potential source of DOM producing harmful DBPs. However, some articles investigate the photo-irradiation impact on DOM (9%) and comparing method of DOM characterization (3%), e.g., size exclusion chromatography (SEC) and two-stage differential characterization of water matrix and DOM during chlorination.

A large diversity of DOM sources, e.g., surface water (71%), algae/bacterial DOM (24%) and leaf leachate (10%), were investigated (summarized in the SI), with a DOM concentration indicated in 98% of the selected articles ranging from 0.03 to 1,000 mg C L$^{-1}$ with a mean (and standard deviation) of 22 (± 123) mg C L$^{-1}$.

Figure 2. A) Cumulative number of articles from 2009 to 2022 that have been included in this critical review (n$_{\text{articles}}$ = 42). The number of articles published each year is indicated above the bars. B) Country of origin and number of selected articles published by each country.
3. EXPERIMENTAL AND MODES OF OPERATION CONSIDERATIONS

**Excitation–emission matrix (EEM) spectra and PARAFAC**

Sample preparations methods prior to EEM spectra acquisition reported in the selected papers were typically filtration, dilution and/or pH adjustment of water samples. Dilution, up to a specific ultraviolet (UV) absorbance ($\text{UV}_{254}\text{ cm}^{-1} < 0.05$), may be used to avoid the inner-filter effect, i.e., absorbance signal saturation at high DOM concentrations.50 Two differing strategies were adopted concerning pH: the first was to acquire EEM spectra at native pH and the second strategy was to reduce the pH. Reducing pH of samples has been shown to promote the dissociation of DOM-metal complexes and thus minimize the quenching of fluorescence.51, 52 However, acidic (–COOH, –OH) and basic (–NH$_2$) DOM functional groups are very sensitive to pH.53 Lowering the pH has been demonstrated to significatively alter EEM spectra,54, 55 therefore pH adjustment should be used with caution and potential EEM spectra alteration should be properly evaluated.

EEM spectra are typically recorded for excitation and emission wavelengths higher than 240 and 300 nm, respectively, where shorter wavelengths are potentially associated with alteration of the signal to noise ratios for most commercial fluorimeters.35 The pre-processing of EEMs spectra, i.e., blank subtraction, smoothing data, inner-filter effect, aims to i) correct bias, ii) remove scatter lines and iii) normalize spectra within a data set. Commonly used PARAFAC toolboxes were DOMFluor,$^{56}$ and drEEM,$^{31}$ present in 62% and 17% of the selected articles, respectively. However, a multitude of toolboxes exist where 7 toolboxes have been found in the selected articles (detailed in SI). Data workflows using different toolboxes but similar parameter during the pre-processing and PARAFAC analysis have been demonstrated to be nearly identical
for several datasets but analyses made under a unique workflow are more reliable. Although the methods for pre-processing and PARAFAC are well established, there is considerable variation apparent across the studies reviewed (detailed in SI), making it challenging when comparing data. Altogether, this highlights the need of sharing raw EEM spectra as supplementary material of a research paper to make accessible and reusable the data for further collaborative and comparison studies. In addition, several useful best practice aspects were reviewed and are fully described in the SI to guide effort on EEM-PARAFAC for future studies.

**Disinfection byproducts (DBPs) formation potential**

Two strategies were reported in the selected papers to evaluate the DBPs formation potential. The most widely adopted strategy, used in 83% of the 42 selected articles, was to collect samples at several stages of the disinfection treatment chain and investigate the DBP formation potential under controlled laboratory conditions. The second strategy, used in 17% of the selected articles, was to collect samples directly at the drinking water treatment plant. This strategy, whilst less costly in laboratory work, and allowing for large-scale field survey investigations, may have some limitations on data interpretation as the DBP formation potential and the EEM spectra may not originate from the same point in time. DBP formation potentials were investigated for chlorination on 81% of the selected articles, therefore our review is mainly focused on chlorination.

Two standard DBP formation potential laboratory methods exist which may be employed depending on the goal addressed by the study. The Standard Methods for Examination of Water and Wastewater number 5710-B used high free chlorine residual (3–5 mg Cl₂ L⁻¹) at the end of an extended incubation period (7 days) at temperature (25 ± 2°C) and pH (7.0 ± 0.2) controlled.
condition to estimate the maximum concentration of DBPs for a given raw water sample. Due to the high free chlorine dose and longtime exposure, this method is not best suited to simulate DBP concentration in tap water and is more commonly used to investigate the effectiveness of water-treatment process options. Conversely, the method called uniform formation condition (UFC) using low free chlorine residual (1.0 ± 0.4 mg Cl₂ L⁻¹) under a quick incubation period (24 hours) at temperature (20 ± 1°C) and pH (8.0 ± 0.2) controlled conditions, offers greater representativity of the average distribution system conditions, i.e., 1–3 days before distribution and low chlorine residual (~0.2 mg Cl₂ L⁻¹).

Several variables influence DBP concentration and formation pathway, such as pH, temperature, incubation time and disinfection dose. pH governs the speciation of reactive chlorine (pKa, HOCℓ/−OCl, 25°C = 7.54) and consequently the kinetics of the reaction toward a specific DOM functional group. Longer incubation periods generally increase the DBPs concentration. As some DBPs might be unstable in aqueous solution, e.g., HAAs, HANs, and HALs, the use of an appropriate quenching agent is needed to preserve DBP concentrations and make accurate DBP measurement. Ascorbic acid is a widely used quencher and has been demonstrated to be suitable for a large spectrum of halogenated DBPs, e.g., THMs, HAAs, HKs, HALs and HANs (SI, Table S3).

4. ESTABLISHED RELATIONSHIPS BETWEEN PARAFAC COMPONENTS AND DBP FORMATION POTENTIAL

From the 42 selected articles, 202 empirical linear relationships between PARAFAC components and DBP formation potential (Figure 3) were found across the selected articles, 126 with a strong linear relationship (Pearson correlation coefficient ≥0.7), and 76 with a moderate
linear relationship (Pearson correlation coefficient ≥0.5; SI, Table S2). In the selected articles, PARAFAC components offer a superior surrogate for quantification of DBP formation potential in comparison to classical parameters, such as UV absorbance or fluorescence peak picking.\textsuperscript{15, 16, 48, 65-68} However, articles exist where PARAFAC components improve only marginally the prediction of DBP formation potential.\textsuperscript{20, 69-71}

Overall, a larger proportion of relationships between C-DBP classes (Figure 3) and humic/fulvic-like components compare to other components were found in the selected articles as follow: 73%, 71%, 71%, 76% for the THMs, HAAs, HKs and HALs, respectively. In contrast, a similar proportion of relationships between N-DBP classes and humic/fulvic-like versus protein-like, i.e., tyrosine/tryptophan-like, were observed. Comparisons between the selected articles are not straightforward as EEM-PARAFAC as DBPs formation potential were not performed analogously across the selected articles. In addition, caution remains in the transferability of established relationships to predict DBPs formation potential beyond the training dataset used to create them. This is because model evaluation and performance on an independent dataset were never made on the selected articles and Pearson correlation coefficient may not represent a transferable metric to compare two models.\textsuperscript{72} Consequently, occasionally strong relationships might be co-incidental rather than causal,\textsuperscript{73} or established relationship specific for one DOM source only.\textsuperscript{74, 75} Despite this criticism, high-resolution mass spectroscopy (HR-MS) demonstrated that PARAFAC components summarized 39% of the identified DOM species and 59% of the peak intensity including species that are highly unlikely to fluoresce,\textsuperscript{76} which highlights the capabilities of PARAFAC components to characterize DOM precursors and predict DBP formation potential.
Figure 3. Relative percentage of linear relationship ($R^2 \geq 0.5$) described in the 42 selected articles in function of the disinfection byproduct (DBP) class and the PARAFAC component. Number of relationships ($n_{total} = 202$) per DBP classes are expressed in parentheses.

**THM and HAA classes** reported in the selected articles contained 11 regulated DBPs (SI, Table S4) including four THMs, i.e., trichloro- (TCM), bromodichloro- (BDCM), dibromochloro- (DBCM) and tribromomethane (TBM) and five HAAs, i.e., monochloro-, dichloro-, trichloro-, monobromo-, dibromo- acetic acid with an addition of four unregulated HAAs, i.e., bromochloro-, bromodichloro-, dibromochloro-, and tribromo-acetic acid under concern. It is noteworthy that trichloromethane was the dominant DBP compound, with up to 92% of total THMs and up to 47% of the total organic halogen load, observed from chlorination/chloramination in drinking water treatment plants when bromide concentrations were low. According to their major presence, THMs and HAAs were investigated in 90%
and 50% of the selected articles, respectively (SI, Table S5) and accounted together for 52% of the total established linear relationships (Figure 3). Humic/fulvic-like components play a significant role in the THMs and HAAs formation potential and exhibit strong relationships (Table 2). Humic/fulvic-like components, which contain a complex mixture of aromatic and aliphatic hydrocarbon structures with functional groups including amide, carboxyl, hydroxyl and ketone, were postulated within the selected articles, as a major precursor of THMs and HAAs, originating in surface water from fresh plant or leaf litter leachate (ref. in Table 1). This observation is consistent with electrophilic attack on carbonyl functional groups, such as aldehydes, ketones and carboxylic acids, which is thought to be one of the major pathways to produce THMs and HAAs. Conversely, fluorescence measurements support a similar reactivity contribution by all components, as a shift to shorter emission wavelength and decrease in fluorescence intensity are observed in the selected articles. This observation has more impact on the protein-like region than humic/fulvic-like region with higher chlorine doses and longer reaction times. To account for the contribution of all PARAFAC components, multivariable PARAFAC component indices, e.g., sum of all components, humic-like divided by tryptophan-like, etc. have been established and show substantial relationship improvement in comparison to a single PARAFAC component model. This confirms that several compounds with different fluorophore regions contribute as a part of DBPs precursor.

**Brominated (Br-DBPs) and iodinated DBPs (I-DBPs)** are formed when Br\(^-\) and I\(^-\) are present in the water. These species react quickly with hypochlorous acid (HOCl) to form hypobromous (HOBr) or hypoiodous acid (HOI) which might further react with DOM under the same pathway as HOCl. However, HOBr reacts typically up to three orders of magnitude quicker than
HOCl and has a very high reactivity to phenol groups.\textsuperscript{87} Therefore, the incorporation yield of Br into THMs is around 50\% compared to 5–10\% for Cl.\textsuperscript{89} The proportion to form Br-THMs and Br-HAAs species under environmental conditions may be estimated by the ratio between Br\textsuperscript{−} and dissolved organic carbon (DOC) content molar ratios ([Br\textsuperscript{−}]/[DOC]).\textsuperscript{90} Interestingly, the ratio between the variation of bromide and PARAFAC components (\(\Delta\text{Br}/\Delta\text{CPARAFAC}\)) before and after chlorination exhibits a strong relationship with Br-DBP formation potential, e.g., Br-THMs, Br-HAAs and Br-HANs. Conversely, a single PARAFAC component model does not express any significant relationship.\textsuperscript{91} Similar to Br-DBP species, multivariable PARAFAC components indices express a good relationship with I-DBPs formation potential.\textsuperscript{74}

Other C-DBPs like HKs and HALs were intensively investigated in 33\% and 19\% of the selected articles, respectively (SI, Table S5). Similar to THMs and HAAs, they exhibited high number of moderate and strong relationships with humic/fulvic-like components, with 71\% and 76\% of the total relationships per DBP class, respectively (Figure 3). PARAFAC components were particularly suitable as a surrogate for these DBPs as 75\% and 63\% of the investigated relationships exhibited moderate or strong relationships (Table 2 & SI, Table S5) for HKs and HALs, respectively. Overall, other C-DBPs share the same relationship as THMs and HAAs described above, where some relationship could be explained for the HKs\textsuperscript{92} and the HALs\textsuperscript{93} which are reaction intermediates for the formation of THMs.
Table 2. Established Relationships between Formation Potential and PARAFAC Components per Disinfection Method and Disinfection Byproduct (DBP) Classes.

<table>
<thead>
<tr>
<th>DBP species formation potential&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PARAFAC components&lt;sup&gt;b&lt;/sup&gt;</th>
<th>correlation coefficients (R&lt;sup&gt;2&lt;/sup&gt;)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>chlorine – trihalomethanes (THMs)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trichloromethane (TCM)</td>
<td>hum/ful/m-</td>
<td>≥0.71**(5), ≥0.52**(3)/≥0.70**</td>
<td>15, 17, 66,</td>
</tr>
<tr>
<td></td>
<td>hum/tyr/tryp/multi</td>
<td>0.57**/≥0.77*(2)</td>
<td>68, 70, 81,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>94</td>
</tr>
<tr>
<td>dibromochloromethane (DBCM)</td>
<td>tyr/tryp</td>
<td>0.54**/0.50**</td>
<td>95</td>
</tr>
<tr>
<td>TCM, bromodichloromethane (BDCM), DBCM</td>
<td>hum/ful</td>
<td>0.96*/0.92*</td>
<td>96</td>
</tr>
<tr>
<td>TCM, BDCM, DBCM, tribromomethane</td>
<td>hum/ful/m-</td>
<td>≥0.71**(12),</td>
<td>48, 67, 71,</td>
</tr>
<tr>
<td></td>
<td>hum/tyr/tryp/multi</td>
<td>≥0.52*(8)/≥0.70*(4)/0.82*,</td>
<td>75, 80, 81,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.66*/0.95/0.84**, ≥0.52 (5)/0.95</td>
<td>90, 94, 97-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99</td>
</tr>
<tr>
<td><strong>chlorine dioxide – THMs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCM</td>
<td>hum</td>
<td>≥0.92 (4)</td>
<td>65</td>
</tr>
<tr>
<td><strong>ozone – THMs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCM</td>
<td>multi</td>
<td>0.91**</td>
<td>100</td>
</tr>
<tr>
<td>BDCM</td>
<td>multi</td>
<td>0.90**</td>
<td>100</td>
</tr>
<tr>
<td><strong>chlorine – haloacetic acids (HAAs)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dichloroacetic acid (DCAA)</td>
<td>hum/ful/multi</td>
<td>0.58**/≥0.53**(2)/0.61**</td>
<td>17, 101</td>
</tr>
<tr>
<td>trichloroacetic acid (TCAA)</td>
<td>hum/ful/multi</td>
<td>0.73**, 0.52**/0.68**(2)/0.70**</td>
<td>17, 101</td>
</tr>
<tr>
<td>monochloroacetic acid (MCAA), DCAA, TCAA</td>
<td>hum/m-</td>
<td>≥0.73*(4), ≥0.51</td>
<td>66, 99</td>
</tr>
<tr>
<td></td>
<td>hum/tyr/tryp/multi</td>
<td>(2)/0.64*/0.71*/0.87*/0.91*</td>
<td>96</td>
</tr>
<tr>
<td>DCAA, TCAA, bromochloroacetic acid (BCAA)</td>
<td>hum/ful</td>
<td>0.93*/0.89*</td>
<td>96</td>
</tr>
<tr>
<td>DCAA, TCAA, dibromoacetic acid (DBAA)</td>
<td>hum/multi</td>
<td>0.72**, 0.67**/≥0.69**(2)</td>
<td>86</td>
</tr>
<tr>
<td>MCAA, DCAA, TCAA, monobromoacetic acid (MBAA), DBAA</td>
<td>hum/tyr/tryp/multi</td>
<td>≥0.81**(4), ≥0.55 (2)/0.86/0.74/0.86</td>
<td>16, 85, 90,</td>
</tr>
<tr>
<td>MCAA, DCAA, TCAA, BCAA, MBAA, DBAA</td>
<td>hum/m-hum/tryp</td>
<td>≥0.71**(2), 0.65*/≥0.56*(2)/0.54*</td>
<td>98, 102</td>
</tr>
<tr>
<td>Process Type</td>
<td>Compound Name</td>
<td>Concentration</td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>----------------------------------------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td><strong>MCAA, DCAA, TCAA, MBAA, DBAA,</strong></td>
<td>tribromoacetic acid, BCAA, bromodichloroacetic</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ozone – HAAs</strong></td>
<td>DCAA</td>
<td>multi 0.90**</td>
<td></td>
</tr>
<tr>
<td><strong>chlorine – haloketones (HKs)</strong></td>
<td>DCAA</td>
<td>hum 0.71</td>
<td></td>
</tr>
<tr>
<td>1,1-dichloro-2-propanone (DCP)</td>
<td>hum/tyr/tryp</td>
<td>≥0.86** (2), ≥0.52**/0.54**/0.60**</td>
<td></td>
</tr>
<tr>
<td>1,1,1-trichloro-2-propanone (TCP)</td>
<td>hum/ful/tryp</td>
<td>≥0.71*(5), ≥0.56 (3)/0.81*/0.72**</td>
<td></td>
</tr>
<tr>
<td>DCP, TCP</td>
<td>hum/m-hum/tyr/tryp/multi</td>
<td>≥0.72*(2)/0.71*/0.71*/0.76*/0.86*</td>
<td></td>
</tr>
<tr>
<td><strong>chlorine – haloacetaldehydes (HALs)</strong></td>
<td>2,2,2-trichloroethane-1,1-diol (chloral hydrate)</td>
<td>hum/ful</td>
<td></td>
</tr>
<tr>
<td><strong>UV+chlorine – HALs</strong></td>
<td>chloral hydrate</td>
<td>≥0.72*(6), ≥0.50**(2)/0.52**</td>
<td></td>
</tr>
<tr>
<td><strong>chlorine – halogenated furanones (X-furanones)</strong></td>
<td>4-chloro-3-dichloromethyl-2H-furan-5-one</td>
<td>hum</td>
<td></td>
</tr>
<tr>
<td><strong>monochloramine – iodinated DBPs (I-DBPs)</strong></td>
<td>dichloroiodomethane, bromochloroiodomethane, dibromoiiodomethane, chlorodiiodomethane</td>
<td>multi</td>
<td></td>
</tr>
<tr>
<td><strong>chlorine – carbonaceous disinfection byproducts (C-DBPs)</strong></td>
<td>monochloroacetic acid, dichloroacetic acid, trichloroacetic acid, TCM, DCP, TCP</td>
<td>hum/m- hum/tyr/tryp/multi</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCM, BDCM, DBCM, TBM, dichloroacetic acid, trichloroacetic acid, dibromooacetic acid</td>
<td>0.78*/0.70**/0.72*/0.88*/0.92*</td>
<td></td>
</tr>
<tr>
<td><strong>chlorine – halonitromethanes (HNMs)</strong></td>
<td>Trichloronitromethane (chloropicrin)</td>
<td>hum/m- hum/tyr/tryp/multi</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.65**/0.85*</td>
<td></td>
</tr>
<tr>
<td><strong>chlorine – haloacetonitriles (HANs)</strong></td>
<td></td>
<td>0.73*(4)/0.75*/≥0.71*(2)/≥0.80*(2), 0.65**/0.85*</td>
<td></td>
</tr>
<tr>
<td>Compound</td>
<td>PARAFAC Components</td>
<td>R²</td>
<td>References</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------------------</td>
<td>------</td>
<td>------------</td>
</tr>
<tr>
<td>bromochloroacetonitrile</td>
<td>hum/m-hum/tyr/tryp</td>
<td>0.76**/0.79**/0.59**/0.57**</td>
<td>95</td>
</tr>
<tr>
<td>dichloroacetonitrile</td>
<td>hum/tyr/tryp</td>
<td>≥0.83*(3), 0.60/0.52**/0.66**</td>
<td>99, 104</td>
</tr>
<tr>
<td>dichloroacetonitrile, bromochloroacetonitrile</td>
<td>hum/ful</td>
<td>0.90*/0.85*</td>
<td>96</td>
</tr>
<tr>
<td>dichloroacetonitrile, trichloroacetonitrile</td>
<td>hum/tyr/tryp</td>
<td>0.55/0.6/0.89**, 0.64</td>
<td>48, 75</td>
</tr>
<tr>
<td>dichloroacetonitrile, trichloroacetonitrile</td>
<td>hum/m-hum/tyr/tryp/multi</td>
<td>0.56*/0.64*/0.64*/0.78*/0.80*</td>
<td>66</td>
</tr>
<tr>
<td>dichloroacetonitrile, trichloroacetonitrile</td>
<td>tryp</td>
<td>≥0.85**(2)</td>
<td>67</td>
</tr>
<tr>
<td><strong>monochloramine – N-nitrosamines (NAs)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-nitrosodiphenylamine (NDPhA)</td>
<td>hum/multi</td>
<td>0.6/0.6</td>
<td>106</td>
</tr>
<tr>
<td>N-nitrosomorpholine</td>
<td>multi</td>
<td>0.55</td>
<td>106</td>
</tr>
<tr>
<td><strong>UV+monochloramine – cyanide (CNX)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cyanogen chloride</td>
<td>hum/tryp</td>
<td>0.91/0.79</td>
<td>21</td>
</tr>
<tr>
<td><strong>chlorine – nitrogenous disinfection byproducts (N-DBPs)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dichloroacetonitrile, bromochloroacetonitrile</td>
<td>hum/m-hum</td>
<td>0.50**/0.62**</td>
<td>95</td>
</tr>
<tr>
<td>trichloronitromethane, trichloroacetonitrile</td>
<td>hum/m-hum/tyr/tryp/multi</td>
<td>0.77*/0.80*/0.78*/0.86*/0.94*</td>
<td>66</td>
</tr>
</tbody>
</table>

Established linear relationships accounting for similar DBP species and disinfection methods (separated by a forward slash). This table is an extract summary of the report data framework (SI, Table S2). aSum of the DBPs formation potential for the reported species used to establish the relationship. bFluorescence maximal intensity of hum, ful, m-hum, tyr and tryp PARAFAC components which indicate humic-, fulvic-, microbial humic-, tyrosine- and tryptophan-like, respectively, used to establish the relationship. In addition, ‘multi’ indicates relationships using multivariable PARAFAC component indices, e.g., sum of all components, humic divide by tryptophan. The excitation–emission values associated for each PARAFAC component are reported in Table 1. cStrong linear relationship (R² ≥0.7) and moderate linear relationship (R² ≥0.5) between similar DBPs species and PARAFAC component are differentiated by a coma. Number of established relationships are expressed in parenthesis, * and ** indicate p-values of ≤0.05 and ≤0.01, respectively. In the case of several significant relationships, only the highest p-value is reported.
N-DBP classes were not as well represented in the 42 selected papers as C-DBPs. This group of DBPs was represented by HNMs, HANs, NAs and CNX in 31%, 31%, 5% and 2% of the selected articles (Figure 3), respectively. Despite the low number of articles for these DBP classes, noteworthy observations could still be mentioned. Overall, the total number of established relationships with humic/fulvic-like PARAFAC components versus protein-like were almost identical across N-DBP classes (Figure 3 & Table 2). In addition, none of the selected articles reported the PARAFAC component associated with wastewater or nutrient-enriched waters identified in a previous study ($\lambda_{ex}/\lambda_{em}$: 350/428 nm). This observation is surprising because, humic/fulvic-material has generally a low organic nitrogen content, <5% N/C mass ratio, compared to wastewater treatment or algal bloom, which have up to 20% N/C. In addition, protein-like components, which contain a high amino acid and N-organic compounds content might both serve preferentially as N-DBP formation precursors. In the context of identified algae/bacterial DOM sources, 72% of the successfully established relationships for chlorination are between protein-like components and HNMs and HANs formation potential, which is in agreement with the N-DBPs formation precursors described above. In contrast, other DOM sources, e.g., leaf leachate, natural soil/water organic matter, showed no preferential relationship with humic/fulvic-like or protein-like components and the formation of N-DBPs. These observations support the assumption that protein-like components are not always the main precursor of N-DBPs and that their formation pathways do not always involve similar amine precursor. Some N-DBP formation pathways may involve inorganic nitrogen or chloramine as the nitrogen source which react with humic/fulvic-like component to produce N-DBPs, but this cannot be concluded from PARAFAC modeling alone.
5. FUTURE RESEARCH NEEDS

DBP classes and disinfection methods

From the 42 selected articles, 90% and 50% of the investigated relationship include the THM and HAA groups, respectively (SI, Table S3). In addition, most of the articles focus on chlorine disinfection (81% of the selected articles). THMs are estimated to represent up to 7% and 47% of the total organic halogen content for chloramination and chlorination, respectively.78 Altogether, these observations highlights the need for more DBP species and disinfection methods to be investigated in further studies. For instance, emerging DBP classes, including HKs, HALs, HNMs, HANs, HAMs, NAs, I-HAAs and I-THMs need to be screened more frequently with advanced analytical methods to potentially established relationships with PARAFAC components.5,111 The potentially greater carcinogenic concerns associated with HANs and I-HAAs means these two classes should be prioritized in future investigations.12 In addition, various disinfectant methods might have, from a mechanistic point of view, various precursors and formation pathways with contrasting effects on a single PARAFAC component.21 Moreover, NAs, HANs and HAMs were recently identified to be a bigger concern for drinking water treatment plants when chloramine is used instead of chlorine.111-113 However, PARAFAC established relationships exist only with chlorine for HANs48, 66, 67, 75, 95, 96, 99, 104 and were investigated without success for the highly concerning NAs, N-nitrosodimethylamine (NDMA)81, 106 and HAMs.21, 91 This latest observation suggest a limitation of the fluorescence sensitivity coming from the additive properties of the Lamber-Beer’s Law31 where major DBPs species might mask the fluorescence measurement capability to detect minor DBPs species. However, these limitations should be evaluated and clearly demonstrated in future studies, e.g., by using sensitivity analysis.
Toward a unified model for DBP formation potential prediction

Recurrent strong linear relationships between PARAFAC components and DBP classes have been established in the selected articles. However, linear models developed with one particular DOM source were not necessarily that good at predicting DBP formation with different DOM sources. One potential reason is that DOM originating from different sources may exhibit very contrasting chemical composition. Interestingly, better relationships have been established using the humification index (HIX) to classify DOM and develop specific linear models for contrasted HIX samples. To overcome the challenges associated with a better understanding of DOM chemical composition, additional methods such as HR-MS, size-exclusion chromatography, and fractionation, might help to interpret the complexity of fluorescence EEMs across different DOM sources. Therefore, one current challenge would be to transpose site-specific DOM sources specific empirical relationships between PARAFAC component and DBPs formation potential into more global models considering DOM source variability. To overcome this challenge, robust statistical and computational data processing are needed. Machine learning (ML) algorithm such as self-organizing maps, neural networks, classification and regression tree analysis, already in use for fluorescence data might provide significant improvement of the DBP formation prediction. ML has the distinct advantage in taking into account non-linear relationships, interaction between variables, i.e., not independent, diverse variable types, e.g., continuous and discrete, and do not rely on predetermined physical based rules or assumptions on the given dataset. However, the success of the deployment of the ML algorithm relies on having sufficient contrasting data to train and validate properly an algorithm to recognize the DOM variability and complexity. This highlights the need for sharing raw EEM spectra to be able to perform unified analysis under the same workflow.
Worthy to note that people already share PARAFAC models under the online repository OpenFlor.\textsuperscript{121}

Implications and perspective for online continuous monitoring

Two approaches have been explored to perform online continuous monitoring in the literature: (i) to select some excitation–emission pairs from the PARAFAC analysis and use it to predict online DBPs formation potential with a portable relatively unexpensive in-situ UV light-emitting diodes (LED) fluorescence sensor.\textsuperscript{99, 122, 123} (ii) to obtain a full EEM spectra using a low cost fluorometer connected to the water with in-situ fiber-optic sensor\textsuperscript{124} and develop new, faster PARAFAC algorithms\textsuperscript{125, 126} to process data continuously. Currently, the method (i) is the most cost-effective method where UV-LEDs offer a narrow spectral emission centered on a specific wavelength and are commercially available for humic-like (Peak C; $\lambda_{\text{ex}}/\lambda_{\text{em}}$: 320–370 ± 15 / 450–490 ± 30 nm) and tryptophan-like (Peak T; $\lambda_{\text{ex}}/\lambda_{\text{em}}$: 270–285 ± 30 / 340–350 ± 30 nm) fluorophores. An exhaustive list of current commercially available probes with studies examples, and overview of pitfall and key practical aspects relative of the use of UV-LEDs have been reviewed elsewhere.\textsuperscript{38, 127} Custom made UV-LED systems have also been explored by some authors\textsuperscript{122, 128} and are expecting to grow with the rapid development of new UV-LEDs wavelength available on the market.\textsuperscript{129} In addition, there are several practical challenges, reviewed by Henderson\textsuperscript{37} to perform continuous online fluorescence measurement in surface water. The fluorescence signal may be affected by turbidity, pH, inner-filter effect at high absorbance, temperature, soluble metal content and presence of quenching elements, which require calibration and post-processing corrections. Current probes on the market integrate absorbance, i.e., inner-filter effect, turbidity, and temperature correction. However some
deviation might be observed under high suspended sediment loads or rapid changes in
temperature, which might require site specific calibration using bench scale fluorometers to
correct the continuous online measurement.\textsuperscript{130, 131} In particular, inner-filter effect should be
carefully evaluated by measuring the EEMs spectra with and without inner-filter effect
correction, since a 5\% change is considered to introduce higher bias than measurement
uncertainty.\textsuperscript{132} Despite the limit of having only one wavelength per sensor and the need to cross
validate results with bench top fluorescence, high-frequency sampling opens a new road to
explore rapid shifts in DOM source, concentration, and characterization in surface water.

6. CONCLUSIONS
To the best of our knowledge, the present work represents the first critical review to collate
and analyze associations between observed PARAFAC components and the potential formation
of specific DBP classes during water disinfection. Predicting DBPs formation from the optical
properties of DOM during water treatment presents new opportunities for online continuous
monitoring for raw water quality and treatment process optimization with a view to minimizing
harmful concentrations of DBPs in consumer tap water. In this regard, our review has
demonstrated that EEM-PARAFAC techniques offer considerable promise as a low-cost optical
surrogate for multiple classes of DBPs. From the 42 selected articles, we found a total of 202
linear relationships (126 with a correlation coefficient $\geq 0.7$) between formation of 10 DBP
classes and PARAFAC components. From the critical review, the specific findings can be
summarized as follow:

- Strong linear relationships were established between formation of C-DBP classes, i.e.,
  THMs, HAAs, HALs and HKs, and humic/fulvic-like PARAFAC components. In
contrast, the formation of N-DBP classes exhibited strong linear relationship across all fluorophore regions suggesting the contribution of several compounds as part of the N-DBPs precursor. However, exception is rule out for the case of algae/bacterial DOM source which exhibit strong relationship between N-DBPs and protein-like PARAFAC component only.

- Relationships using multivariable PARAFAC component indices, e.g., sum of components, humic-like divide by tryptophan-like, to predict DBPs formation potential showed stronger relationships compared to a single PARAFAC component model. This demonstrates the DOM complexity and confirming the contribution of different DBP precursors.

- The shift to shorter wavelength and decrease in fluorescence intensity are widely reported across all fluorophore regions during the process of chlorination. This observation demonstrates the contribution of multivariable PARAFAC components are necessary to predict the formation of one DBP class.

ASSOCIATED CONTENT

Supporting Information. If the manuscript is accepted the Supporting Information (SI) will be available free of charge at the ACS website

https://pubs.acs.org/doi/XXXXXXX

The following files are available.

S1: Additional information on the literature survey, experimental and modes of operation conditions (PDF)
Table S2: Report data framework, containing description of the selected articles and metadata associated used for the critical review (XLSX).

AUTHOR INFORMATION

* Corresponding author. e-mail: john.weatherill@ucc.ie and bodroz@bluewin.ch phone: +353 21 490 4578

ORCID Number

Elena Fernandez-Pascual: 0000-0002-4116-7734

Boris Droz: 0000-0002-3942-704X

Jean O’Dwyer: 0000-0001-7852-4662

Connie O’Driscoll: 0000-0002-1727-2964

Emma H. Goslan: 0000-0003-0367-226X

Simon Harrison: 0000-0002-2383-8995

John Weatherill: 0000-0001-9803-8397

Author Contributions

The manuscript was written through contributions of all authors. CRediT authorship contribution statement: EFP: Conceptualization, Data curation, Formal Analysis, Writing – original draft. BD: Data curation, Conceptualization, Methodology, Writing – original draft. JW: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft. JOD: Conceptualization, Formal Analysis, COD EHG and SH: Writing – review & editing. All authors have given approval to the final version of the manuscript. The authors declare no conflict of interest.
ACKNOWLEDGMENT

This work was carried out as part of the Proactive Optical Monitoring of Catchment Dissolved Organic Matter for Drinking Water Source Protection (PRODOM) project which is entirely funded by the Irish Environmental Protection Agency (Grant No. 2019-W-MS43) as part of the EPA Research Program 2021-2030.

REFERENCES


67. Ma, C.; Xu, H.; Zhang, L.; Pei, H.; Jin, Y., Use of fluorescence excitation-emission matrices coupled with parallel factor analysis to monitor C- and N-DBPs formation in drinking water


77. Munch, D. J.; Munch, J. W.; Pawlecki, A. M. *EPA Method 552.2, Rev. 1.0: Determination of haloacetic acids and dalapon in drinking water by liquid–liquid extraction, derivatization and gas chromatography with electron capture detection; U.S. EPA: Cincinnati, OH, 1995.*


100. Siddique, M. S.; Xiong, X.; Yang, H.; Maqbool, T.; Graham, N.; Yu, W., Dynamic variations in DOM and DBPs formation potential during surface water treatment by


106. Maqbool, T.; Zhang, J.; Qin, Y.; Ly, Q. V.; Asif, M. B.; Zhang, X.; Zhang, Z., Seasonal occurrence of N-nitrosamines and their association with dissolved organic matter in full-
scale drinking water systems: Determination by LC-MS and EEM-PARAFAC. *Water Res.* 2020, 183, 116096.


FOR TOC ONLY