PREDICTING WATER TREATMENT CHALLENGES FROM SOURCE WATER NATURAL ORGANIC MATTER CHARACTERIZATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL AND ENVIRONMENTAL ENGINEERING

Lauren E. Bergman

B.S. Civil Engineering, University of Michigan
M.S. Civil and Environmental Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA 15213

August 2016

ABSTRACT

Natural Organic Matter (NOM), a pervasive component of natural waters, presents many challenges for water treatment systems. Its complex and heterogeneous nature makes NOM difficult to characterize and highly variable in its effect in water treatment. Two specific water treatment challenges caused by NOM and dependent on its character are disinfection by-product (DBP) formation and organic fouling in pressure-driven membranes. Many NOM characterization methods exist and have shown success in highly controlled laboratory settings; however, evaluating their effectiveness in full-scale systems to predict DBP formation and membrane fouling remains an ongoing challenge. Fluorescence NOM Excitation Emission Matrices (EEM) are hypothesized to be effective in NOM characterization because they capture the complexity and heterogeneity of the NOM in data-rich measurements that are unique to each individual sample.

The objective of this work was to assess the utility of fluorescence EEM and other NOM characterization techniques for predicting DBP formation and membrane fouling in full-scale treatment systems. The review of current literature on NOM characterization and use in predicting water treatment challenges revealed patterns among NOM characterizations and water treatment outcomes – namely, high molecular weight, hydrophobic, aromatic NOM leads to increased DBP formation, while hydrophilic NOM with low aromaticity leads to increased organic fouling. Multiple reports from laboratory studies indicating the success of fluorescence measurements in characterizing DBP formation and membrane fouling suggest evaluation at full-scale treatment plants is warranted. The two field studies presented in this dissertation each address one of the major treatment challenges outlined – DBP formation and membrane fouling.

The DBP formation field study incorporated source water and finished water samples from six treatment plants along the Monongahela River in southwestern Pennsylvania to create a regional watershed model. Fluorescence measurements of the source water were used successfully to classify finished water DBPs according to various targets using classification trees. The membrane fouling study incorporated samples of the raw source water and treated water at various treatment stages within a full-scale two-pass (two-stage) reverse osmosis membrane treatment plant. Fluorescence measurements were successful in distinguishing between high fouling and low fouling periods within the plant, however, they were not capable of tracking treatability of source water throughout the pre-treatment steps. The results of the two field studies indicate that fluorescence measurements have utility in NOM characterization for full-scale treatment plant operations, but more research is needed in determining which specific signals are useful in online fluorescence detection and in assessing the broader applicability of these techniques to other geographical regions with different water qualities.

ACKNOWLEDGEMENTS

This research was made possible by many generous funding sources – the NEEP IGERT fellowship funded by the NSF, the CIT Dean's Fellowship, the PITA grant and additional support from Aquatech Inc., the Bradford and Diane Smith Graduate Fellowship, the Northrop Grumman Fellowship, an ARCS scholarship funded by Carol Heppner, Kathy Testoni and Maureen Young, and the WWOAP David A. Long Scholarship.

Thank you to my committee members, Jeanne VanBriesen, Kim Jones, Mitch Small and Dave Dzombak, for your involvement in my research and the time and effort you have taken in being a part of this dissertation. The invaluable advice and expertise you provide have helped shape my research and dissertation. I would especially like to thank the chair of my committee and my advisor, Jeanne. Jeanne is the reason I came to Carnegie Mellon and is the reason I completed the program, despite doubting myself (many times) along the way. You have been an extraordinary advisor and mentor, and I am lucky to have been able to work with you.

I also want to thank all the people that have helped me behind the scenes with the everyday tasks that sometimes seem like the most difficult ones – Ron Ripper, Maxine Leffard, Andrea Rooney, Hannah Diecks, Cornelia Moore, and Jodi Russo. Ron, none of this research would have been possible without you. From teaching me lab techniques and proper operation of lab equipment, to ordering chemicals for me while on lab exchanges, to buying a new refrigerator to chill all my (many) samples, to collecting sample shipments when I'm out of town, you have been a tremendous help throughout my PhD. I would not have this completed dissertation without you and I really appreciate everything you have done. Maxine, thank you for always keeping me on

track, whether it's registering for courses, meeting administrative deadlines, encouraging me to participate in the fun social programming you put together, or just being a friendly face around the department, I really appreciate all your hard work. Andrea, Hannah, Cornelia and Jodi, thank you for helping me to reserve rooms, ship packages across the country, track sample deliveries for me, print posters, complete reimbursements, and answer my endless administrative questions.

Thank you to our collaborators at Aquatech – C. Ravi, Mahesh Kumar, Ron Lewis, Stuart McGowan, and Joe Rayfield. Your help collecting samples, providing data, and corresponding with me over email and phone about plant operations, as well as your financial support made this project possible. Additionally, thank you to Clint Noack, Clint Mash, Dana Peck and David Bergman for statistical and programming assistance. From help with the PARAFAC Matlab program to learning the basics of R to discussing statistical concepts, your time, effort, thought and patience (I know it took a lot of patience) have contributed immensely to my research.

Thank you to all my friends that have been there along the way, including those from 207C, the VanBriesen research group, the CEE department, and others from around Pittsburgh. Whether it's having tea together, exploring Pittsburgh, going to a Buccos game, critiquing each other's presentations, travelling to conferences together, helping each other in the lab, going out for Friday lunch, or simply knowing there's a friend there when I need one, you have helped make my Pittsburgh experience unforgettable. I will miss spending time together and seeing each other regularly, but I know true friendships like these can last the distance and time.

Thank you to my family for your endless love and support. This process has been difficult and trying at times and it always helps to know you're cheering for me. And last, I want to thank my incredible husband, Dave. The list of things I want to thank you for is endless, but among the most important are your love, support, and advice. You helped me code in R (including one stint lasting 4 straight days), taught me many data analytics techniques (which have proven to be critical components of my PhD research), proof-read networking emails, critiqued my resume and cover letters, packed and moved me across the country (multiple times), cheered me on when I was feeling discouraged, and celebrated my accomplishments with me. Needless to say, this dissertation wouldn't be possible without you, and I can't thank you enough.

DISSERTATION COMMITTEE

Jeanne M. VanBriesen (Chair)
Duquesne Light Company Professor
Department of Civil and Environmental Engineering
Carnegie Mellon University
Director, Center for Water Quality in Urban Environmental Systems (WaterQUEST)

David A. Dzombak
Hamerschlag University Professor and Department Head,
Department of Civil and Environmental Engineering
Carnegie Mellon University

Kimberly L. Jones
Professor and Chair
Department of Civil and Environmental Engineering
Howard University

Mitchell J. Small
H. John Heinz Professor
Department of Civil and Environmental Engineering
Department of Engineering and Public Policy
Carnegie Mellon University

LIST OF ACRONYMS

AUC: Area Under the Curve
BIF: Bromine Incorporation Factor
C1/C2/C3: Component 1/Component 2/Component3
CF: Cleaning Filter
DOC: Dissolved Organic Carbon
EEM: Excitation Emission Matrix
NOM: Natural Organic Matter
PARAFAC: Parallel Factor Analysis
RO: Reverse Osmosis
ROC: Receiver Operator Characteristic
SUVA: Specific Ultraviolet Absorbance
SV: Site Validation
SW: Source Water
THM/TTHM: Trihalomethanes/Total Trihalomethanes
TOC: Total Organic Carbon
UF: Ultrafiltration Membrane
UV/UV_{254}: Ultraviolet/Ultraviolet absorbance at 254 nm

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION, PROBLEM IDENTIFICATION, AND RESEARCH OBJECTIVES

1.1 Introduction

Natural Organic Matter (NOM) is a universal component of natural aquatic systems, but presents significant challenges for water treatment operations. Not only does it degrade the aesthetics by altering the taste, color and odor, NOM contributes to the formation of toxic disinfection by-products and increases the operational cost of membrane treatment due to fouling. NOM is a heterogeneous mixture of carbon-based materials that contains a range of molecular weights, functional groups, molecular structures, and elemental compositions (Wong *et al.*, 2002; Matilainen *et al.*, 2011; Owen *et al.*, 1995). These different organic carbon compounds with diverse properties have highly variable effects on water treatment systems (Tran *et al.*, 2015; Owen *et al.*, 1995; Ivancev-Tumbas, 2014).

In the present work, the presence of natural organic matter and its effect on water treatability is explored through consideration of disinfection by-product formation and membrane fouling control. Drinking water disinfection is a critical component of water treatment – inactivating most pathogenic microorganisms present in source waters and ensuring safe water is delivered to consumers. However, the strong oxidizing agents used for disinfection react with organic matter that is not fully removed in treatment steps to form disinfection by-products (DBPs). DBPs are associated with adverse health effects, such as bladder cancer and low birth weight (Cantor *et al.*, 2010; Villanueva *et al.*, 2004; Danileviciute *et al.*, 2012; Kumar *et al.*, 2014). Regulation of disinfection includes consideration of the balancing of risk from microbial contaminants and the risk from the DBPs that form (EPA, 2010).

Disinfection enables the use of freshwater sources for water consumption, but these sources are increasingly under stress due to global population growth and climate change. Membrane technology, specifically reverse osmosis membrane treatment, plays an important role in augmenting limited freshwater resources through desalination and treatment of brackish waters. The principal challenge facing the membrane separation process is fouling, generally characterized as a loss of performance in the membrane system. Organic fouling, a reduction in hydraulic permeability due to accumulation of organic foulants on the membrane, is a concern in all membrane treatment processes and often precedes more severe biological fouling. Membrane fouling increases operational costs through the additional pressure required to maintain a constant flux through the membrane despite reduced permeability. Overall, fouling adds to the already high costs of membrane treatment, limiting the use of membrane technology.

The diversity of NOM structures and fractions makes it (1) difficult to characterize NOM simply, yet comprehensively and (2) difficult to connect NOM character to water treatment challenges. Significant research has attempted to address these issues, yet the challenge of developing a NOM characterization technique that can effectively capture its complexity and relate it to downstream problems in water treatment is ongoing. Accurate predictions of how the NOM in the source water will affect downstream water treatment operations could greatly improve the economical provision of safe and clean water to consumers. Specifically, understanding the connection between NOM character and adverse water treatment outcomes would help operators identify problems in advance and implement additional pre-treatments to remove harmful NOM

prior to treatment, making the finished water safer and the entire treatment system more cost-effective.

1.2 Problem Identification and Research Objectives

There are many methods currently available for measurement and characterization of NOM; however, their utility for treatment system management and optimization is unclear. NOM characterization must be improved in order to understand and control formation of harmful disinfection byproducts and fouling of reverse osmosis membranes.

This dissertation assesses the utility of a specific NOM characterization technique, fluorescence Excitation Emission Matrices (EEM), in each of these different, but important water treatment challenges. Fluorescence EEM have been proposed to address these characterization challenges because they capture the fluorescence character of NOM with data-rich measurements that are unique to each individual sample (Stedmon and Bro, 2008). Along with Parallel Factor Analysis (PARAFAC), EEM can be decomposed into a few representative components that can be incorporated into statistical models used to predict water treatment challenges (Stedmon *et al.*, 2003b; Stedmon and Bro, 2008; Bro, 1997). Given the success of EEM-PARAFAC components in bench-scale and lab-scale water treatment studies (Pifer and Fairey, 2012; 2014; Johnstone *et al.*, 2009; Peiris *et al.*, 2010b; Peiris *et al.*, 2010a), it is expected that this NOM characterization technique will also provide useful results for full-scale treatment plants experiencing NOM challenges from natural waters. These results will be essential in making progress towards implementation of online fluorescence monitoring of influent water in full-scale systems.

There are three research objectives:

1. To assess NOM measurement techniques, with an emphasis on fluorescence measurements, and their use in predicting DBP formation and membrane fouling through a review of published studies;

2. To create watershed-level DBP formation prediction models using fluorescence NOM measurements that define treatability of the water source, with a focus on relevant regulatory and operational parameters; and

3. To link fluorescence NOM measurements to observed fouling events in a full-scale membrane treatment plant and track changes in NOM due to pre-treatment using fluorescence.

1.3 Structure of the Dissertation

The dissertation is made up of five chapters, including an introduction, a literature review, two research papers that are intended for publication in peer-reviewed journals (one has been accepted and one is in preparation), and a conclusion. Chapter 1, the introduction, provides the motivation for the research along with an overview of the dissertation. Chapter 2, the literature review, provides the background necessary for the two research papers, including an overview of natural organic matter characterization and how it has been used in disinfection by-product and membrane fouling studies. Chapter 3 focuses on predicting basin-wide finished water DBP targets based on source water NOM characterization using classification trees. Chapter 4 focuses on the application of NOM characterization for predicting fouling events and treatability under various pre-treatments in a full-scale reverse osmosis membrane treatment plant. Chapter 5, the conclusion, summarizes the major findings presented in the dissertation and the potential for future work.

Chapter 2

REVIEW OF FLUORESCENCE ORGANIC CARBON CHARACTERIZATION FOR ENGINEERED WATER TREATMENT SYSTEMS

2.1 Abstract

Natural organic matter (NOM) in source water leads to many water treatment challenges, including disinfection by-product formation and organic fouling in membranes. Extensive research to minimize these two challenges is ongoing. However, given the highly complex and heterogeneous nature, characterizing NOM for application in water treatment systems remains a challenging task. This review provides an overview of NOM measurement and characterization techniques that are often used in water treatment plants and studies, with a focus on fluorescence measurements, and outlines current knowledge of how these relate to disinfection by-product (DBP) formation and membrane fouling. Patterns of NOM characterization found within the literature are described, including NOM fractions that are "highly reactive" in DBP formation and NOM fractions that are commonly identified as "foulants." Further, fluorescence measurements have shown success in many studies in characterizing DBP formation and membrane fouling in bench-scale and laboratory-scale studies. Pre-treatment, commonly used to reduce NOM in the treatment plant, is also discussed as well as how it affects various NOM fractions and how it has been employed in DBP and membrane fouling studies. Finally, this overview of NOM characterization for specific water treatment challenges highlights important gaps and inconsistencies where further research is needed.

2.2 Introduction

Understanding the complex, heterogeneous nature of organic carbon in water and identifying specific organic components that can negatively affect treatment operations is a critical step in improving water treatment. Natural Organic Matter (NOM), a mixture of compounds, is found in all natural waters and varies in composition depending on the source (Frimmel, 1998 ; Baghoth *et al.*, 2011; Cabaniss and Shuman, 1987; Sierra *et al.*, 1994; Nissinen *et al.*, 2001; Goldman *et al.*, 2014; Jacob Daniel Hosen *et al.*, 2014). The character of NOM affects many water treatment processes, including conventional surface water treatment unit operations, the formation of carcinogenic disinfection byproducts, and organic fouling in membrane treatment plants (Bieroza *et al.*, 2009; Sanchez *et al.*, 2013; Pifer and Fairey, 2012; Pisarenko *et al.*, 2013; Rodriguez *et al.*, 2007; Kennedy *et al.*, 2008; Zhang *et al.*, 2014; Shao *et al.*, 2014; Yamamura *et al.*, 2014).

2.3 Natural Organic Matter Characterization

One common method of analyzing organic carbon from natural samples is to measure the Total Organic Carbon (TOC). The Wet-Dry Combustion Method was first developed by Pickhardt *et al.* (1955), and today TOC is commonly measured on combustion or UV/persulfate analyzers by oxidizing samples and measuring the oxidation products. While TOC does not provide information about the character of the sample, it provides a quantitative measure of the organic carbon present in the sample. Dissolved Organic Carbon (DOC) is measured the same way on samples that have been filtered through a 0.45µm filter.

Ultraviolet (UV) absorbance and specific ultraviolet absorbance (SUVA; UV absorbance normalized by DOC) have long been used in NOM characterization because they provide more information about the character of the organic carbon (Weishaar *et al.*, 2003; Traina *et al.*, 1990; Chin *et al.*, 1994; Korshin *et al.*, 1997). Some studies have found high correlations between TOC/DOC and UV absorbance (Edzwald *et al.*, 1985; Shao *et al.*, 2014); however, UV absorbance is related to the aromaticity rather than the quantity of the NOM (Weishaar *et al.*, 2003; Traina *et al.*, 1990; Chin *et al.*, 1994). In addition to UV and SUVA, other UV-based measurements, such as ratios of absorbance at different UV wavelengths, UV absorbance spectra slope, and differential absorbance provide additional information about the NOM character (Korshin *et al.*, 1997; Roccaro *et al.*, 2015; Louie *et al.*, 2013; Lavonen *et al.*, 2015; Roccaro *et al.*, 2008; Roccaro *et al.*, 2009).

Since NOM is made up of many different components, fractionation is often the first step in analysis. Size exclusion chromatography (SEC), liquid chromatography (LC), or dialysis can be used to separate by size (Li *et al.*, 2014b; Chen *et al.*, 2014a; Vuorio *et al.*, 1998; Rausa *et al.*, 1991; Nissinen *et al.*, 2001; Gloor and Leidner, 1979; Chin *et al.*, 1994; Kennedy *et al.*, 2008). High Performance SEC allows for determination of molecular weights and polydispersity of the NOM within the sample and can be used to determine the changes in size distribution that occur throughout water treatment (Gloor and Leidner, 1979; Nissinen *et al.*, 2001; Vuorio *et al.*, 1998). Hydrophobic and hydrophilic NOM fractionation is commonly performed using XAD resins and membrane separation (Hua *et al.*, 2015; Hua and Reckhow, 2007a; Gray *et al.*, 2011; Kennedy *et al.*, 2005; Kitis *et al.*, 2002; Li *et al.*, 2014a; Yamamura *et al.*, 2014; He and Hur, 2015; Wong *et al.*, 2002), and humic/fulvic fractionation is also performed using resins or other

centrifugation/acidification extraction techniques (Reckhow *et al.*, 1990; Miller and Uden, 1983; Babcock and Singer, 1979; Coble, 1996; Hua *et al.*, 2015). Wong *et al.* (2002) demonstrated the ability of size and hydrophobic/hydrophilic NOM fractionation to distinguish among multiple water sources.

In an effort to capture the complexity of NOM in one comprehensive measurement, fluorescence characterization of NOM has become increasingly popular; the measurement provides a simple way to quickly characterize the NOM within each sample. Excitation-Emission Matrices (EEM) provide a unique fingerprint of the organic matter in a sample (Pifer and Fairey, 2012; Pifer *et al.*, 2011; Stedmon *et al.*, 2003a; Stedmon and Markager, 2005). Fluorescence techniques for organic carbon characterization have been used in disinfection byproduct (DBP) studies (Hua *et al.*, 2006a; Pifer and Fairey, 2012) and in membrane fouling studies (Chen *et al.*, 2014a; Choi *et al.*, 2014; Peiris *et al.*, 2010a; Peiris *et al.*, 2010b; Peiris *et al.*, 2013). Each water sample EEM provides fluorescence intensities for many pairs of excitation and emission wavelengths. And each EEM shows a three-dimensional plot of intensity values versus excitation wavelengths and emission wavelengths from the organic matter in the sample.

Given the large amount of data captured within sample EEM, multiple analytical techniques have been developed to make fluorescence EEM data accessible for further data analysis, including (1) Peak Picking, (2) Fluorescence Regional Integration, (3) Principal Component Analysis, and (4) Parallel Factor Analysis. Peak picking is used as a way to extract a smaller amount of information from the fluorescence EEM by selecting the maximum of each main fluorescence signal (usually one or two) for each sample EEM. With peak picking, the fluorescence intensity

of the EEM maximum (peak) as well as the location of the peak can be used to describe sample fluorescence character. Peak picking has been used to evaluate differences in organic matter (Coble, 1996; He and Hur, 2015), but results have a high level of uncertainty (Korak *et al.*, 2013). Fluorescence Regional Integration (FRI) was developed to summarize the total EEM signal by integrating the volume under the EEM plot (Chen *et al.*, 2003). FRI has been used for advanced organic matter characterization and identifying specific fractions of interest (Li *et al.*, 2013; He *et al.*, 2013; He and Hur, 2015). Principal component analysis has also been used in some fluorescence EEM studies because it enables the use of the entire sample EEM while summarizing the whole fluorescence dataset into a few representative values. Principal component analysis of sample EEM has been used successfully to relate fluorescence signals to DBP formation and membrane fouling (Peleato and Andrews, 2015; Chen *et al.*, 2014a; Peiris *et al.*, 2010a; Peiris *et al.*, 2010b).

Parallel Factor Analysis (PARAFAC) has become a widely-used statistical analysis tool for EEM data because it provides a summary of large datasets by determining a few representative components of the multi-dimensional dataset. Further, PARAFAC is able to handle three-dimensional EEM data (Bro, 1997), and PARAFAC components represent actual fluorophores present in the EEM dataset (Stedmon and Bro, 2008). PARAFAC for EEM analysis can be used to determine variations in a multi-dimensional matrix and to specifically identify the independent variables responsible for variations in large sets of multivariate data (Harshman and Lundy, 1994; Bro, 1997). Equation 2.1 is the governing equation for PARAFAC, as used in EEM applications

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \qquad \textbf{2.1}$$

In Equation 2.1, x_{ijk} represents the fluorescence intensity of one element in the three way array, X. In terms of the EEM model, *i* is the sample, *j* is the emission wavelength, *k* is the excitation wavelength, *a* is the concentration, *b* is the emission spectra, *c* is the excitation spectra, *f* is a fluorophore (component), *F* is the total number of fluorophores, and *e* is the residual, or additional variability in the data set that is not captured in the model. The developed model aims to minimize the sum of the squared residuals (Stedmon *et al.*, 2003b; Stedmon and Bro, 2008). Essentially, the total signal is composed of the sum of the individual fluorophore signals, which are made up of the concentration, emission wavelength, and excitation wavelength. Multiple studies have developed classifications of the fluorescence signals as a means to distinguish them and identify NOM fractions that may be responsible for the signals. Some of the commonly used classifications are presented in Table 2.1

Table 2.1: Classification of EEM fluorescence signals by NOM fraction

Region (EX/EM nm)	**Classification**	**Reference**
EX = 200 – 250 EM = 280 – 380	Aromatic Protein	(Chen *et al.*, 2003)
EX = 200 – 250 EM = 380 – 540	Fulvic Acid-like	(Chen *et al.*, 2003)
EX = 250 – 330 EM = 280 – 380	Soluble Microbial By-Product Protein-like	(Chen *et al.*, 2003) (Coble, 1996) (Her *et al.*, 2003)
EX = 250 – 400 EM = 380 – 540	Humic Acid-like Fulvic Acid-like	(Chen *et al.*, 2003) (Coble, 1996) (Her *et al.*, 2003) (Lochmuller and Saavedra, 1986)

Although the groupings in Table 2.1 are widely used and are helpful in classifying fluorescence signals, a fluorescence signal alone cannot confirm the presence of a specific organic fraction because a particular signal may be comprised of one or a sum of multiple organic fluorophores (Stedmon and Bro, 2008; Coble, 1996). A major limitation in the EEM-PARAFAC method of characterizing organic carbon is the inability to link resultant components with specific NOM fractions. Li *et al.* (2014b) used liquid chromatography and size exclusion chromatography along with EEM-PARAFAC analysis to determine that EEM-PARAFAC components could not be used to identify organic species in NOM because some compositionally different species exhibited the same fluorescent signals. Coble (1996) reported "humic-like" fluorescence signals come from a combination of different fluorophores.

Although fluorescence EEM are limited in their fundamental characterization of NOM fractions, they have been used in differentiating among other NOM properties. Cuss and Gueguen (2014) found that changes in fluorescence were associated with differences in molecular weight. Further, EEM-PARAFAC components are often correlated with DOC and UV_{254} (Baghoth *et al.*, 2011; Shao *et al.*, 2014; Johnstone *et al.*, 2009). In terms of using EEM signals for source identification, there have been some contradictory findings. Sierra *et al.* (1994) and Coble (1996) found that it ocean and freshwater samples showed distinct fluorescence signals, while McKnight *et al.* (2001) found very similar fluorescence peaks between ocean and freshwater fulvics. EEM, however, have shown promise for water treatment studies, demonstrating the ability to track NOM changes throughout a treatment train, which is important in addressing treatability concerns associated with DBP formation and membrane fouling (Baghoth *et al.*,

2011; Sanchez *et al.*, 2013; Shao *et al.*, 2014; Peleato *et al.*, 2016). Ratios of EEM-PARAFAC components also provide insight into the relative contribution of different NOM fractions (Baghoth *et al.*, 2011; Shao *et al.*, 2014). Baghoth *et al.* (2011) used humic-like to protein-like component ratios to track the change in humic/protein NOM ratios throughout treatment and determine which NOM fractions were preferentially removed in each treatment process. Carstea *et al.* (2014) also used humic-like to protein-like component ratios to describe the relative contribution of rural to urban water sources and therefore the relative impact of anthropogenic activities. Given their ability to differentiate among multiple NOM samples, along with their relatively easy and inexpensive operation, EEM have potential for use in many engineering applications, including prediction of DBP formation and membrane fouling.

2.4 Background on Disinfection By-Product Formation

Disinfection is an important component in drinking water treatment because it keeps water safe for consumers by inactivating many pathogenic microorganisms found in the source water. However, as a result, toxic disinfection by-products (DBPs) form when disinfectants oxidize NOM in the source water. DBPs have been linked to adverse health effects, such as bladder cancer and low birth weight (Danileviciute *et al.*, 2012; King and Marrett, 1996; Kumar *et al.*, 2014; Villanueva *et al.*, 2004). Further research has found that the ability to metabolize trihalomethanes and thereby increase the odds of developing bladder cancer is based on a specific gene that a portion of the population carries (Cantor *et al.*, 2010).

Research has identified hundreds of different disinfection by-products (Richardson *et al.*, 2007; Boorman *et al.*, 1999; Richardson *et al.*, 2000) and the specific DBPs formed in water treatment depend on many different variables, among them (1) the type of disinfectant used, (2) the presence of dissolved ions, and (3) the character of the NOM. Chlorine is a widely used disinfectant and leads to many halogenated DBPs, including trihalomethanes (THMs) and haloacetic acids (HAAs), two of the DBP classes currently regulated by the EPA (EPA, 2006; Durmishi *et al.*, 2015; Rathburn, 1996b; Nokes, 1999; Liang and Singer, 2003; Amy *et al.*, 1987; Singer *et al.*, 2002). Additional chlorine disinfection by-products include haloacetonitriles, chloral hydrate, haloketones, and chlorophenols (Roccaro and Vagliasindi, 2010; Miller and Uden, 1983; Reckhow *et al.*, 1990; Oliver and Lawrence, 1979; Chu *et al.*, 2012). Despite the challenges associated with DBP formation, chlorine remains the most commonly used disinfectant (Siedel *et al.*, 2005).

Alternative disinfectants, including chlorine dioxide, chloramine, and ozone, are used in some treatment plants in an effort to control THMs and HAAs (Richardson *et al.*, 2000; Tian *et al.*, 2013; Lu *et al.*, 2009; Richardson *et al.*, 1994); however, these alternative disinfectants result in other species of disinfection by-products. Chlorine dioxide (ClO_2) produces chlorite and chlorate from NOM oxidation, in addition to multiple species of carboxylic acids, chloro-benzenes, and halopropanones (Korn *et al.*, 2002; Richardson *et al.*, 2000; Richardson *et al.*, 1994; EPA, 2010). Chloramination, the use of monochloramine (NH_2Cl), leads to nitrogenous DBPs (N-DBPs), such as haloacetonitriles and nitrosodimethylamine (NDMA), which research has shown are more toxic than carbon-based DBPs (i.e. HAAs) (Sakai *et al.*, 2015; Muellner *et al.*, 2007). Chloramination also produces THMs and HAAs, but to a lesser extent (Tian *et al.*, 2013; Lu *et*

al., 2009). Ozone (O_3) is known to produce multiple species of aldehydes, ketones, and ketoacids instead of halogenated by-products (Richardson *et al.*, 2000; Karnik *et al.*, 2005), and leads to formation of bromate, a regulated DBP, in areas experiencing higher bromide loading, such as coastal areas (Gyparakis and Diamadopoulos, 2007; Moslemi, 2012; EPA, 2010; Haag and Holgne, 1983). Additionally, increased concentrations of brominated DBPs have been observed when ozone and chlorine are used together (Mao *et al.*, 2014).

The presence of dissolved ions, especially bromide, in source waters is also a concern because bromide increases the rate of DBP oxidation reactions and leads to more toxic brominated DBPs (Plewa *et al.*, 2002; Richardson *et al.*, 2007; Richardson *et al.*, 2003). As discussed previously, bromide is usually only a concern in coastal areas where ground waters and surface waters may experience sea water intrusion and as a result an increase in dissolved salts, including bromide (Gyparakis and Diamadopoulos, 2007; Ged and Boyer, 2014). However, with new energy extraction activities, such as unconventional hydraulic fracturing, that produce wastewater high in dissolved salts, there are new sources of bromide to inland waterways (Wilson and Van Briesen, 2013; States *et al.*, 2013). As a result, disinfection by-products formed in the region may show shifts towards more brominated forms since bromide in source water increases brominated DBP concentration (Nokes, 1999; Cowman, 1996; Chowdhury *et al.*, 2010; Watson *et al.*, 2015; Navalon *et al.*, 2008). In addition to bromide, iodide in the source water can lead iodide-containing DBPs with even higher toxicity (Allard *et al.*, 2015; Plewa *et al.*, 2004; Hua *et al.*, 2006b; Criquet *et al.*, 2012).

2.5 Natural Organic Matter and Disinfection by-Products

Natural organic matter (NOM) is the main precursor to DBP formation and its character is also an important input variable in DBP formation and speciation. The high level of NOM variability among sources has been shown to result in high variability of DBP formation and speciation (Weiss *et al.,* 2013; Kitis *et al.,* 2002). While TOC and DOC are used to measure the amount of organic matter present, they rarely provide adequate quantitative prediction of DBPs formed. An extensive literature review by Chowdhury (2009) demonstrated the importance of TOC and DOC as model input parameters for DBPs – with most successful Trihalomethane (THM) and Haloacetic Acid (HAA) models incorporating either TOC or DOC. However, there are many different reports of the relationship between TOC/DOC and DBP formation potential; some studies report high correlations (Edzwald *et al.*, 1985; Amy *et al.*, 1987; Rook, 1976), while others' results show that the two variables are uncorrelated (Li *et al.* (2014a) or that only some DBP classes are correlated (Chen and Westerhoff, 2010). The literature suggests that TOC/DOC is an important variable in determining DBP formation, but alone is not successful in making predictions.

UV absorbance at 254 nm has also been used in many DBP studies to predict formation. Studies report correlation between UV absorbance and THM formation potential (Amy *et al.,* 1987; Kitis *et al.,* 2002; Roccaro *et al.,* 2015; Roccaro *et al.*, 2008); however, changes in highly variable NOM limit its applicability (Abouleish and Wells, 2015; Edzwald *et al.*, 1985; Shao *et al.*, 2014). Models incorporate UV, SUVA, or even UV-TOC composite terms as input to generate DBP predictions (Korn *et al.*, 2002; Amy *et al.*, 1987; Chowdhury, 2009). Research on the relationship between SUVA and DBP formation suggests aromatic carbon structures (NOM

fractions that also absorb UV) are more reactive with chlorine and therefore lead to increased DBP formation (Hua *et al.*, 2015; Kitis *et al.*, 2001; Kitis *et al.*, 2002; Awad *et al.*, 2016). This is further confirmed by experimental results showing chlorine consumption increasing linearly as the percent of aromatic carbon increases in a treated water (Reckhow *et al.*, 1990). UV/SUVA, however, is limited as a DBP formation potential surrogate in low aromatic source water (Li *et al.*, 2014a). Although not highly predictive of overall DBP formation, low SUVA values indicate another issue within DBP formation – bromine incorporation. Studies have found that under lower SUVA values, bromine experiences higher incorporation into DBPs (Kitis *et al.*, 2001; Kitis *et al.*, 2002). UV absorbance may show improved THM formation potential prediction over DOC because it captures the NOM characteristics that are relevant to THM formation.

Other UV absorbance parameters, such as ratios of absorbance at different UV wavelengths, the slope of the UV absorbance spectra and differential absorbance, have also been used successfully in DBP formation studies. The ratio of absorbance at 253 nm to 203 nm wavelengths was found to be highly correlated with chloroform formation (Korshin *et al.*, 1997). Further, the slope of the UV spectra between 280 nm and 350 nm was found to be related to percent aromaticity and formation of total haloacetic acids (THAA) and total trihalomethanes (TTHM) (Roccaro *et al.*, 2015). The utility of UV spectral slopes in DBP formation studies agrees with other studies that show that differences in UV slopes indicate differences in NOM composition (Louie *et al.*, 2013). Differential absorbance has also been used to track DOM changes when DOC is low and in DBP predictive studies (Lavonen *et al.*, 2015; Roccaro *et al.*, 2008; Roccaro *et al.*, 2009).

Fractionation of NOM (i.e. hydrophobic vs hydrophilic content, molecular weight, and humic vs fulvic fractions) has been used extensively to better understand the relationship between NOM character and DBP formation. Hydrophobic fractions are generally more reactive and therefore produce more DBPs than hydrophilic fractions (Kitis *et al.*, 2002). More specifically, the hydrophobic fraction is more reactive with chlorine and therefore produces more chloroform and TCAA; whereas the hydrophilic fraction is more reactive with bromide and therefore produces more Br-DBPs (Li *et al.*, 2014a; Hua and Reckhow, 2007a). The hydrophobic fraction, a halogenated DBP precursor, was also found to have a higher humic content and more aromatic structures (Hua *et al.*, 2015; Wong *et al.*, 2002). The hydrophilic fraction also contributes to DBP formation, but to different DBP classes and to a lesser extent than the hydrophobic fraction (Hua and Reckhow, 2007a). Although the hydrophobic fraction produces more DBPs under chlorination and chloramination, contradictory results were found by Hua *et al.* (2015) whose experiments showed that hydrophilic fractions had higher chlorine demands than hydrophobic ones. Given than chlorine consumption is related to NOM-DBP reactivity, measured as aromatic content (Reckhow *et al.*, 1990), it is expected that the more reactive hydrophobic fractions would have higher chlorine demands.

Molecular weight (MW) and Humic/Fulvic fractionation also provide insight into DBP formation potential. Higher MW fractions were found to produce more DBPs (high MW fractions were more reactive), however there was higher bromine incorporation with lower MW fractions (Kitis *et al.*, 2002). Like the hydrophobicity and chlorine demand results, Hua *et al.* (2015) also found unexpected molecular weight and chlorine demand results – higher chlorine demands were found with smaller MW NOM fractions. There is also some disagreement about

the effect of humic and fulvic fractions on DBP formation and speciation. Reckhow *et al.* (1990) found that humic acids produced more DBPs than fulvic acids due to a higher humic acid chlorine consumption, while Miller and Uden (1983) found that fulvic acids produced more DBPs than humic acids, however, these contradictory results were due to the fact that the humics used in experimentation had fewer activated aromatic structures . Chloroform concentration, as well as chlorine consumption, was found to increase linearly with humic acid concentration when excess chlorine was present, while less chloroform formation was observed with fulvic acids (Babcock and Singer, 1979). The observed association between higher SUVA values, hydrophobicity, and higher molecular weight NOM fractions, suggests that there are more aromatic structures in hydrophobic and high MW NOM fractions (Hua *et al.*, 2015).

The accumulation of results from various NOM fractionation studies and their associated DBP formation potentials provides evidence for two main NOM fractions: (1) high reactivity and DBP formation potential and (2) low reactivity and DBP formation potential. Figure 2.1 illustrates the relationship between NOM characteristics and resulting DBP formation from chlorine disinfection, based on general themes found in the literature.

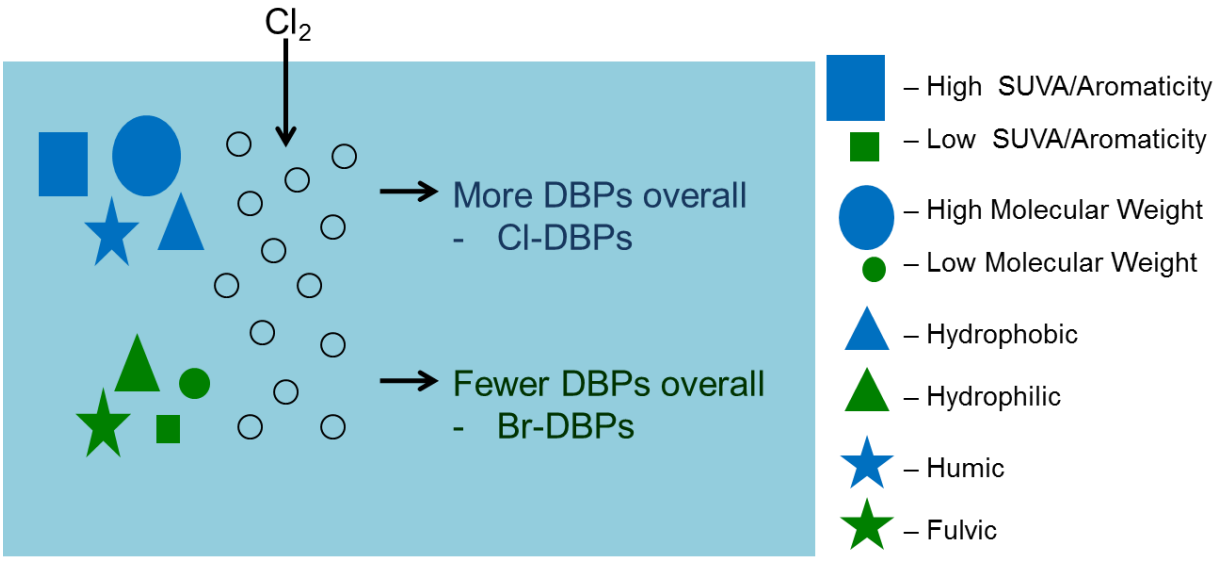

Figure 2.1: Illustration of common NOM characterization and subsequent DBP formation patterns from chlorine disinfection found in the literature. According to the literature, aromatic, high molecular weight, hydrophobic and humic NOM leads to increase DBP formation, especially chlorinated forms. Whereas less aromatic, low molecular weight, hydrophilic and fulvic NOM fractions result in fewer DBPs overall, but produce more brominated species.

High DBP reactivity is characterized by higher aromatic content (higher SUVA values), higher molecular weights and hydrophobicity while lower DBP reactivity is characterized by lower aromatic content (lower SUVA values), lower molecular weights and hydrophilicity (Li *et al.*, 2014a; Hua *et al.*, 2015; Kitis *et al.*, 2002). These differences in NOM character have also been found to result in differences in speciation of DBPs. Hua *et al.* (2015) found that high molecular weight, hydrophobic fractions ("high DBP reactivity") are related to uncharacterized DBP species, while low molecular weight, hydrophilic fractions ("low DBP reactivity") are related to regulated DBPs, such as THM and HAA acids. These reported relationships, however, are not constant across all water samples. For example, (Kitis *et al.*, 2002) tested multiple source waters and found that one source water showed a clear relationship between aromaticity and molecular weight (i.e. an increase in SUVA was correlated with an increase in molecular weight), however,

the other source water did not exhibit the same trend. Inconsistencies such as these highlight the need to further investigate other methods for characterizing NOM. Current techniques lack the consistency and reliability necessary to provide input for process control.

Fluorescence EEM offer another method of NOM characterization that provides data-rich quantitative measurements of NOM within an aqueous sample. While they are not perfect representations of NOM fractions, EEM NOM signals (from PARAFAC and PCA components), such as those identified in Table 2.1, have been used successfully to predict DBP formation in laboratory studies. Humic-like EEM-PARAFAC components have been found to be highly correlated with TTHM formation potential and higher chlorine reactivity (Pifer and Fairey, 2014; Yang *et al.*, 2015b; Pifer and Fairey, 2012; Ma *et al.*, 2014). Meanwhile, Johnstone *et al.* (2009) found that both the marine EEM-PARAFAC humic-like and protein-like fluorescence signals were predictive of chloroform and trichloroacetic acid formation in treated water. Furthermore, EEM-PCA protein-like fluorescence signals were found to provide improved predictions of both THM and HAA formation in laboratory tests of natural samples (Peleato and Andrews, 2015).

2.6 Background on Organic Fouling in Membranes

Organic fouling is caused by a build-up of adsorbed natural organic matter (NOM) on the membrane surface or in the membrane pores, which, over time can lead to bacterial growth on the surface and eventually, biological fouling (Martínez *et al.*, 2015; Arora and Trompeter, 1983; Herzberg and Elimelech, 2007; Rukapan *et al.*, 2015; Nam *et al.*, 2013; Zhao *et al.*, 2010).

The build-up of NOM on the membrane, and eventual bacterial growth, increases the osmotic pressure across the membrane, which reduces hydraulic permeability of the membrane. Additionally, the organic fouling layer that develops on the membrane surface reduces the solute rejection in reverse osmosis membranes, which are designed to remove dissolved mono-valent ions, resulting in lower quality permeate water (Hoek and Elimelech, 2003; Hoek *et al.*, 2002; Song and Elimelech, 1995; Schäfer *et al.*, 2000). In porous microfiltration (MF), ultrafiltration (UF), and nanofiltration (NF) membranes, organic fouling results from the build-up of organic matter in the membrane pores and on the surface; whereas with (non-porous) reverse osmosis (RO) membranes, organic fouling is a result of the organic build-up on the membrane surface (Rukapan *et al.*, 2015; Nam *et al.*, 2013). The abundance and composition of organic matter in source water affects the structure of the fouling layer and consequently, the amount of flux decline that occurs during fouling (Ang *et al.*, 2011; Zhao *et al.*, 2010; Tiraferri and Elimelech, 2012; Airey *et al.*, 1998; Zhu and Elimelech, 1997; Tang *et al.*, 2007).

In membrane systems that operate under a constant pressure, such as bench-scale systems in a laboratory setting, membrane fouling is observed as a loss of flux over time. However, in membrane systems that operate under a constant flux, such as full-scale plants that need to meet a daily water demand, fouling is quantified by the additional applied pressure required to maintain water flux. Backwashing and chemical cleaning are often used to reduce fouling in the membranes and are effective in prolonging the life of the membrane; however cleaning cannot regain all of the hydraulic permeability lost to fouling (Nam *et al.*, 2013; Grelot *et al.*, 2010; Rukapan *et al.*, 2015; Ang *et al.*, 2011). To reduce organic fouling in membranes, pre-treatment, such as coagulation and in the case of reverse osmosis membranes, ultrafiltration and

microfiltration, is commonly used (Brehant *et al.*, 2002; Rukapan *et al.*, 2015; Vial and Doussau, 2002; Bonnélye *et al.*, 2008; Lorain *et al.*, 2007; Guastalli *et al.*, 2013).

2.7 Natural Organic Matter and Membrane Fouling

While NOM is well-known as the main driver of organic fouling in pressure-driven membrane systems, TOC and DOC are generally not good predictors of membrane fouling (Shao *et al.*, 2014; Yamamura *et al.*, 2014; Pramanik *et al.*, 2016). Yamamura *et al.* (2014) found that various NOM fractions with the same TOC exhibited different fouling behavior, suggesting that organic fouling is dependent on the character of the NOM, rather than the quantity. Further, there is uncertainty of the relationship between membrane fouling and UV/SUVA. UF membrane experiments show correlations of SUVA and salt rejection (Cho *et al.*, 2000) while MF bench scale experiments did not show correlations between UV_{254} and membrane fouling resistance (Pramanik *et al.*, 2016). Further, Myat *et al.* (2014) used UV_{254} values to track differences in membrane foulants, although Amy (2008) indicates that only low SUVA values are an indicator of high fouling potential. Figure 2.2 provides an illustration of the relationship of various NOM fractions and membrane fouling. The figure shows which fractions have been linked to an increase in fouling based on studies in the published literature.

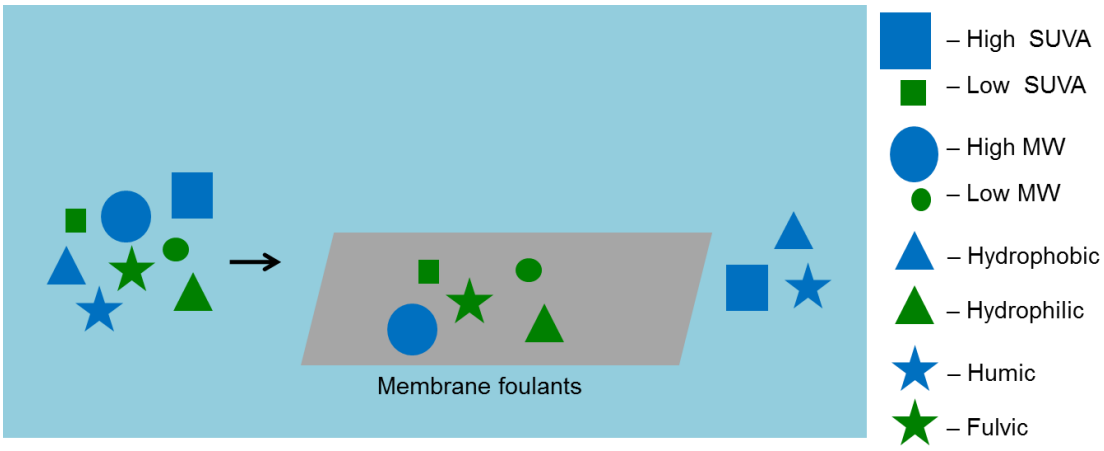

Figure 2.2: Illustration of the common NOM characterizations and subsequent organic fouling patterns found in the literature. According to the literature, less aromatic, hydrophilic, humic, high molecular weight and low molecular weight fractions have been associated with increased fouling in membranes.

Investigating specific NOM fractions, such as hydrophobic/hydrophilic and different molecular weights, provides additional insight into the relationship between NOM and organic fouling. Bench-scale UF and MF experiments show that the hydrophilic fraction fouls membranes more than hydrophobic and transphilic fractions (Yamamura *et al.*, 2014; Kennedy *et al.*, 2005; Gray *et al.*, 2011). Although the ability to reject hydrophilic/hydrophobic NOM fractions is dependent on the hydrophilicity/hydrophobicity of the membrane surface (Shan *et al.*, 2016; Diagne *et al.*, 2012; Zodrow *et al.*, 2009). Howe and Clark (2002) found that smaller particles (colloidal) contributed more to fouling than larger particulate matter (> 0.45 μm) in UF and MF systems. In contrast, other studies have found that larger NOM fractions, such as biopolymer and humics, contributed more to fouling than smaller polymers (Pramanik *et al.*, 2016; Gray *et al.*, 2011). In general, membrane fouling is exacerbated by the "low DBP reactivity" NOM fractions – those with low SUVA and more hydrophilic in nature (Yamamura *et al.*, 2014; Kennedy *et al.*, 2005; Amy, 2008).

In membrane fouling studies, protein-like EEM PARAFAC and PCA components have been identified as predictive of high fouling events (Shao *et al.*, 2014; Chen *et al.*, 2014a). However, in other fouling studies, tryptophan-like and microbial byproduct-like EEM-PARAFAC signals have been correlated with fouling (Yu *et al.*, 2014; Choi *et al.*, 2014). Furthermore, Peiris and colleagues found that "colloidal/particulate matter" EEM-PCA components were correlated with reversible fouling, but that "humic-like" and "protein-like" components were correlated with irreversible fouling (Peiris *et al.*, 2010a; Peiris *et al.*, 2013). Additionally, microbial humic-like and tryptophan-like EEM-PARAFAC components have been found to be associated with organic fouling in membrane bioreactors (Hur *et al.*, 2014). Shao *et al.* (2014) used the humic-like to protein-like component ratios to determine the relative composition of foulants in a membrane system.

2.8 Pre-treatment and mitigation of water treatment challenges

Given that NOM is the cause of many water treatment challenges, including DBP formation and membrane fouling, removal of NOM is critical. Removal of NOM can be effective in mitigating DBP formation and membrane fouling, but should be used strategically since pre-treatment can add signficantly to the cost of clean water and different methods preferentially remove specific NOM fractions (Zhang *et al.*, 2015; Sanchez *et al.*, 2013; Kitis *et al.*, 2001; Pifer and Fairey, 2012; Lavonen *et al.*, 2015; Brehant *et al.*, 2002; Babcock and Singer, 1979; Owen *et al.*, 1995; Peleato *et al.*, 2016). Figure 2.3 shows an illustration of preferential removal for three categories of pre-treatment – coagulation, activated carbon (granular, powder, and biological), and resins (ion exchange and mesoporous adsorbent). Overall, each of the pre-treatments reduce DOC of

the influent water, and subsequently DBP formation and membrane fouling, but each pre-treatment also shows preferential removal of certain NOM fractions, as shown in the illustration.

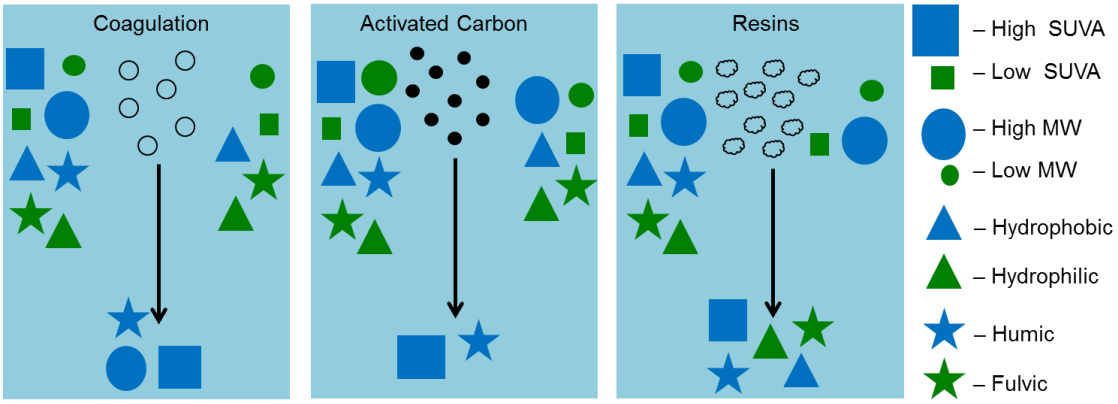

Figure 2.3: Illustration of preferential removal of NOM fractions by various pre-treatments, based on the literature. According to the literature, coagulation preferentially removes aromatic, high molecular weight and humic fractions, activated carbon preferentially removes aromatic and humic fractions, and resins remove aromatic, hydrophobic, hydrophilic, humic and fulvic fractions.

Coagulation, a commonly used surface water treatment for removing particulate matter reduces NOM and DBP formation (Babcock and Singer, 1979; Owen *et al.*, 1995). Alum coagulation preferentially removes high SUVA NOM fractions (Kitis *et al.*, 2001) and larger fractions at pH 6 (Pifer and Fairey, 2012). Coagulation removes some PARAFAC component signals better than others, particularly the humic-like signals associated with DBP formation (Sanchez *et al.*, 2013; Pifer and Fairey, 2012; Lavonen *et al.*, 2015). Coagulation is also effective in removing polysaccharide-like and protein-like NOM that is responsible for membrane fouling (Amy 2008). Further, enhanced coagulation is used by many surface water treatment plants throughout the United States to meet DBP regulations (Archer and Singer, 2006). Following the 1998 release of the Stage 1 Disinfection Byproduct Rule (DBPR), the EPA set "enhanced coagulation" as an NOM removal treatment technique for plants struggling to meet the Maximum Contaminant

Level (MCL) for DBPs (EPA, 1999). Studies show 9 – 73% removal of DOC with enhanced coagulation and improvements in DOC removal when enhanced coagulation is coupled with powder activated carbon (Uyak *et al.*, 2007; Kristiana *et al.*, 2011; Wang *et al.*, 2013). Further, enhanced coagulation treatments show preferential removal of high MW and high UV absorbing compounds (Kristiana *et al.*, 2011; Archer and Singer, 2006; Uyak *et al.*, 2007).

Sorbents, such as activated carbon and anion exchange resins, are also sometimes used in treatment to remove NOM. Like alum coagulation, granular activated carbon (GAC) has shown preferential removal of NOM with high SUVA values (Kitis *et al.*, 2001). Do *et al.* (2015) found that GAC was effective in removing "humic-like" EEM-PARAFAC signals that were correlated to DBP precursors. Powder activated carbon (PAC) has demonstrated superior removal, compared to anion exchange and polymeric resins, in removing protein-like fluorescence signals in NOM that are associated with fouling in UF membranes (Shao *et al.*, 2014). However, Amy (2008) indicated that PAC is overall not very effective in reducing fouling. Biological activated carbon (BAC) was found to effectively remove biopolymers that primarily lead to organic fouling and therefore helps to mitigate fouling; however over time the BAC showed reduced removal of humics (Pramanik *et al.*, 2016).

Additionally, ion exchange is effective in removing NOM, again with preferential removal of certain fractions (Shao *et al.*, 2014; Sanchez *et al.*, 2013; Hsu and Singer, 2010; Jutaporn *et al.*, 2016). Ion exchange, specifically magnetic ion exchange (MIEX) resins, has been suggested as an effective pre-treatment for DBP control because it reduces both DOC and bromide concentrations (Hsu and Singer, 2010), as well as humic-like substances (Bazri *et al.*, 2016).

Ion exchange resins were also found to be effective in removing UV-absorbing NOM fractions, and were equally effective in removing charged hydrophobic and hydrophilic fractions as well as humic and fulvic fractions, but were not as effective in removing large NOM fractions (Bolto *et al.*, 2002; Cornelissen *et al.*, 2008). Mesoporous adsorbent resin (MAR) was found to be more effective in mitigating organic fouling in ultrafiltration membranes than powder activated carbon because MAR removes the NOM fractions that deposit on the membrane surface (foulants) and reduce water permeability (Li *et al.*, 2016).

2.9 Application of Fluorescence NOM Characterization and Future Work

Fluorescence EEMs have been used successfully in many applications requiring advanced characterization of NOM and show promise for implementation in full-scale water treatment systems, such as for online fluorescence detection of influent water. Stedmon *et al.* (2011) found that some EEM-PARAFAC fluorescent components were indicative of microbial contamination in ground water, and therefore fluorescence monitoring of influent water could alert operators to this issue. The use of online fluorescence detection has also been suggested by multiple DBP and membrane fouling studies (Roccaro and Vagliasindi, 2010; Korshin *et al.*, 1997; Shutova *et al.*, 2014; Jutaporn *et al.*, 2016). There has been some success in the development of accurate online fluorescence detectors and in the use of such devices in monitoring for upstream pollutants in a reservoir (Chen *et al.*, 2014b; Liu *et al.*, 2014).

Given the many water treatment challenges associated with NOM, monitoring technologies to mitigate operational challenges are increasingly important. Advanced warning of water

treatment challenges provides operators with a greater opportunity to preemptively mitigate these issues. Applying additional pre-treatment and/or changing operational conditions to address these issues on an as needed basis also allows for a more cost-effective method for delivering safe, clean water to consumers.

Chapter 3

APPLICATION OF CLASSIFICATION TREES FOR PREDICTING DISINFECTION BY-PRODUCT FORMATION TARGETS FROM SOURCE WATER CHARACTERISTICS[1]

3.1 Abstract

Formation and speciation of disinfection by-products (DBPs) depends on source water constituents. Many studies have sought to model the formation of DBPs using both source water and in-plant operational data, and while sometimes highly predictive of DBP formation, these models are limited in their applicability. To create regional models that could apply to multiple plants within a watershed, classification trees were used to predict finished water DBP parameters from source water constituents collected at multiple locations in a watershed. Data were from a field study conducted in the Monongahela River in southwestern, PA from May, 2010 to September, 2012 incorporating six different sites. Classification trees were used to predict violation of, or compliance with, four threshold values that have regulatory and operational significance, namely: the Total Trihalomethanes Maximum Contaminant Level (regulatory standard of 80 μg/L); 80% of the Total Trihalomethanes Maximum Contaminant Level (64 μg/L); a Bromine Incorporation Factor (BIF) of 0.75; and 50% Brominated Trihalomethanes by mass. The classification trees demonstrated accuracies of 76% to 83%. Fluorescence measurements were selected in all classification trees, demonstrating their utility in DBP predictive models. Further, model validation using data from each collection site demonstrated the potential use of classification models across this spatially variable region for

[1] This chapter has been published in Environmental Engineering Science as Bergman, L., Wilson, J., Small, M., VanBriesen, J.M. (**2016**) "Application of classification trees for predicting disinfection by-product formation targets from source water characteristics."

drinking water plants unable to collect their own source water data. Thus, classification trees provide a valuable tool for creating watershed-level source water-based DBP models.

3.2 Introduction

Drinking water disinfection protects consumers from waterborne pathogens; however, it contributes to the formation of harmful disinfection byproducts (DBPs). Disinfection by-products form when natural organic matter (NOM), found in natural waters, is oxidized by disinfectants necessary for control of pathogenic microorganisms. The highly complex and variable NOM present in water poses a challenge for drinking water treatment because the nature of the NOM affects the speciation as well as the extent of DBP formation (Reckhow *et al.*, 1990; Kitis *et al.*, 2002; Liang and Singer, 2003; Singer *et al.*, 2002; Abouleish and Wells, 2015). DBP formation is further complicated by the presence of other ions in the source water (Singer and Chang, 1989), most notably, bromide. Source water bromide leads to increased formation of DBPs, among them brominated DBP species (Richardson *et al.*, 2003; Chowdhury *et al.*, 2010; Watson *et al.*, 2015; Navalon *et al.*, 2008), which are more toxic than the chlorinated forms (Plewa *et al.*, 2002; Richardson *et al.*, 2003; Richardson *et al.*, 2007). DBP exposure, through ingestion of drinking water or inhalation of compounds volatilized during indoor use of disinfected water, has been linked to adverse health effects, such as bladder cancer (King and Marrett, 1996; Kumar *et al.*, 2014; Danileviciute *et al.*, 2012; Villanueva *et al.*, 2004; Cantor *et al.*, 2010). To protect the public health, certain classes of DBPs are regulated by the US Environmental Protection Agency (EPA, 2006).

The high observed variability of DBP formation and speciation in drinking water has been the subject of extensive research. Differences in the type of disinfectant used are responsible for

some of the differences observed in DBP speciation (Mao *et al.*, 2014; Pisarenko *et al.*, 2013; Montesinos and Gallego, 2013; Hua and Reckhow, 2007b; Tian *et al.*, 2013). Additionally, seasonal changes in temperature and chlorine demand, oxidant reaction time, and water residence time within the distribution system, all affect DBP formation (Rodriguez *et al.*, 2007; Rodriguez *et al.*, 2004; Hua and Reckhow, 2012; Chen and Weisel, 1998; Sakai *et al.*, 2015; Allard *et al.*, 2015; Sohn *et al.*, 2006). Furthermore, the variability in NOM, particularly the humic/fulvic content, the aromaticity, and the hydrophobic and hydrophilic fractions, have been linked to variability in DBP formation and speciation (Reckhow *et al.*, 1990; Kitis *et al.*, 2002; Liang and Singer, 2003; Lu *et al.*, 2009; Hua and Reckhow, 2007a; Singer *et al.*, 2002).

Since disinfection byproduct formation and speciation is dependent on the *nature* of the organic matter present in the source water, multiple methods for quantifying and characterizing NOM have been assessed, including: total organic carbon (TOC), dissolved organic carbon (DOC), and ultraviolet absorbance at 254 nm (UV_{254}) (Chen and Westerhoff, 2010; Amy *et al.*, 1987; Harrington *et al.*, 1992; Korn *et al.*, 2002; Sohn *et al.*, 2004; Abouleish and Wells, 2015; Awad *et al.*, 2016; Weishaar *et al.*, 2003). A composite term, $SUVA_{254}$ (UV absorbance normalized by DOC) is frequently used in DBP studies (Edzwald *et al.*, 1985; Kitis *et al.*, 2002; Hua *et al.*, 2015) because it has been shown to be a good indicator of chlorinated DBP formation (Mayer *et al.*, 2015; Li *et al.*, 2014a; Kitis *et al.*, 2001), and in some cases better than TOC in treatment plant operational control (Najm *et al.*, 1994). However, UV_{254} and $SUVA_{254}$ may be less useful for DBP formation and speciation prediction when NOM is of low molecular weight and low aromaticity (Ates *et al.*, 2007; Li *et al.*, 2014a). While $SUVA_{254}$ may be predictive of certain

classes of DBPs, in some datasets, it has also shown weak correlations with trihalomethanes (THM), a commonly observed and regulated class of DBPs (Hua *et al.*, 2015).

Excitation Emission Matrices (EEM) are gaining attention as an improved method for predicting DBP formation because they provide a large amount of data to capture the complexity and heterogeneity of NOM (Pifer and Fairey, 2012; Pifer *et al.*, 2011; Stedmon *et al.*, 2003a; Stedmon and Markager, 2005; Baghoth *et al.*, 2011; Awad *et al.*, 2016). Differential absorbance and fluorescence, as well as differential log-transformed absorbance and fluorescence have shown promise as DBP predictive tools as studies have shown high correlations between these NOM measurements and multiple DBP species (Roccaro *et al.*, 2008; Roccaro *et al.*, 2009; Roccaro and Vagliasindi, 2010; He *et al.*, 2015). To convert EEM for further analysis and use in predictive models, while incorporating all the data obtained from EEM, Parallel Factor Analysis (PARAFAC) is often used because it simplifies large, multi-dimensional data into a few representative components, similar to Principal Component Analysis (Harshman and Lundy, 1994; Stedmon and Markager, 2005; Murphy *et al.*, 2013). Studies have shown promise for the use of EEM-PARAFAC components in predicting DBP formation (Yang *et al.*, 2015a; Pifer and Fairey, 2014; Johnstone *et al.*, 2009; Sakai *et al.*, 2015; Yang *et al.*, 2015b). Further, research by Pifer and Fairey (2012) on EEM coupled with PARAFAC has demonstrated that EEM-PARAFAC components may be better at predicting DBP formation than $SUVA_{254}$. Other research has illustrated the unique ability of EEM-PARAFAC components to differentiate NOM among sources when using sampling from multiple sites (Cabaniss and Shuman, 1987; Sierra *et al.*, 1994; He and Hur, 2015). Pifer and Fairey (2014)'s success in developing strong correlations between EEM-PARAFAC components and DBP formation potential of natural raw

water samples chlorinated and measured in the lab, provides motivation for using similar NOM characterizations for predicting DBP formation in full-scale treatment plants across a watershed.

DBP formation has been modeled mainly using linear regressions (both with untransformed and log transformed variables) that are based on source water characteristics and in-plant operational data (Sadiq and Rodriguez, 2004; Chowdhury, 2009; Ged *et al.*, 2015). The use of in-plant parameters and site specific attributes often limits the applicability of models to different sites or conditions (Ged *et al.*, 2015; Chowdhury, 2009; Westerhoff *et al.*, 2000; Nokes, 1999; Regli *et al.*, 2015). Recently, an extensive literature review and statistical analysis identified few models where the standard errors of the predicted DBP concentrations were less than the maximum contaminant level (MCL) allowable in drinking water (Ged *et al.*, 2015). Thus, while DBP models are useful to understand general trends in the relationships among source water, operational conditions, and DBP formation, they are not particularly useful to a utility in predicting their future compliance state should conditions in the source water change.

A watershed model that provides general predictions of DBP formation and speciation based on source water constituents would be a valuable tool, particularly for plants unable to develop their own site-specific models, and for assessing the impacts of source water changes on multiple drinking water plants within a region. Such wide-spread source water changes might occur due to anthropogenic discharges, such as those observed in the Allegheny River due to oil and gas wastewater discharges (States *et al.*, 2013; Weaver *et al.*, 2015), or due to climate change (Li *et al.*, 2014c). A three year multi-treatment plant field study in the Monongahela River in southwestern Pennsylvania provided source and finished water quality data for the development

of models to assess the utility of extensive organic carbon characterization to predict DBPs under changing conditions. To avoid the use of in-plant data not regularly collected by these utilities and to increase the effectiveness of source water parameters as finished water predictors, multiple NOM characterization techniques were incorporated into the present analysis to more accurately capture the complexity of the NOM as a DBP precursor. Source water constituents alone were used to create decision making models that provide broader, more widely applicable results. Trihalomethanes were the focus of the study because they are the most problematic class of regulated DBPs in the Monongahela River (Handke, 2008).

The goals were (1) to create watershed-level models that broadly define the treatability of the source water, and (2) to provide generalized results so that they are more useful for decision makers (treatment plant operators and regulators) within the region. To make the models useful for decision makers, classification techniques were employed to make predictions of exceedance of four threshold values – the Total Trihalomethanes (TTHM) maximum contaminant level (MCL) of 80 μg/L, 80% of the TTHM MCL (64 μg/L), a Bromine Incorporation Factor (BIF) of 0.75 (corresponding to a 25% molar concentration), and 50% THM brominated by mass. Classification trees were explored in this study because they are easy to interpret and can incorporate multiple trends within a dataset, unlike regression analysis which works when there is a single relationship throughout the dataset. The flexibility of classification trees to incorporate multiple trends is advantageous in a regional watershed model where many different source water constituents exhibit different behaviors. Classification trees have been used successfully to predict specific operational decisions in drinking water treatment plants, such as drinking water advisories (Harvey *et al.*, 2015; Murphy *et al.*, 2016) and coagulant use (Bae *et al.*, 2006).

Additionally, regression trees (used to predict continuous variables) have been used in other DBP formation studies (Trueman *et al.*, 2016) and in broad-scale prediction of multi-national disease burden (Green *et al.*, 2009). Thus, the models described here are designed to enable assessment of how source water variability affects finished water quality and are designed to span a watershed rather than be specific to a single intake location. These techniques can be applied to other regions where anticipated source water changes have the potential to affect finished water DBPs.

3.3 Materials and Methods

Field Site and Sample Analyses

Data for this analysis were from a field study that included six drinking water treatment plants along the Monongahela River in southwestern Pennsylvania (Wilson and Van Briesen, 2013; Wilson, 2013). Samples included in the current analysis (N = 111) span the period May, 2010 to September, 2012, and represent weekly to monthly sampling, depending on season. The six plants, labeled A through F, in order from upstream (southern-most site) to downstream (northern-most site), are shown in Figure 3.1. Two locations were sampled at each of the six plants – from the source water intake in the river and from the finished water leaving the plant after all treatment steps. All plants in the study use chlorine disinfection and two of the plants (Sites C and D) apply chlorine prior to coagulation (pre-chlorination).

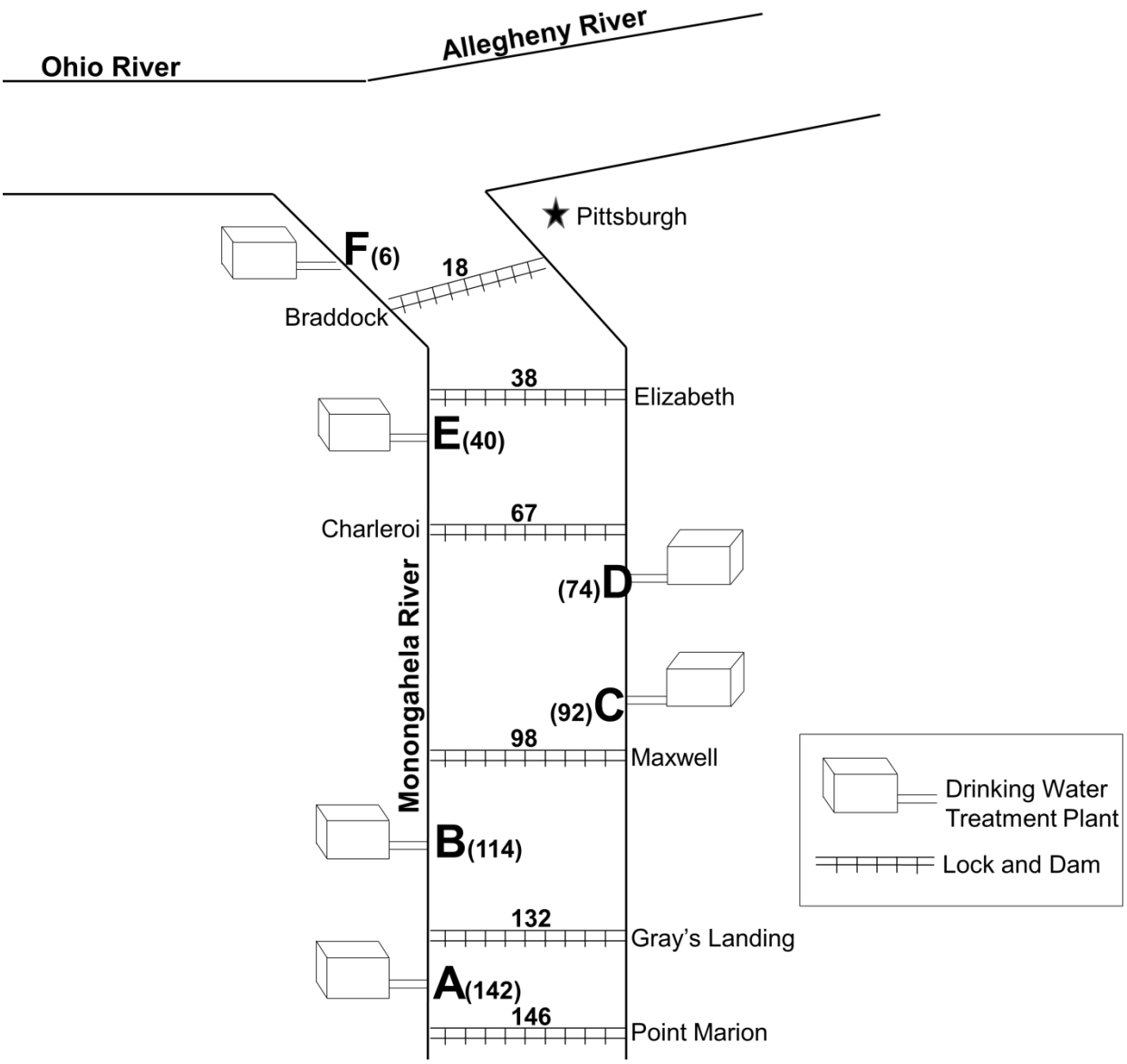

Figure 3.1: Schematic of Monongahela River sampling locations. Schematic shows the bank location of six drinking water plants (A through F), the corresponding locations along the river (in kilometers) upstream of its confluence with the Allegheny River, and locations of lock and dam structures that control river flow.

Source water geochemical data for this field study were previously published (Wilson and Van Briesen, 2013), including concentrations of bromide, chloride, and sulfate. In addition to those data, source water sample analyses included DOC, UV_{254}, and EEM. DOC was measured for samples that were passed through a 0.45 μm filter on a Total Organic Carbon Analyzer (O I

Analytical, College Station, TX) and UV_{254} was measured on a Cary 300 Bio UV Visible Spectrophotometer (Santa Clara, CA). EEM were measured on a Fluoromax-4 Spectrofluorometer (Horiba, Kyoto, Japan). For finished water, the four trihalomethane species (chloroform, bromodichloromethane, dibromochloromethane, and bromoform) were measured using Standard Method 551.1 (EPA, 1995). Missing and below detection data were imputed using log-normal distributions of the known data (Helsel, 1990).

Excitation Emission Matrices and Parallel Factor Analysis

EEM were measured for the 111 samples with the excitation spectra ranging from 200 to 500 nm with a 2 nm step size and with the emission spectra ranging from 300 to 600 nm with a 5 nm step size. A blank sample (MilliQ water measured with the same EEM parameters) was subtracted from each sample EEM to remove the fluorescent signal from water. Any negative values generated in the blank subtraction (mostly from small variations in the water fluorescence) were set to zero. The fluorescence signal was calibrated by converting to Raman units – normalizing all elements in the EEM by the Raman water peak. Specifically, each fluorescence intensity was divided by the integral of the fluorescence intensities under the water peak (EX = 350 nm, EM = 371 – 428 nm) (Lawaetz and Stedmon, 2009). Once all the EEM data were processed, they were analyzed via PARAFAC using the DOMFluor toolbox (http://www.models.life.ku.dk/algorithms) created by Stedmon and Bro (2008). Component data are provided in Table A1 Appendix A.

PARAFAC can be used to simplify large, multi-dimensional datasets by identifying the independent variables responsible for variations in the data (Harshman and Lundy, 1994; Bro, 1997). The advantage of using PARAFAC for an EEM dataset, over other statistical techniques,

is that it can handle multi-dimensional data and produces components that represent real physical phenomena (Stedmon 2003, 2008). PARAFAC uses 3-way decomposition to identify the underlying fluorophores present in multiple EEM samples within the data set. In a simple, dataset with just a few fluorophores, a correct PARAFAC analysis identifies PARAFAC components that represent the individual fluorophores. However, in a more complex mixture, where there are likely many fluorophores, PARAFAC components represent groups of fluorophores with similar fluorescent activity (Stedmon 2003, 2008). Two outliers– Site D on 9/7/2011 and Site A on 6/23/2011 – were identified in the PARAFAC model and removed, leaving 109 instances in the dataset. The validated PARAFAC model produced 3 components, which together sum to the total fluorescence intensity within each sample (Stedmon 2003, 2008). The components generated by the PARAFAC model are representative of the major organic carbon fluorescent groups within the dataset. The three resultant PARAFAC components are referred to as C1, C2, and C3, and the total fluorescence intensity is referred to as F_{max}. The components (C1, C2, C3), the total fluorescence F_{max}, and the ratios of each PARAFAC component to Fmax ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$) are used as model inputs in the study to evaluate both the main fluorescence signals as well as the relative contribution of each fluorescence signal.

Calculating DBP Composite Values

From the experimental data, total trihalomethanes (TTHM) were calculated as the sum of the four individual Trihalomethane species – chloroform ($CHCl_3$), bromodichloromethane ($CHBrCl_2$), dibromochloromethane ($CHBr_2Cl$), and bromoform ($CHBr_3$), each measured as concentrations in μg/L.

Two different methods were used to measure the relative contribution of brominated species to TTHM – Bromine Incorporation Factor (BIF) and percent brominated THM. BIF, a molar-based value, is measured and incorporated in the analysis because source water bromide (and subsequently hypobromous acid) is expected to increase the rate of TTHM formation (Acero *et al.*, 2005; Gallard *et al.*, 2003), thus, increasing the molar total THM present in the finished water. Percent brominated THM by mass is also incorporated because the molar mass of bromide is higher than that of chloride, and thus brominated THMs by virtue of their higher mass increase the likelihood of exceedance of the mass-based TTHM standard by more than would be predicted on a molar basis.

BIF was first developed by Gould *et al.* (1983) and is used frequently to describe the finished water quality, in terms of the DBPs formed (Rathburn, 1996a; Kawamoto and Makihata, 2004; Elshorbagy, 2000; Francis *et al.*, 2010; Tian *et al.*, 2013; Chang *et al.*, 2001). BIF is calculated according to the equation (3.1)

$$BIF = \frac{0*[CHCl_3] + 1*[CHBrCl_2] + 2*[CHBr_2Cl] + 3*[CHBr_3]}{[CHCl_3] + [CHBrCl_2] + [CHBr_2Cl] + [CHBr_3]} \qquad (3.1)$$

where each term represents the molar concentration of the species. BIF can range from 0 (all chloroform) to 3 (all bromoform), with values closer to 3 representing a more brominated TTHM sample. A threshold of 0.75 (25% molar fraction of brominated THMs) was chosen to bisect the data.

Percent brominated THM (shown in equation 3.2) has been used recently to assess the relative contribution of brominated-DBPs to the total regulated TTHM (States *et al.*, 2013).

$$\%\ Brominated = \frac{[CHBrCl_2]+[CHBr_2Cl]+[CHBr_3]}{[CHCl_3]+[CHBrCl_2]+[CHBr_2Cl]+[CHBr_3]} * 100\% \quad (3.2)$$

A threshold of 50% brominated THMs was chosen to bisect the dataset and provide a measure of the relative contribution of Br-THMs to TTHM, by mass.

Statistical Analyses

R (RCoreTeam, 2015), a statistical programming language was used to create regression and classification tree models. Regression models, both with untransformed and log-transformed variables, were used to predict numerical finished water characteristics of interest – TTHM concentration, $CHCl_3$ concentration, $CHBrCl_2$ concentration, $CHBr_2Cl$ concentration, $CHBr_3$ concentration, BIF, and percent brominated TTHM by mass as a function of source water parameters.

A backward step-wise regression was used to choose a subset of variables based on the Akaike Information Criteria (AIC) for both sets of regressions (Akaike, 1974). Regressions using log-transformed variables were tested, in addition to those with untransformed variables, because environmental data are often highly skewed, exhibiting multiplicative, order-of-magnitude relationships, and previous DBP studies have shown success in creating log-transformed predictions (Amy *et al.*, 1987; Rathburn, 1996b; Sohn *et al.*, 2004). Regressions were evaluated based on their adjusted R^2 values and Residual Standard Errors (RSE).

Classification trees are used to classify instances within a dataset by the binary response variable through stratification of the dataset. The data are split for each predictive input variable, with branches chosen sequentially to minimize the misclassification rate in the resulting response variable subsets. The first split is based on the most predictive variable, and subsequent splits are added based on previous or new input variables if these variables are needed to improve the classification according to the response variable. Classification trees are especially useful when the relationship between response and input variables changes over different portions of the input domain, whereas regression models fit a single relationship over an entire domain. Confusion matrices (4x4) and Receiver Operator Characteristic (ROC) curves are used to summarize the overall performance of each classification tree. The confusion matrices show the number of true positives, true negatives, false positives, and false negatives for each tree, which are used to calculate the sensitivity, specificity, and accuracy. The sensitivity (true positive rate), specificity (true negative rate), and accuracy (rate of correctly classified instances) provide an indication of the fit of the model. High sensitivity, specificity, and accuracy values, as well as relatively similar sensitivity and specificity values indicate a good fit and balanced result that minimizes both false positives and false negatives. ROC curves show the trend of true positives (sensitivity) to false positives (1 – specificity). A greater the area under the curve (AUC), obtained from an ROC curve that approaches the top left corner of the plot more closely, indicates a more predictive model. The decision trees and ROC curves were created in R using the Rpart and ROCR packages (RCoreTeam, 2015; Chambers and Hastie, 1992; Sing *et al.*, 2005). The decision trees were pruned using a minimum split of 25 (i.e. at least 25 observations

must be present in a node, otherwise any further downstream branches are pruned), and validated using a 10-fold cross validation, with instances randomly partitioned into each of the 10 subsets.

Table 3.1: Summary of variables used in regression and classification models. Measured source water parameters are used as input variables. Measured finished water parameters serve as the basis for regression and classification model response variables. Threshold values are used to create binary response variables for classification models.

Source Water	Finished Water	Threshold Values
Br (mg/L) DOC (mg/L) UV_{254} (cm^{-1}) C1 C2 C3 F_{max}	Total Trihalomethanes (μg/L) – TTHM Chloroform (μg/L) – $CHCl_3$ Bromodichloromethane (μg/L) – $CHBrCl_2$ Dibromochloromethane (μg/L) – $CHBr_2Cl$ Bromoform (μg/L) – $CHBr_3$	TTHM MCL (80 μg/L) 80% TTHM MCL (64 μg/L) BIF of 0.75 (25% Br-THM by mol) 50% Brominated THM (by mass)

A summary of the variables used in the regression and classification models is presented in Table 3.1. While fluorescence is not usually routinely monitored by plant operators, new research supporting online fluorescence monitoring of NOM may encourage future implementation of such technology by treatment plants (Roccaro *et al.*, 2009; Roccaro and Vagliasindi, 2010; Shutova *et al.*, 2014). The four binary response variables were chosen because they provide important information about the quality of the water and can be used by operators and regulators to make decisions. The TTHM MCL is a threshold value that regulators have set as an allowable limit of TTHM concentration in drinking water at the point of consumption (EPA, 2006). As an enforceable regulation, operators must manage treatment plant operations so as to not exceed the TTHM MCL at all points in the water distribution system. Eighty percent of the TTHM MCL,

corresponding to a concentration of 64 μg/L, was also chosen as a threshold value because it is commonly used as a target for finished water TTHM in the plant to maintain regulatory compliance throughout the system (Roberson *et al.*, 1995; Becker *et al.*, 2013). BIF and percent brominated THM indicate the relative presence of brominated THM species, which may represent more significant health concerns (Plewa *et al.*, 2002; Richardson *et al.*, 2003). The threshold values of BIF and percent brominated were set to represent a moderate distribution of brominated THMs. BIF usually stays below 0.3 (on a 0 to 3 scale) in the Mississippi, Missouri, and Ohio Rivers (Rathburn, 1996a).

3.4 Results and Discussion

Variability of Finished Water Trihalomethanes

TTHM were measured in the finished water at each of the six drinking water treatment plants. The boxplots in Figure 3.2 show the range of TTHM levels at each of the six sampling locations in the Monongahela River.

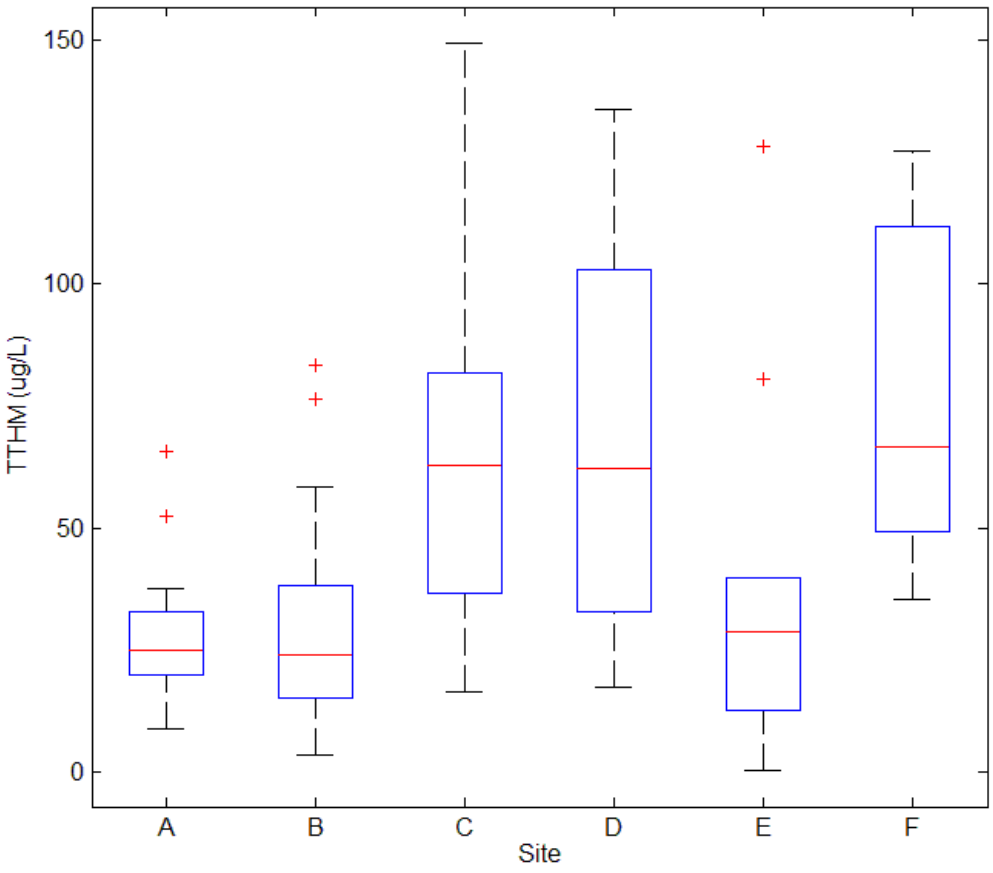

Figure 3.2: Boxplots of TTHM (µg/L) at each of the six sampling sites. Plots show median values, 75th and 25th quartiles (upper and lower ends of the box), minimum and maximum (non-outlier) values (ends of whiskers), and outliers (+ signs).

Differences among sites are statistically significant (ANOVA test p-value of 1.05×10^{-36}). Post-hoc t-tests indicate significant ($p < 0.05$) differences between all site pairs except Sites C and D and Sites A and B. Sites C, D, and F have higher median levels of TTHMs as well as a larger ranges of TTHM levels. The high variability in the river across many sites is not surprising, especially since the river is navigationally-controlled by a series of locks and dams that create pools, which can show significant variation in source water quality (Wang *et al.*, 2015).

Variation in TTHM at different sites has been widely reported in prior work (Obolensky and Singer, 2005; Obolensky and Singer, 2008; Francis *et al.*, 2009). Sites C and D have some of the highest TTHM levels, as would be expected since these sites apply chlorine ahead of the coagulation and filtration steps. The TTHM levels in Sites C and D may also be similar because they are in the same pool of the river (see Figure 1), making their source water quality likely more similar to each other.

Variability of Bromide in the Source Water

The presumed consistency of the single river source was a primary reason for selection of the field study sites at multiple plants using similar processes and all using free chlorine for disinfection. As discussed previously, bromide is an important source water component to consider because bromide in the source water leads to more brominated DBPs (Richardson *et al.*, 2003; Chowdhury *et al.*, 2010; Watson *et al.*, 2015; Plewa *et al.*, 2002). Bromide was expected to be fairly consistent across the six sites throughout the three-year field study; however, as reported by Wilson and Van Briesen (2013), significant changes in bromide concentration were observed during 2011-2013 in this river.

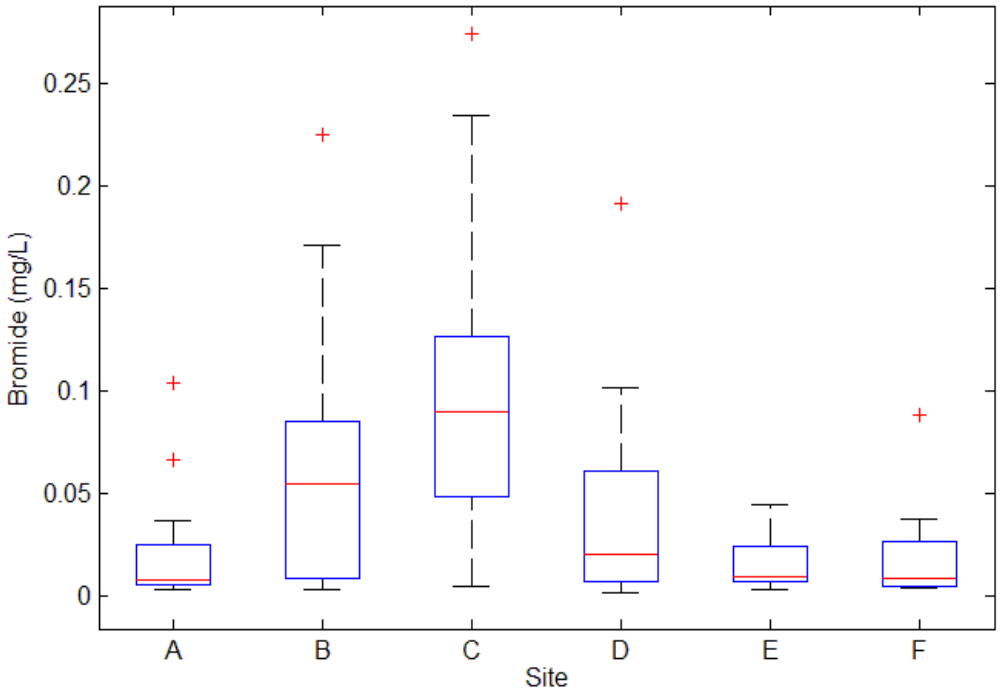

Figure 3.3: Boxplots of source water bromide concentration (mg/L) at each of the six sampling sites along the Monongahela River. Plots show median values, 75th and 25th quartiles (upper and lower ends of the box), most extreme non-outlier values (ends of whiskers, and outliers (+ signs).

In addition to temporal variation, bromide in the river also shows spatial variation. Figure 3.3 shows a high level of variability of bromide across the six sampling locations (ANOVA test p-value of 2.9×10^{-5}). The high variability of the bromide suggests that it is a potential cause of the high variability in the finished water TTHM, compounding the challenge in assessing the role of NOM characterization in TTHM prediction. Although bromide is a known DBP precursor and plays an important role in DBP formation, bromide and TTHM levels across all sites demonstrate a poor linear relationship, with an R value of 0.06. This is consistent with many prior studies that report bromide concentration alone is not predictive of finished water DBP

concentrations (Sakai *et al.*, 2015; Chowdhury *et al.*, 2010; Al-Omari *et al.*, 2004; Kulkarni and Chellam, 2010).

Variability in Organic Source Water Characteristics

Organic precursors were analyzed, using commonly measured criteria, including DOC, UV_{254}, as well as through fluorescence EEM, which were analyzed using PARAFAC analysis. Boxplots of DOC and UV_{254} throughout the 3-year study at each of the six plants can be found in Figure A1 in Appendix A. In general, DOC is very stable across the sites. UV_{254} appears to be slightly more variable, but an ANOVA test indicates that mean UV_{254} values are not significantly different across sites (p-value = 0.22). NOM is a well-known precursor for DBP formation, and UV_{254} and DOC are often included in DBP prediction models (Edzwald *et al.*, 1985; Reckhow *et al.*, 1990; Kitis *et al.*, 2002). However, these parameters are not correlated with TTHM in this data set (R=0.12 for DOC, 0.08 for UV_{254})). While DOC and UV_{254} provide some insight into organic carbon, their stability across multiple sites and seasons suggests these parameters are not providing enough information about variability to account for variability in observed TTHM in finished water in the plants.

Table 3.2: Fluorescence maxima (emission and excitation) for the three PARAFAC components - C1, C2, and C3.

Component	Emission Maxima (nm)	Excitation Maxima (nm)
C1	440	346
C2	385	314
C3	495	394

The EEM-PARAFAC analysis of the 109 sample EEM yielded 3 components, C1, C2, and C3. Fluorescence maxima for the three components are shown in Table 3.2. All three components are found in the humic acid-like region, according to Chen *et al.* (2003). Further, Sakai *et al.* (2015) found that EEM with fluorescence signals in the "humic acid-like" region are highly correlated with TTHM formation. The three plots in Figure 3.4 provide visual representations of the resultant PARAFAC components. Prior to considering the components as input modeling variables, their stability across sites was evaluated. Boxplots that illustrate the variability of the PARAFAC components and total fluorescence intensity, F_{max}, at each of the six sites throughout the three-year study can be found in Figure A2 in Appendix A.

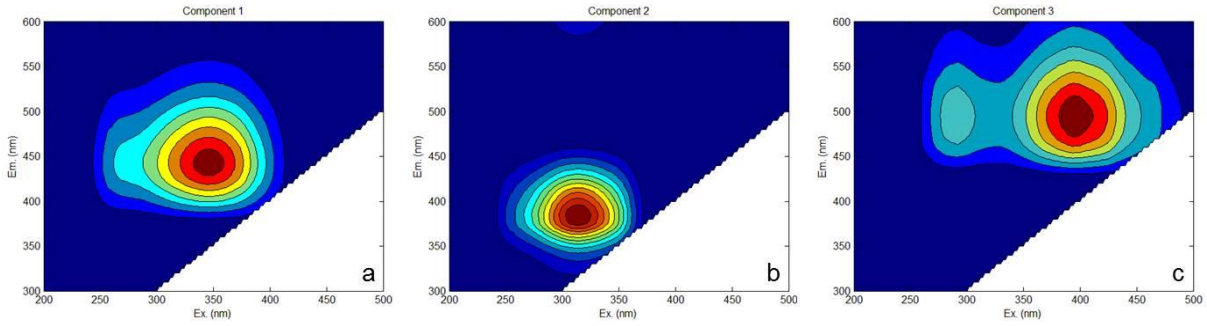

Figure 3.4: EEM of 3 Components resulting from the EEM-PARAFAC analysis as follows: (a) C1, (b) C2, and (c) C3.

The four fluorescence characterizations – C1, C2, C3, and F_{max} – show some similar patterns at multiple sites. For example, Sites A and F and Sites D and E show similar central tendencies for each of the four fluorescence parameters. Overall, there is high variability in component values and F_{max} across the six sites, which is confirmed by ANOVA tests for each of the four fluorescence characterizations. ANOVA tests for C1, C2, C3, and F_{max} across the sites produced significant p-values, 0.04, 0.003, 0.01, and 0.01, respectively. Although PARAFAC components

show promise as DBP predictive parameters individually, they demonstrate poor linear fits with TTHM (R^2 values of 0.10, 0.14, 0.07, and 0.11 for C1, C2, C3, and F_{max}, respectively). Previous work by (Pifer and Fairey, 2014) indicated high correlations between PARAFAC components and TTHM formation potential measured in the lab; however, direct prediction from a single component or F_{max} was not successful with these field samples.

Regression Analysis

Source water constituents (i.e., NOM and bromide) are expected to influence DBP formation, and thus, have the potential to predict concentrations of THM species. In the present work, the utility of expanded NOM characterization along with bromide to predict THMs was examined. Operational characteristics were specifically excluded from modeling to ascertain if models could be developed to account for source water variability throughout the region, independent of plant-specific operational characteristics.

Linear regressions were first developed for seven different response variables – TTHM, Chloroform ($CHCl_3$), Bromodichloromethane ($CHBrCl_2$), Dibromochloromethane ($CHBr_2Cl$), Bromoform ($CHBr_3$), BIF, and Percent Brominated – using multiple input variables, including bromide, DOC, UV_{254}, and EEM-PARAFAC components. The untransformed and log-transformed variable regression models were statistically significant (F statistic p-value < 0.05), but showed poor to moderate R^2 values, ranging from 0.07 to 0.44 for untransformed variable regressions and 0.10 to 0.28 for the log transformed variable regressions. Complete results and further discussion are presented in Appendix A.

Classification Trees

Classification trees were used to predict whether four key threshold values related to finished water DBPs – the TTHM MCL, 80% of the MCL, a BIF of 0.75, and 50% Brominated THMs by mass – would be met. Two classification trees were created for each of the four binary response variables (based on the four threshold values) – one incorporating the three PARAFAC components (C1, C2, C3) and one incorporating the ratios of each PARAFAC component to the total fluorescence intensity ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$) as well as the total fluorescence intensity, F_{max}. ROC curves for all 8 classification trees are shown in Figure 3.5. Figure 3.5a shows the ROC curves for the two THM threshold trees (TTHM MCL and 80% of the TTHM MCL) and Figure 3.5b shows the ROC curves for the two brominated threshold trees (0.75 BIF and 50% Br-THM).

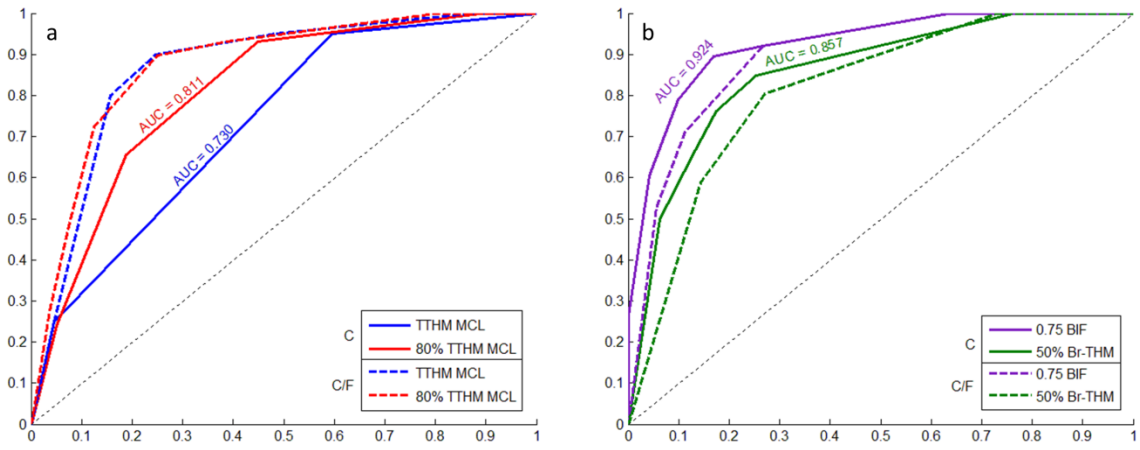

Figure 3.5: Plot of Receiver Operator Characteristic (ROC) Curves for the classification trees. The TTHM MCL and 80% TTHM MCL (64 µg/L) trees are shown in (a) and the 0.75 BIF and 50% Br-THM trees are shown in (b). The ROC curves for the component trees (C) are drawn in solid lines and the ROC curves for the component ratio (C/F) trees are drawn in dashed lines. Each response variable is designated by a different color, as shown in the legend. The dotted black line at Y = X shows a curve based on a random selection. AUC values are shown for the component trees in each plot.

The plots in Figure 3.5a show that incorporating component fractions provides stronger predictions than components for the two THM thresholds, and that when incorporating components, a better prediction is obtained for 80% of the TTHM MCL (64 μg/L) than for TTHM MCL. The plots in Figure 3.5b show that incorporating components provides a stronger prediction than with component fractions for the two brominated thresholds, and that a better prediction is obtained for 0.75 BIF than for 50% Br-THM. Overall, the 0.75 BIF component tree provides the strongest predictions of all eight trees, while the TTHM MCL component tree provides the weakest predictions.

Table 3.3: Summary of Classification Tree Performance. The Table shows the AUC (area under the ROC curve) value, accuracy, sensitivity, and specificity for the classification trees that use components (C1, C2, C3) as fluorescence inputs and for the classification trees that use component ratios and total fluorescence ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}) as fluorescence inputs for all 4 response variables – TTHM MCL, 80% of the TTHM MCL, BIF of 0.75, and 50% Brominated THM.

	COMPONENTS				COMPONENT RATIOS			
Response Var.	**AUC**	**Acc.**	**Sens.**	**Spec.**	**AUC**	**Acc.**	**Sens.**	**Spec.**
TTHM MCL	0.730	0.83	0.25	0.96	0.867	0.83	0.8	0.84
80% MCL	0.811	0.77	0.66	0.81	0.875	0.83	0.72	0.88
0.75 BIF	0.924	0.83	0.61	0.96	0.894	0.8	0.53	0.94
50% Br-THM	0.857	0.8	0.76	0.83	0.815	0.76	0.8	0.73

A summary of the performance of all eight classification trees – component and component ratio trees for predicting exceedance of each of the four threshold values – is shown in Table 3.3. The AUC values range from 0.730 to 0.924 and the accuracy values range from 0.76 to 0.83. Most of the trees have high and fairly similar sensitivity and specificity values (except for the component

TTHM MCL tree and the component ratio 0.75 BIF tree), which means that the trees provide fairly balanced results. To evaluate the added value of fluorescence measurements, AUC values were determined for trees without fluorescence measurements. Based solely on DOC, UV_{254}, and bromide, AUC values are 0.60 for TTHM MCL, 0.561 for 80% TTHM MCL, 0.893 for 0.75 BIF, and 0.759 for 50% Br-THM. All of these additional trees used the same minimum split as the 8 classification trees incorporating the fluorescence measurements (25), except for the TTHM MCL tree which used a minimum split of 15 because a tree could not be created beyond a single node at a larger minimum split. The AUC values for trees without fluorescence measurements are overall worse than those for trees that incorporate fluorescence measurements, except for the 0.75 BIF, which gave similar results both with and without fluorescence measurements (AUC = 0.894 for the component ratio tree and AUC = 0.893 for the tree that omits fluorescence variables). These results indicate that in general, fluorescence measurements improve classification tree predictions.

Predicting TTHM Concentrations in Excess of the Maximum Contaminant level (MCL). The classification trees that predict exceedance of the TTHM MCL Regulation (TTHM concentration of 80 μg/L) are shown in Figure 3.6 – Figure 3.6a is the tree that uses components as inputs (C1, C2, C3) and Figure 3.6b is the tree that uses component ratios and total fluorescence ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}) as inputs.

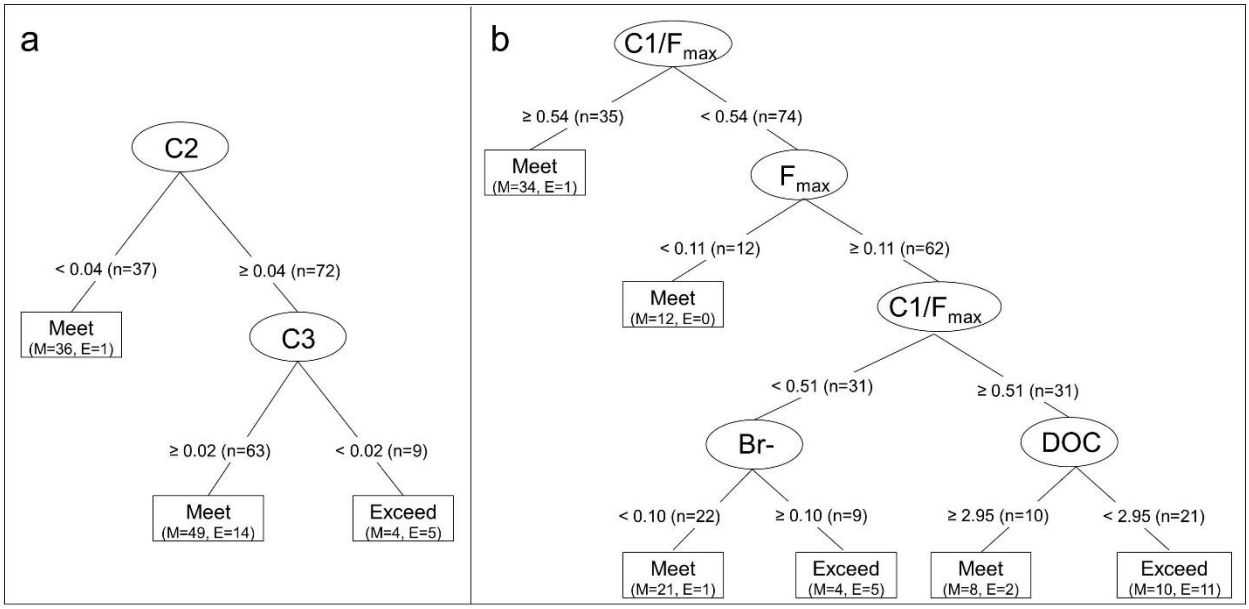

Figure 3.6: Classification Trees created in R that predict whether the TTHM MCL Threshold is exceeded based on source water characteristics, including bromide, DOC, UV_{254}, and component sub-groups: (a) the three PARAFAC components (C1, C2, C3); and (b) the component ratios and total fluorescence intensity ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}). The input parameters are drawn in ovals and the terminal nodes (indicating whether the TTHM MCL will be met or exceeded) are drawn in rectangles. Branches are labeled with the split of the input parameters and the number of instances (n) pertaining to the split. Terminal nodes are labeled with the overall outcome ("Meet" or "Exceed") and the number of instances that actually meet (M) or exceed (E) the threshold.

Classification trees provide good fits of the dataset, as demonstrated by the high accuracy values and generally high sensitivity and specificity values. Though the two trees performed similarly in accurately classifying instances, the component ratio tree (Figure 3.6b) is more balanced in its classified outcomes, with nearly equal sensitivity and specificity values. The component tree (Figure 3.6a), on the other hand has a very high specificity (true negative rate) and very low sensitivity (true positive rate) because the tree slightly under-predicts exceeding the MCL, according to Table 3.3. The component classification tree classified very few instances as "exceed," only 9 out of 109, though in reality 20 instances exceeded the MCL.

The classification tree that uses components as inputs identifies C2 and C3 as the most important variables in predicting TTHM MCL exceedance, with C2 being the dominant input variable. According to the tree, instances with low C2 values (< 0.04) are likely to meet the TTHM MCL. Outcomes for instances with high C2 values (≥ 0.04) depend on C3 values. Instances with high C2 values *and* high C3 values (≥ 0.02) are likely to meet the MCL, while instances with high C2 values and *low* C3 values (< 0.02) are likely to exceed the MCL. The classification tree that uses component ratios and total fluorescence intensity as inputs identifies $C1/F_{max}$, F_{max}, bromide concentration, and DOC as the most important variables, with $C1/F_{max}$ being the dominant input variable. According to the tree, when the $C1/F_{max}$ ratio is high (≥ 0.54), instances are likely to meet the TTHM MCL. At lower $C1/F_{max}$ values (< 0.54), F_{max} is used to determine the outcome. Low $C1/F_{max}$ and low F_{max} values (F_{max} < 0.11) generally meet the MCL. Instances are more likely to exceed the MCL when $C1/F_{max}$ is low, F_{max} is high, and bromide concentration is high (≥ 0.10), or when Cl/F_{max} values are moderate (0.51 – 0.54), F_{max} is high, and DOC is low (< 2.95).

A major difference between the two trees is the set of input variables included in each tree. The component classification tree incorporates only two fluorescence measurements (C2 and C3), while the component ratio classification tree incorporates two fluorescence measurements ($C1/F_{max}$ and F_{max}), DOC, and bromide concentration. Despite these differences, both trees show a preference for fluorescence NOM measurements over DOC and UV_{254}, based on order of appearance in the tree and overall inclusion in the tree. Fluorescence measurements have also been found to be superior to SUVA in other studies when DOC is low (Lavonen *et al.*, 2015). The inclusion of bromide in only one tree and at the bottom of the tree indicates that NOM

characterization is more important than bromide concentration in predicting TTHM regulatory outcomes in this system, despite significant variability of bromide in the source water. The behavior of TTHM formation due to bromide concentration (increased likelihood of exceeding the MCL at higher bromide concentrations) is consistent with previous studies that found that increases in bromide concentration result in increased TTHM (Chowdhury *et al.*, 2010; Hua *et al.*, 2006b; Navalon *et al.*, 2008).

Predicting TTHM in excess of 80% of the Maximum Contaminant Level (MCL). The classification trees that predict exceedance of 80% of the TTHM MCL (64 μg/L) are shown in Figure 3.7 – Figure 3.7a illustrates the component classification tree (incorporating C1, C2, C3) and Figure 3.7b illustrates the component ratio tree (incorporating $C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, and F_{max}).

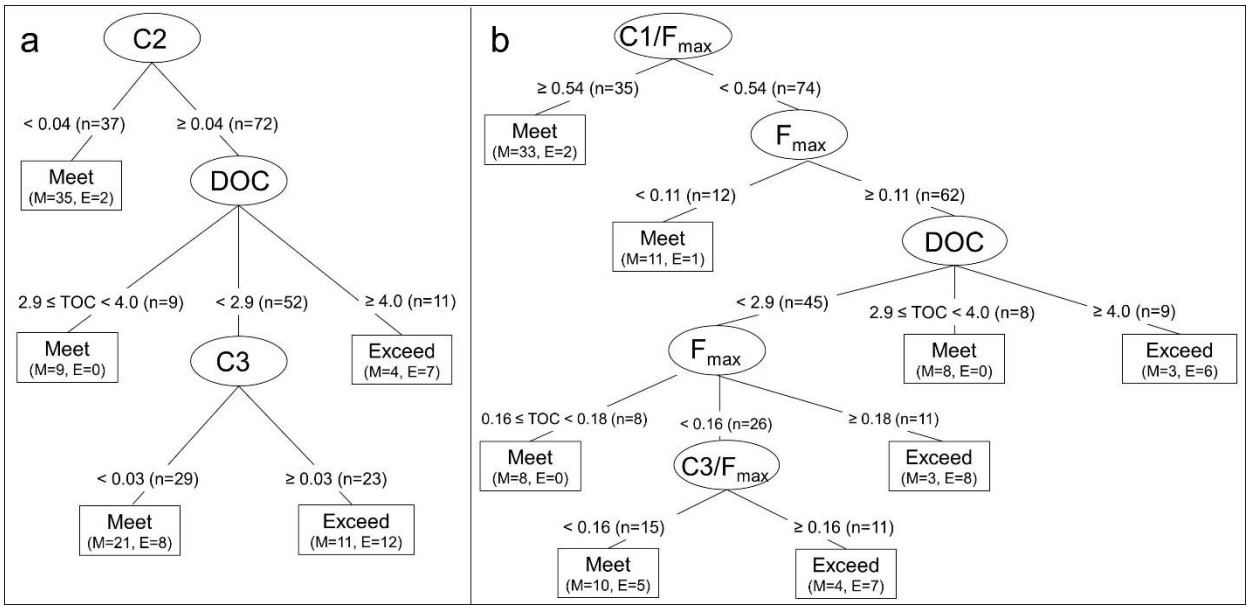

Figure 3.7: Classification Trees created in R that predict whether the 80% of the TTHM MCL (64 μg/L) is exceeded based on source water characteristics, including bromide, DOC, UV_{254}, and component sub-groups: (a) the three PARAFAC components (C1, C2, C3); and (b) the component ratios and total fluorescence intensity ($C1/F_{max}$, $C2/F_{max}$,

C3/F_{max}, F_{max}). The input parameters are drawn in ovals and the terminal nodes (indicating whether the TTHM MCL will be met or exceeded) are drawn in rectangles. Branches are labeled with the split of the input parameters and the number of instances (n) pertaining to the split. Terminal nodes are labeled with the overall outcome ("Meet" or "Exceed") and the number of instances that actually meet (M) or exceed (E) the threshold.

The 80% MCL (64 μg/L) classification trees look similar to the TTHM MCL trees in that most of the same input variables were used. Both of the component trees incorporate C2 and C3 and the C2 split occurs at the same cut-off value, however, the 80% MCL tree also incorporates DOC. Both component ratio trees incorporate C1/F_{max}, F_{max}, and DOC, and the C1/F_{max} and first F_{max} splits occur at the same cut-off values, however, the TTHM MCL tree incorporates bromide while the 80% MCL ratio tree incorporates C3/F_{max}. Of the four classification trees related to the regulatory TTHM MCL threshold (Figures 3.6a, 3.6b, 3.7a, 3.7b), only one incorporates bromide, indicating that it is not as important as NOM characterization in determining whether or not the regulatory thresholds will be met. Though bromide has been found to increase DBP formation, many of the studies that report bromide being an important precursor in DBP formation incorporate synthetic laboratory samples that have higher concentrations of bromide than those found in these natural waters (Richardson *et al.*, 2003; Chowdhury *et al.*, 2010; Watson *et al.*, 2015; Hua *et al.*, 2006b; Navalon *et al.*, 2008; Hua and Reckhow, 2012; Chang *et al.*, 2001). Additional discussion of the 80% TTHM MCL classification tree is found in Appendix A.

Predicting BIF Values in Excess of 0.75. The classification trees that predict exceedance of the 0.75 BIF threshold are shown in Figure 3.8 – Figure 3.8a illustrates the component classification

tree (incorporating C1, C2, C3) and Figure 3.8b illustrates the component ratio tree (incorporating $C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, and F_{max}).

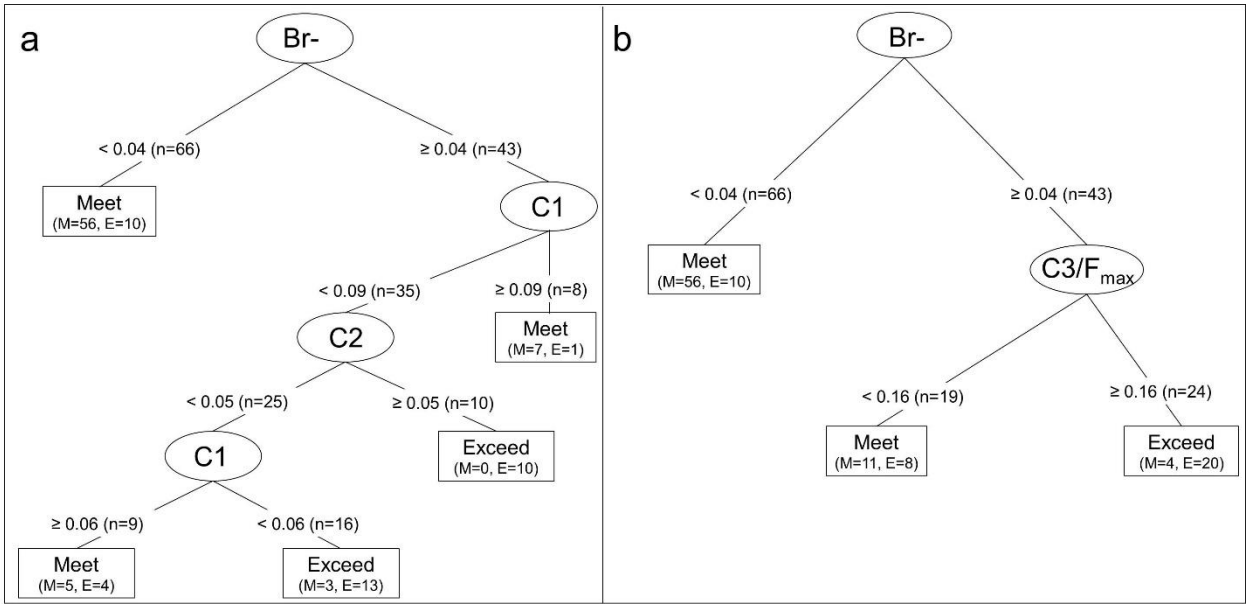

Figure 3.8: Classification Trees created in R that predict whether the 0.75 BIF (25% molar bromination) threshold is exceeded based on source water characteristics, including bromide, DOC, UV_{254}, and component sub-groups: (a) the three PARAFAC components (C1, C2, C3); and (b) the component ratios and total fluorescence intensity ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}). The input parameters are drawn in ovals and the terminal nodes (indicating whether the TTHM MCL will be met or exceeded) are drawn in rectangles. Branches are labeled with the split of the input parameters and the number of instances (n) pertaining to the split. Terminal nodes are labeled with the overall outcome ("Meet" or "Exceed") and the number of instances that actually meet (M) or exceed (E) the threshold.

The component classification tree (Figure 3.8a) identifies bromide concentration, C1, and C2 as the most important variables, while the component ratio classification tree (Figure 3.8b) identifies bromide and $C3/F_{max}$ as the most important variables. In both classification trees, bromide is the first variable, meaning that it is the most indicative of the outcome behavior – exceeding or meeting the 0.75 BIF threshold. The inclusion of bromide as the dominant variable in both classification trees is consistent with previous research that found that bromide in the source water contributes to increased BIF in finished water (Rathburn, 1996a).

Predicting THM Bromination in Excess of 50%. The classification trees that predict exceedance of 50% brominated THM (by mass) are shown in Figure 3.9. The component classification tree is illustrated in Figure 3.9a and the component ratio classification tree is illustrated in Figure 3.9b.

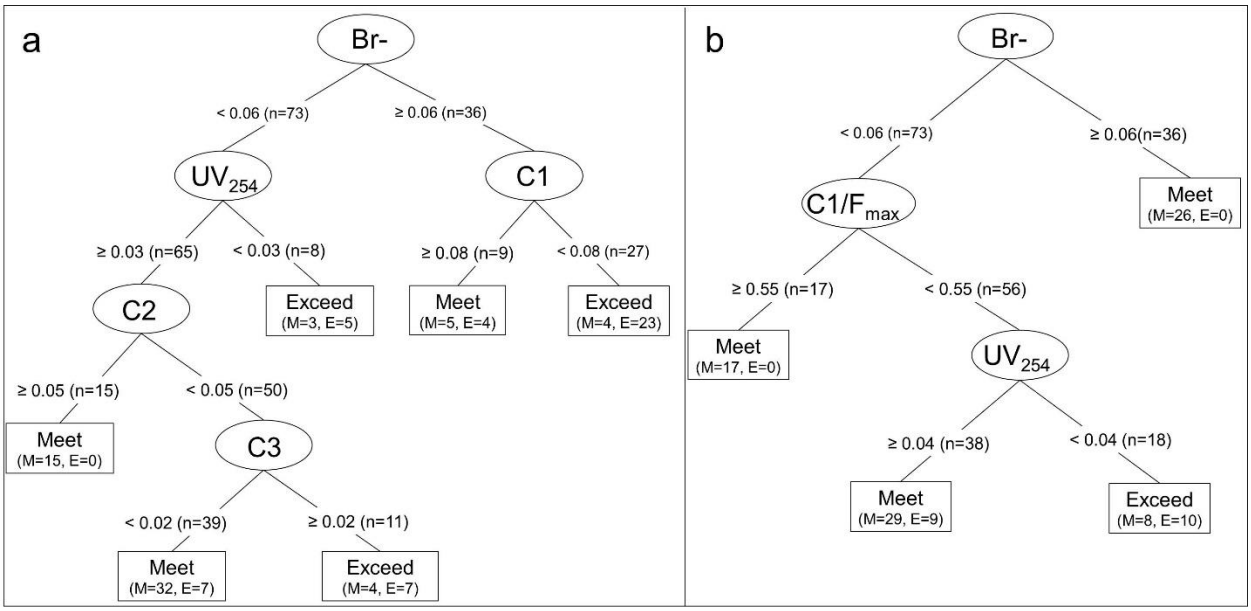

Figure 3.9: Classification Trees created in R that predict whether the 50% Brominated THM (by mass) threshold is exceeded based on source water characteristics, including bromide, DOC, UV_{254}, and component sub-groups: (a) the three PARAFAC components (C1, C2, C3); and (b) the component ratios and total fluorescence intensity ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}). The input parameters are drawn in ovals and the terminal nodes (indicating whether the TTHM MCL will be met or exceeded) are drawn in rectangles. Branches are labeled with the split of the input parameters and the number of instances (n) pertaining to the split. Terminal nodes are labeled with the overall outcome ("Meet" or "Exceed") and the number of instances that actually meet (M) or exceed (E) the threshold.

The component classification tree identifies bromide, UV_{254}, C1, C2, and C3 as the most important input variables, and the component ratio classification tree identifies bromide, UV_{254}, and $C1/F_{max}$ as the most important input variables for predicting whether the 50% brominated THM by mass threshold will be exceeded. The results indicate that exceedance of the 50%

brominated THM threshold is dependent on both bromide and NOM characterization, with bromide being the most important. Further, DOC is not included in either tree, indicating that the characterization of NOM is more important than the quantity in brominated THM formation (by mass), like the 0.75 BIF classification tree results. Both 50% Br-THM classification trees show unexpected results– in three of the four the exceedance scenarios contain lower bromide levels (< 60 μg/L). It was expected that exceedances would more often occur in the high bromide branches of the trees (≥ 60 μg/L) because higher bromide shifts DBP towards brominated species (Richardson *et al.*, 2003; Watson *et al.*, 2015; Chang *et al.*, 2001). However, the unexpected results may be due to a more complex relationship between bromide and NOM in DBP formation.

The inclusion of fluorescence measurements in all 8 classification trees, in addition to the higher AUC values for trees that include fluorescence measurements, demonstrates that fluorescence measurements are valuable parameters when classifying instances based on exceeding or meeting TTHM or Br-THM thresholds. All four component trees (Figures 3.6a, 3.7a, 3.8a, 3.9a) include C2 and at least one other component (C1 or C3). In the TTHM component trees (Figures 3.6a and 3.7a), C2 is the most important input variable. C2 has a similar peak to one of the two peaks in a PARAFAC component identified in another study (EM/EX = 381/219(304)), which was found to be highly correlated with chloroform formation in a multivariate linear regression (Johnstone *et al.*, 2009). In the present study, chloroform is the dominant THM species. Three of the four component ratio trees (Figures 3.6b, 3.7b, and 3.9b) include $C1/F_{max}$, and in all three of the trees, higher $C1/F_{max}$ ratios (≥ 0.54) increase the likelihood of meeting the threshold. Finally, seven of the eight classification trees identify more than one NOM measurement as important

input variables. The use of multiple NOM characterizations within the classification trees demonstrates the need for multiple NOM characterization techniques for effectively capturing the complexity and heterogeneity of NOM for predictive models.

Model Validation Across Sites

To further evaluate the robustness of the classification trees across a spatially variable dataset, additional classification trees were created on subsets of sites. The additional models, referred to as Site Validations (SV), were performed by creating models based on 5 of the 6 sites (training dataset) and then tested on the one remaining site (testing dataset). Successful model generation from the site validations would suggest that a model created from multiple sites within a specific geographic region (such as the dataset used in this study) could be applied to other sites within the region that were not originally incorporated into the model. Table 3.4 presents a summary of the accuracy values within the testing dataset for the classification tree Site Validation models that use the components (C1, C2, C3) as inputs. Also contained in the summary are accuracy values for the models presented previously that were generated on the entire dataset (referred to as "initial").

Table 3.4: Summary of Accuracy Results for the Site Validation Classification Trees using components (C1, C2, C3). Results are shown for the initial models (Initial) and the six site validation (SV) models for each of the four response parameters.

Model	TTHM MCL	80% MCL	0.75 BIF	50% Brominated
Initial	0.83	0.77	0.83	0.80
SV 1	0.75	0.75	0.81	0.50
SV 2	0.82	0.74	0.82	0.65
SV 3	0.68	0.53	0.26	0.63
SV 4	0.60	0.50	0.75	0.70
SV 5	0.70	0.80	0.60	0.50
SV 6	0.50	0.20	0.70	0.80

Overall, the Site Validation models in Table 3.4 show fairly high accuracy results. Except for 80% MCL SV 6 and 0.75 BIF SV 3 models, the accuracy values for the SV models are 0.50 or higher. Each of the four parameters have at least three Site Validation models that correctly classify 65% or more of the test instances.

The same site cross validations were performed for the classification tree models that used the component ratios and total fluorescence ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}) as inputs. A summary of the results from these Site Validation Classification Models is presented in Table 3.5.

Table 3.5: Summary of Accuracy Results for the Site Validation Classification Trees using component ratios and total fluorescence ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}). Results are shown for the initial models (Initial) and the six site validation (SV) models for each of the four response parameters.

Model	TTHM MCL	80% MCL	0.75 BIF	50% Brominated
Initial	0.83	0.83	0.80	0.76
SV 1	0.56	0.63	0.81	0.44
SV 2	0.88	0.74	0.76	0.65
SV 3	0.68	0.53	0.32	0.32
SV 4	0.65	0.60	0.70	0.65
SV 5	0.80	0.80	0.50	0.50
SV 6	0.40	0.50	0.80	0.80

The Site Validation models in Table 3.5 also show fairly high accuracy results. With the exception of TTHM MCL SV 6, 0.75 BIF SV 3, 50% Brominated SV 1, and 50% Brominated SV 3, the accuracy results for the SV models are 0.50 or higher. Furthermore, each of the four parameters have at least two SV models that correctly classify 65% or more of the test instances. In general, the Site Validation models show lower accuracy values than the initial models because they are developed and tested on a subset of the data.

The site validation models demonstrate a reasonable level of accuracy; many of the site validations have accuracy values comparable to the initial models. Given that these models are fairly predictive across sites, there is potential for use of the models for other sites in the geographic region that were not originally included in the analysis. Additionally, this suggests the general method may provide insights in other geographic regions. Creating a classification model using data from multiple sites in a region may enable application at other drinking water facilities throughout that region.

3.5 Conclusions

Classification techniques demonstrate an improvement in predictive capability compared to regression models for predicting finished water quality based on source water characteristics alone for the dataset used in this study, with 76% to 83% accuracy in correctly classifying instances. The classification trees are able to partition the input space of the explanatory variables to provide predictions that vary across this space. In addition, they are specifically structured and fit to provide optimal prediction of the threshold-defined categories for the dependent variables. Both sets of inputs – components (C1, C2, C3) and component ratios ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}) – demonstrated high sensitivity, specificity, and accuracy results within the classification trees. ROC curves indicated that the 0.75 BIF tree with component inputs was the best model overall.

NOM fluorescence measurements were chosen preferentially over UV_{254} and DOC overall in the classification models, indicating their utility in DBP predictive models. C2 was identified as an

important input variable in all four component classification trees and $C1/F_{max}$ was identified as an important input variable in three of the four component ratio classification trees. Additionally, the use of multiple NOM characterizations within many of the models indicates that multiple NOM characterizations that describe different features of the NOM are necessary for creating robust predictive models. Bromide was used in all Br-THM models (0.75 BIF and 50% Br-THM), but in only one of the TTHM models (TTHM MCL and 80% MCL), indicating that NOM may be more predictive of TTHM regulation than bromide in this region.

The success of the classification trees demonstrates an alternative method for assessing overall treatability of source water within a basin and for broadly predicting the finished water quality from source water characteristics. Classification techniques can be used to create regional source water models for other areas experiencing source water changes to assess potential challenges for compliance with operational and regulatory thresholds of interest.

Chapter 4

FLUORESCENCE CHARACTERIZATION OF ORGANIC MATTER AND FOULING IN A FULL-SCALE REVERSE OSMOSIS MEMBRANE TREATMENT PLANT[2]

4.1 Abstract

Organic Matter in source water is responsible for organic fouling in membranes, reducing water flux and leading to biological fouling. Previous research has demonstrated the need to characterize organic matter for fouling prediction because development of an organic fouling layer on the membrane is dependent on the specific characteristics of the organic matter. A field study was performed at a full-scale reverse osmosis treatment plant that treats secondary wastewater for power plant boilers. Samples were collected at various points within the treatment train and analyzed for multiple water quality measurements, including turbidity, total organic carbon (TOC), conductivity, and fluorescence Excitation Emission Matrices (EEM). Parallel Factor Analysis (PARAFAC) analysis was also performed on the EEM to generate representative fluorescence measurements of the organic matter. Results showed that TOC and fluorescence measurements, both EEM peaks and EEM-PARAFAC components, were effective in differentiating between two observed fouling periods – frequent spikes in differential pressure and steady differential pressure – at multiple locations within the treatment plant. However, none of the water quality measurements were effective in tracking treatability of organic matter throughout pre-treatment. The results provide important information about the relationship between fluorescence organic matter signals and membrane fouling that can be used in future online detection systems.

[2] This chapter has been prepared as a manuscript that will submitted for publication as Bergman, L.E. and VanBriesen, J.M. "Fluorescence Characterization of Organic Matter and Fouling in a Full-Scale Reverse Osmosis Membrane Treatment Plant."

4.2 Introduction

Reverse osmosis (RO) membranes are an important technology in addressing water scarcity because they can effectively treat saline and low quality water. However, a major challenge for RO and other membrane separation processes is membrane fouling, characterized by a loss of water flux through the membrane (Arora and Trompeter, 1983), which increases energy use, reduces salt rejection and increases cost (Hoek and Elimelech, 2003; Hoek *et al.*, 2002; Song and Elimelech, 1995). In full-scale treatment plants, fouling is generally identified by an increase in the differential pressure across the membrane, and fouling associated with the sorption of organic matter to the membrane surface is a universal concern because natural organic matter (NOM) is ubiquitous in natural waters. Organic fouling is also a concern in membrane systems because the organic layer formed on the membrane surface leads to biofilm development and subsequent biological fouling, which significantly reduces flux through the membrane (Martínez *et al.*, 2015; Arora and Trompeter, 1983; Herzberg and Elimelech, 2007; Rukapan *et al.*, 2015). Further, organic fouling has been found to be the primary fouling challenge in full-scale RO membranes treating secondary effluent wastewater (Tang *et al.*, 2016).

Due to the complex nature of organic matter, the extent and severity of organic fouling is dependent on the specific characteristics of the organic matter present in the membrane feed water. For example, polysaccharides have been identified as a major component in organic fouling in multiple bench-scale ultrafiltration (UF) membrane studies (Yu *et al.*, 2014; Kennedy *et al.*, 2005). Other bench-scale UF studies have found that mixtures of organics foul more than individual compounds (Myat *et al.*, 2014; Ang *et al.*, 2011; Gray *et al.*, 2011). Biopolymers,

including carbohydrates and proteins, have also been identified as major foulants in microfiltration (MF), UF, RO, and membrane bioreactor studies (Yamamura *et al.*, 2014; Kennedy *et al.*, 2008; Zhao *et al.*, 2010; Miyoshi *et al.*, 2015). Similarly, microbial organic matter has been identified as a greater contributor to UF membrane fouling than terrestrial organic matter (Jutaporn *et al.*, 2016). Additionally, hydrophilicity and hydrophobicity of foulants and membrane surfaces are reported to affect membrane fouling behavior (Kennedy *et al.*, 2005; Yamamura *et al.*, 2014; Junaidi *et al.*, 2013). Even with the same amount of total organic carbon (TOC), various organic matter fractions have shown different fouling behavior, demonstrating that organic fouling is more related to the character of the organic matter than the quantity (Yamamura *et al.*, 2014).

Many studies have employed fluorescence measurements to characterize the fouling behavior of the influent water because it provides a more comprehensive characterization of organic matter than bulk parameters, such as TOC or turbidity. Both bench-scale and pilot-scale UF studies have linked protein-like fluorescence signals to increased fouling (Yu *et al.*, 2014; Shao *et al.*, 2014; Peiris *et al.*, 2010a; Peiris *et al.*, 2013; Chen *et al.*, 2014a). There have also been some conflicting results concerning humic-like fluorescence signals and bench scale UF foulants. Peiris *et al.* (2010a) found that humic-like signals were correlated to irreversible fouling, while Shao *et al.* (2014) found that foulants with humic-like signals do not contribute to fouling as much as foulants with protein-like signals. In a full-scale RO plant study, the 'microbial by-product-like' fluorescence signals in the brine were most correlated to fouling (Choi *et al.*, 2014).

Peiris and colleagues used the success of correlating fouling behavior to specific fluorescence signals to suggest future applications of online fluorescence monitoring of influent water to provide operators with early warnings of high fouling events about to occur (Peiris *et al.*, 2010a; Peiris *et al.*, 2010b; Jutaporn *et al.*, 2016). Other research concerning online monitoring of UV absorbance spectra and fluorescence spectra have further developed the technology of real-time organic matter tracking within treatment plants (Roccaro *et al.*, 2015; Shutova *et al.*, 2014). Implementation of online fluorescence detection may enable detection of real time changes in organic matter that affect membrane treatment, allowing operators to make real time pre-treatment changes to optimize membrane treatment by removing high-fouling organic matter fractions on an "as-needed" basis. The next step towards online fluorescence monitoring in a full-scale membrane treatment plant for fouling control through pre-treatment changes would be to determine what information fluorescence measurements provide about the treatability and changes in fouling potential associated with pre-treatment steps in a full-scale RO plant.

The goals of the present work were two-fold, (1) to link organic matter measurements, including fluorescence signals, to increased differential pressure at full scale through classification methods, and (2) to track changes in organic matter due to pre-treatment using fluorescence measurements. Identifying the organic matter changes that can be tracked using fluorescence measurements will be an important part of determining whether fluorescence can track treatability in source water organic matter. Further, the use of classification methods in linking organic matter measurements to increased differential pressure events (i.e. fouling events) will enable a determination of which fluorescence parameters most effectively predict fouling and what target thresholds of source water fluorescence measurements should be monitored.

4.3 Methods and Materials

Membrane Treatment Plant Operation

The full scale membrane treatment plant operates with two trains, each with a two-pass system that operates in a two stage configuration, as shown in Figure 4.1. The source water is secondary-treated wastewater effluent designed for reuse as boiler make-up water. Reverse osmosis membrane treatment is designed to remove all dissolved and particulate contaminants, including monovalent ions, to produce high purity water. Aquatech International, the partners in this collaborative study, own and operate the membrane treatment plant.

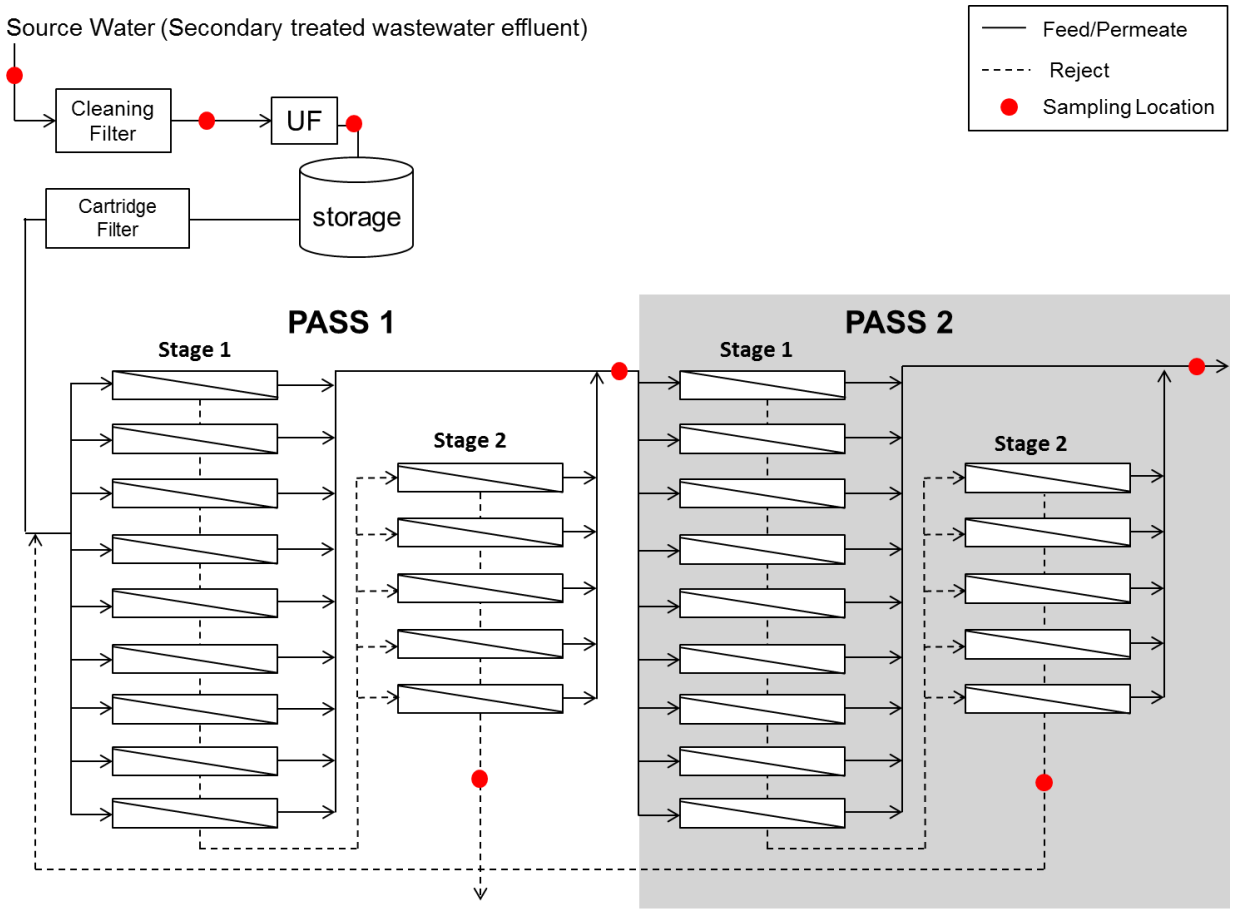

Figure 4.1: Schematic of full-scale membrane treatment plant, from which samples were collected. The schematic illustrates the two-pass, two-stage operation of one of the two trains used at the treatment plant. The red circles indicate locations at which water samples were collected for the study. Feed and permeate flows are depicted by solid black lines and reject flows are depicted by dotted black lines.

Source water (secondary treated wastewater) is first drawn into the plant and passed through two pre-treatments – a 100 μm cleaning filter and an ultrafiltration (UF) membrane. Following pre-treatment, feed water goes through a cartridge filter that feeds the two membrane treatment trains. The plant contains two cartridge filters and two membrane treatment trains, but only one operates as a time. Within each treatment train, there are nine parallel membrane vessels that comprise the Stage 1 configuration and five parallel membrane vessels that comprise the Stage 2 configuration. The reject streams from the nine Stage 1 membranes enter the five Stage 2

membranes as feed to minimize the volume of reject in the system. The permeate from the first pass RO set-up, containing permeate from both the Stage 1 and Stage 2 vessels enters the nine Stage 1 vessels in the second pass as feed.

Sample Collection

Samples were collected at the treatment plant from September 2014 to May 2015 at seven different points within the treatment system (see Figure 2) – at (1) the source water intake, (2) the cleaning filter effluent, (3) the ultrafiltration membrane permeate, (4) the first pass RO permeate (including permeate from both Stage 1 and Stage 2), (5) the first pass RO reject from Stage 2, (6) the second pass RO permeate (again, including permeate from both Stage 1 and Stage 2), and (7) the second pass RO reject from Stage 2. Sample collection began several months after the plant began operation so no samples were collected during the initial start-up time. Duplicate samples were taken at each sampling point. One set of sterile 125 mL bottles were pre-rinsed with each sample and then filled completely with the sample for EEM measurements. Sulfuric acid was first added to the other set of sterile 125 mL bottles to preserve samples, according to EPA Method 415.3(EPA, 2009). Bottles were then filled with sample water at each sampling point for TOC measurements. Samples were shipped on ice to Carnegie Mellon University in Pittsburgh, PA and stored at 4°C and analyzed within allowable hold times. Turbidity and conductivity were measured on site by treatment plant operators as samples were collected.

Organic Matter Measurements

TOC was measured at Carnegie Mellon using a Sievers InnovOx Laboratory TOC Analyzer (GE, Boulder, CO). Excitation Emission Matrices were measured on all samples collected using a Fluoromax-4 Spectrofluorometer (Horiba, Kyoto, Japan) with excitation wavelengths ranging from 200 nm to 550 nm with a 5 nm step-size, and emission wavelengths ranging from 250 nm to 650 nm with a 5 nm step-size. Raman scans and blank EEM were also measured on the same day as the sample EEM. Raman scans were run at an excitation wavelength of 350 nm and emission wavelengths 371 – 428 nm at 1 nm intervals, according to Lawaetz and Stedmon (2009). Blank EEM were measured on Milli-Q water under the same parameters as the sample EEM. Following measurement, blank EEM were subtracted from each sample EEM to remove the spectroscopic effects of water and negative values were set to zero. Blank-subtracted EEM were then converted to Raman Units by normalizing over the area under the Raman scan (Lawaetz and Stedmon, 2009; Murphy *et al.*, 2013).

Parallel Factor Analysis, Classification Methods, and Wilcoxon Rank Sum Tests

Parallel Factor Analysis was performed on EEM of the (1) source water, (2) cleaning filter effluent, (3) UF membrane permeate, (4) RO pass one reject from Stage 2, and (5) RO pass two reject from Stage 2 in Matlab using the DOMFluor Toolbox (http://www.models.life.ku.dk/algorithms) created by Stedmon and Bro (2008) . All permeate samples – RO pass one permeate and RO pass two permeate – were excluded from the PARAFAC analysis because they showed almost no fluorescence signal due to their high purity (very low carbon content). Seventy-nine samples were incorporated into the PARAFAC analysis, which produced three components. Classification Trees were created in R using the *rpart* library

(RCoreTeam, 2015). The trees were cross validated and pruned using a minimum split of 4 instances. Classification Trees are useful for operational decision making because they provide easy to understand results of how the input data affect the outcome of interest. In this dataset, as is characteristic of other environmental dataset, different patterns exist in different subsets of the data. Unlike commonly used regression techniques, classification trees can incorporate the various behaviors into one model. Tests of significance were also performed in R using the Wilcoxon Rank Sum tests (Ng and Balakrishnan, 2004; Bauer, 1972). Wilcoxon Rank Sum tests were used because they are non-parametric, which was important to use in this data set due to the small number of instances and therefore the inability to meet the criteria necessary to use t-tests. The tests were used to determine whether two sample distributions were significantly different from one another. The R *wilcox.test* function provides outputs of (1) the Hodges Lehmann estimator (the estimate of the difference between the sample distributions), (2) the associated p-value, and (3) the 95% confidence interval of the differences between the two sample distributions.

4.4 Results and Discussion

Differential Pressure and Cleaning Events

The plot in Figure 4.2 shows the pattern of differential pressure in stage 1 vessels of the first pass membrane treatment over time.

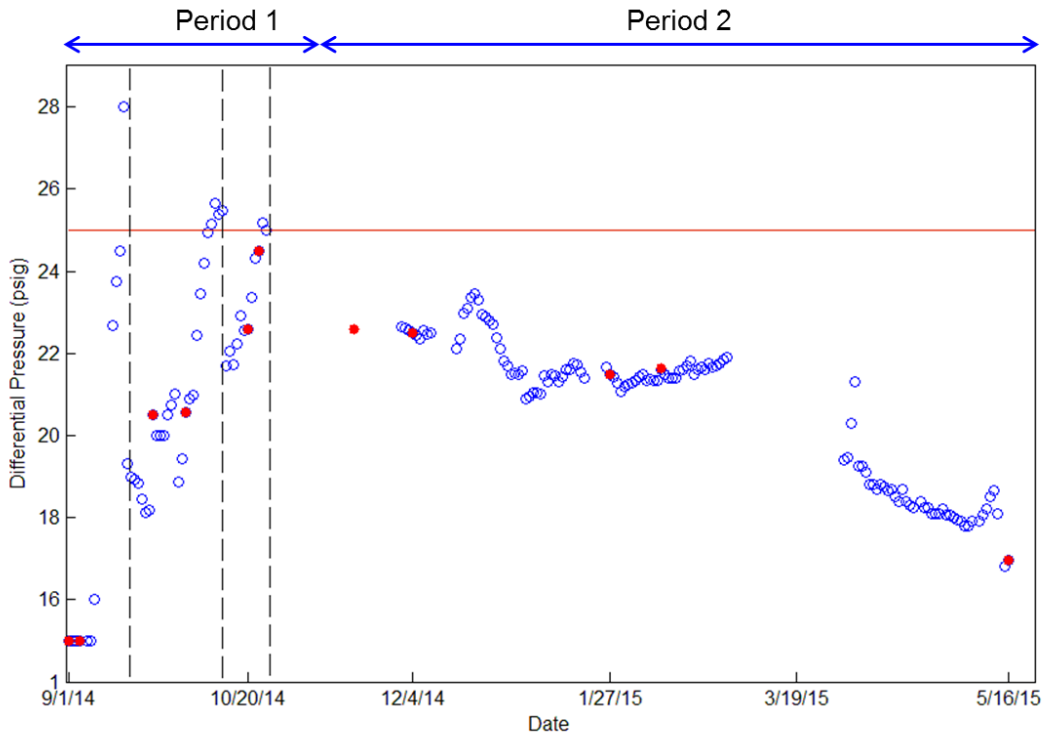

Figure 4.2: Plot of differential pressure in the stage 1 pass 1 membrane vessels over the sampling period. Blue open dots show the differential pressure trend over time, while the red solid dots indicate differential pressure for the times at which samples were collected. The red horizontal line indicates a differential pressure of 25 psig, the cleaning threshold. The vertical purple dashed lines indicate when cleanings most likely occurred, based on the 25 psig differential pressure limit followed by plant operators.

The blue dots show the differential pressure for the whole study period and the red dots show the differential pressure for the dates at which water samples were collected and analyzed. The vertical black dotted lines show when cleaning is needed in the system, based on the general cleaning rule that is followed in this plant – membranes are chemically cleaned when the

differential pressure in the stage 1 pass 1 reaches 25 psig (shown by the horizontal red line). The threshold of differential pressure, 25 psig, was reached three times during the September, 2014 to May, 2015 sample collection period.

During this field study, there were two distinct periods observed – the first one spanning from September, 2014 to October, 2014 and the second one spanning November, 2014 to May, 2015. During the first period, September – October, 2014, a higher frequency of elevated differential pressure (above 25 psig) was observed, indicating that more frequent cleanings were needed. The second period showed more stable differential pressure. From the available data, it is likely that few/infrequent cleanings were required during the second period from November, 2014 to May, 2015. Missing data during November, 2015 and March, 2015 makes it impossible to say for certain that the differential pressure did not exceed 25 psig and require a cleaning during this period; however, even if cleanings occurred during the missing data periods, these would still be less frequent than in the earlier time period. These two distinct periods during sample collection will be referred to as Period 1 and Period 2 throughout the paper. Given that Period 1 and Period 2 show very distinct fouling behavior, differences in source water characteristics were explored to assess whether source water changes were related to fouling changes.

Turbidity, TOC, and Conductivity in Periods 1 and 2. Table 4.1 shows average turbidity, TOC, and conductivity values for the three pre-membrane samples – source water (SW), cleaning filter effluent (CF), and UF membrane permeate (UF) – within each fouling period – Period 1 and Period 2. Turbidity and conductivity are frequently measured at membrane treatment plants to characterize source water for membrane treatment.

Table 4.1: Summary of average Turbidity, TOC, and Conductivity values for the three pre-membrane samples (SW, CF, and UF) for both Period 1 and Period 2

	Turbidity (NTU)		TOC (mg/L)		Conductivity (µS/cm)	
Sample	**P1**	**P2**	**P1**	**P2**	**P1**	**P2**
SW	2.89	15.60	2.59	2.03	323.15	438.44
CF	3.96	7.95	2.53	1.93	357.35	453.40
UF	0.35	0.43	2.24	1.84	367.82	453.22

Turbidity, a bulk measure of optical density, does not differentiate between organic and inorganic constituents. Generally, turbidity is monitored because higher turbidity indicates higher fouling potential of the feed water (Guastalli *et al.*, 2013; Brehant *et al.*, 2002; Lorain *et al.*, 2007). Wilcoxon Rank Sum tests indicate that for turbidity measurements, differences in the source water between Period 1 and Period 2 are significant at the $\alpha = 0.05$ level. However, the CF and UF samples, which are closer to membrane treatment and therefore have a greater impact on fouling, do not have statistically significantly different turbidity measurements between the two periods. Therefore, it is not possible to attribute the difference in turbidity to the difference in fouling behavior because the turbidity of the water that enters membrane treatment is not significantly different between the two fouling periods. The full results of the Wilcoxon Rank Sum Tests can be found in Table B1 in Appendix B.

TOC is also commonly measured in bench-scale studies because it can provide some indication of fouling potential since a higher level of organic matter in the influent is expected to contribute more to organic fouling (Yamamura *et al.*, 2014; Schäfer *et al.*, 2000; Tang *et al.*, 2007). TOC, however, shows statistically significantly different (p-value < 0.05) values between the two

periods for each of the three pre-membrane samples (see Table B2 in Appendix B), indicating this measurement might provide insight into TOC removal and fouling but not insight into how pre-treatment effectiveness alters fouling potential. While in this study, higher TOC was associated with the higher fouling period, previous studies have not identified TOC as a good measure of membrane fouling potential. Shao *et al.* (2014) found even though increases in DOC contributed to increased membrane fouling, removal of DOC and reductions in UF membrane fouling were not proportional. Further, Yamamura *et al.* (2014) demonstrated that even with similar TOC concentrations, different NOM fractions (i.e. hydrophilic and hydrophobic) foul membranes differently.

Conductivity, the measure of dissolved ions, also provides important information about the fouling potential of a given source water because ions contribute to inorganic fouling (scaling). Higher conductivity generally results in more scaling and more frequent cleaning. In the present case, scaling is not expected since the source water is low salinity secondary treated wastewater effluent. Conductivity is considered here because previous work has shown that the increased presence of dissolved ions exacerbates flux decline due to organic fouling (Hong and Elimelech, 1997; Lee and Elimelech, 2006; Saravia *et al.*, 2006; Tang *et al.*, 2007; Gray *et al.*, 2011). Though an important parameter to consider, conductivity has limitations in predicting fouling in full-scale RO plants compared to NOM characterization techniques (Choi *et al.*, 2014). In the present study conductivity does not provide statistically significant differences associated with differences in fouling behavior for any of the three pre-membrane samples (see Table B3 in Appendix B), and therefore is not predictive of fouling behavior.

Fluorescence Characterization of Organic Matter in Periods 1 and 2. Fluorescence characterization was also investigated to determine if it (1) showed statistically significant differences between the two periods, and (2) provided any additional information related to operational control of fouling. Since organic fouling is dependent on the character of the organic matter that comes in contact with the membrane surface, it was expected that a more comprehensive measurement of organic matter in the system that captures more information about the character of the organic matter, as opposed to a bulk measurement, would provide more insight into the unique characteristics of the organic matter that affected differential pressure and caused the two distinct periods. The EEM fluorescence peaks show differences in fluorescence behavior between Period 1 and Period 2; Figure 4.3 shows a plot of EEM peak fluorescence intensity for the pre-membrane samples (SW, CF, UF) over time, with Period 1 and Period 2 separated by a vertical dashed black line.

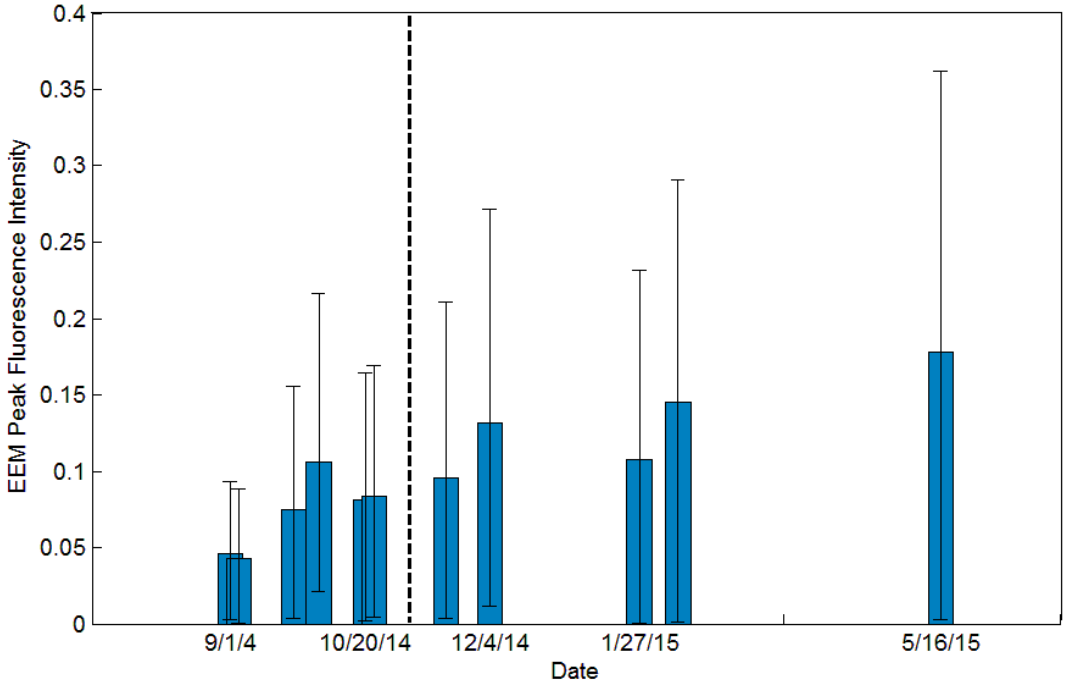

Figure 4.3: Plot of Peak Fluorescence Intensity of the EEM over time. The plot shows peak fluorescence for the pre-membrane samples – bars show the median value and the error bars represent the minimum and maximum. Also shown is a vertical black dotted line, indicating the separation of the two differential pressure periods.

The peak fluorescence intensities overall show distinct differences between Period 1 and Period 2, with Period 2 intensities appearing to be much higher overall. Wilcoxon Rank Sum tests (shown in SI Table S5) confirm the significant difference in peak fluorescence intensity between the two periods. Further, the location of the EEM peak is not indicative of pre-treatment organic matter removal or frequent increases in differential pressure. Most of the peaks occur between EX = 325 – 375 nm and EM = 410 – 450 nm. Further, there is no distinct difference in peak location for samples taken during each of the two differential pressure periods. Other studies have shown that peak location can be used for source identification (Cabaniss and Shuman, 1987; Sierra *et al.*, 1994), and thus, it is not surprising that peak location remained fairly consistent for this single source study.

PARAFAC analysis was then performed on the set of sample EEM to gain a more complete understanding of the fluorescence signals throughout the dataset. The PARAFAC analysis incorporated the three pre-membrane samples, SW, CF, and UF, and the two rejects, RO R1 and RO R2. All three PARAFAC components, referred to as C1, C2, and C3, have maximums within the humic-like region – EX/EM = 360/440 for C1, EX/EM = 395/505 for C2, and EX/EM = 305/375 for C3 (Chen *et al.*, 2003). Boxplots of fluorescence maximums for the three pre-membrane samples, SW, CF, and UF for each of the three PARAFAC components, C1, C2 and C3, in Period 1 and Period 2 are shown in Figure 4.4.

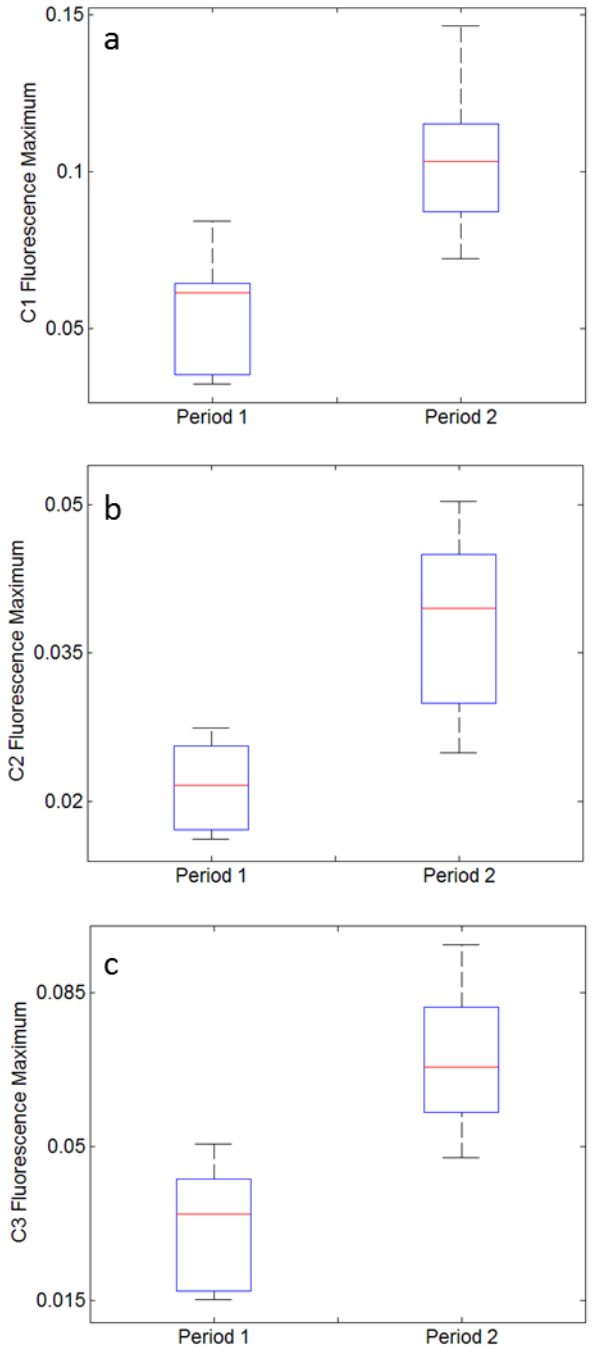

Figure 4.4: Boxplots of component maximum ranges for the three pre-treatment samples (SW, CF, UF) for all three components in each of the two differential pressure periods. Plots shown are: (a) C1, (b) C2, and (c) C3.

Each boxplot contains all three pre-membrane samples (SW, CF, UF) because they showed similar behavior *within* each of the two fouling periods. The sets of boxplots in Figure 4.4 show a significant difference in component maximum fluorescence between the two fouling periods – each of the three PARAFAC components show much lower values in Period 1 than in Period 2, as suggested by the EEM peak fluorescence intensity plot (Figure 4.3).

Wilcoxon Rank Sum Tests (SI Table S6) confirm that differences in component fluorescence between Period 1 and Period 2 are statistically significant. These are somewhat unexpected results (i.e. lower fluorescence during increased fouling) given that fluorescence is related to concentration; however, there are many other organic matter characteristics that also affect the fluorescence activity, such as the molar absorptivity, quantum efficiency, aromatic content, and molecular weight (Stedmon and Bro, 2008; Chen *et al.*, 2003; Cuss and Gueguen, 2014). The distinct differences in EEM-PARAFAC component fluorescent activity between the high fouling period (Period 1) and the low fouling period (Period 2) indicate that organic matter fluorescence characterization can be used to track changes in organic matter that affect membrane performance. In addition to EEM-PARAFAC components, EEM fluorescence peak intensities also show significant differences between Period 1 and Period 2 (SI Table S5).

Classification of Instances Based on Source Water Organic Matter. Given the significant differences in EEM-PARAFAC components between the two periods, classification methods were used to determine if fluorescence characterization could be used for operational control and fouling management. Online detection of organic matter fluorescence for improved operation of treatment systems has been proposed by other researchers (Roccaro *et al.*, 2015; Shutova *et al.*, 2014). For online detection to be useful, it is also necessary to determine threshold values within the various measurements that indicate to operators when to expect a change in organic matter character and subsequent membrane behavior. To determine these threshold values, classification trees were employed. Classification trees were used to classify instances based upon when cleanings need to occur. According to general plant operation/protocol, cleanings occur when the differential pressure in stage 1 pass 1 reaches 25 psig. Using the same Period 1 and Period 2 division, a cleaning binary response variable was created in order to classify instances as "frequent cleanings" or "infrequent cleanings," based on source water characteristics. The classification tree model allows for the determination of which source water characteristics can be used to classify the cleaning frequency behavior in the membrane, and the appropriate division of the source water parameter for each cleaning classification.

Classification trees were created using the three components – C1, C2, and C3 – as well as turbidity and TOC from the three pre-membrane sets of samples – SW, CF, and UF. Only parameters measuring the organic matter were used (i.e. conductivity was excluded) because it is expected that organic matter is the main source of differential pressure increases in this plant. Multiple subsets of input parameters were tested, and it was determined that multiple individual parameters could accurately classify all of the instances as "frequent cleaning" or "infrequent cleaning." Table 4.2 shows a summary of the input parameters that correctly classify all the

instances through a single division of the data, along with the inequality values that separate Period 1 and Period 2.

Table 4.2: Summary of Single Parameter Classifications of the Two Fouling Periods.

Input Parameter	Frequent Cleaning	Infrequent Cleaning
SW C1	< 0.077	≥ 0.077
SW C2	< 0.028	≥ 0.028
SW C3	< 0.055	≥ 0.055
CF C2	< 0.027	≥ 0.027
CF C3	< 0.047	≥ 0.047
CF TOC	≥ 2.22	< 2.22
UF C2	< 0.024	≥ 0.024
UF C3	< 0.045	≥ 0.045

For each of the input parameters in Table 4.2, there is a range of values that the parameter takes within each of the two periods. For example, SW samples associated with “frequent cleaning” have C1 values below 0.077 while SW samples associated with “infrequent cleaning” have C1 values that equal or exceed 0.077. Within this particular dataset, plant operators could use the range limits as threshold values to serve as an indicator in an online monitoring system. These results support the concept that online fluorescence monitoring could be feasible in a full-scale RO plant; however, without widespread sampling of treatment plants with different source waters, it is impossible to know if the particular input parameter and threshold results generated from this study could be applied to other treatment plants. Nevertheless, the method

demonstrated here with field data collection and classification methods could be used at other sites to determine their own thresholds.

These results are in general agreement with those from the Wilcoxon Rank Sum tests and the EEM-PARAFAC Component boxplots. The only input parameters that were selected by the classification trees to make a single parameter tree with one branch split were EEM-PARAFAC components and TOC, which were also the only parameters that showed statistically significant differences between Period 1 and Period 2 using the Wilcoxon Rank Sum test (TOC) and the boxplots. Turbidity and conductivity, which did not show clear differences between the two periods, were not selected by the classification trees. The classification trees, however, showed more selectivity than the other analyses. Although TOC from all three pre-membrane samples (SW, CF, UF) showed statistically significant differences between the periods, only TOC of the CF samples could classify Period 1 and Period 2 with a single division.

Further, C1 is capable of dividing the instances into “frequent cleaning” (Period 1) or “infrequent cleaning” (Period 2) only when assessed in the source water. Overall, the classification results show that monitoring organic matter provides similar information when undertaken at multiple locations within the treatment plant. SW, CF, and UF fluorescence measurements all show distinct divisions of the Period 1 and Period 2 data with clear threshold values. The results support monitoring any of the fluorescence signals at the intake, which is advantageous because it is the furthest from membrane treatment and therefore allows additional time for operational

changes, including incorporating additional pre-treatment steps, to be made prior to membrane treatment.

Tracking Organic Matter Changes throughout Pre-treatment

Given the success of EEM-PARAFAC components and TOC to classify instances based on their fouling behavior, the organic matter characterization techniques were also investigated for their ability to track organic matter changes throughout pre-treatment. Pre-treatments are used to remove organic matter and improve performance of membrane treatment, however, they exhibit preferential removal of certain organic matter fractions and therefore some are more effective than others depending on the character of the organic matter (Kitis *et al.*, 2001; Shao *et al.*, 2014) . The ability to track organic matter removal throughout pre-treatment using organic matter characterization would aid in making online organic matter monitoring more effective.

Turbidity, TOC, and Conductivity throughout Pre-treatment. Table 4.1 shows differences in turbidity, TOC, and conductivity among the three pre-membrane samples (SW, CF, and UF) in addition to differences between Period 1 and Period 2. However, according to Table 4.1, the differences among pre-treatment samples are often less considerable than between the two fouling periods. Further, Wilcoxon Rank Sum Tests (results shown in Appendix B) reveal that many of the differences among pre-treatments are not statistically significant.

Turbidity values showed statistically significant differences between CF and UF samples and between the SW and UF samples in Period 1 (at the α=0.05 level), however, the none of the changes in Period 2 were significant. The higher average turbidity in the CF samples than the

SW samples during Period 1 is unexpected, however these difference between SW and CF are not significant. Even though the samples were taken at the plant in order (i.e. SW was collected first, then CF, and so on), the discrepancy in the turbidity results may be due to the fact that the full-scale plant was operational during the sampling period and therefore it was impossible to ensure that the same column of water is being sampled throughout. Despite the much higher average turbidity in SW than CF for Period 2, there difference is not significant. The average SW Period 1 turbidity was impacted by one very high measurement on 5/16/2015, however the rest of the turbidity measurements are similar to those for CF samples.

TOC concentrations (Table 4.1) show slight decreases throughout pre-treatment, however, the only significant change in concentration is the decrease from SW to UF in Period 1, a result of two pre-treatments. None of the decreases in TOC concentration across individual pre-treatments (i.e. SW to CF) show statistically significant differences. Like TOC, conductivity does not provide useful information about the changes to water quality throughout pre-treatment. Conductivity does not exhibit any statistically significant differences across pre-treatment; and further, shows unexpected increases with treatment.

Fluorescence Characterization of Organic Matter throughout Pre-Treatment. Though turbidity, TOC, and conductivity did not track changes in source water quality throughout pre-treatment, the broad selection of fluorescence measurements in classifying fouling periods (Table 4.2) provides support for its use to track organic matter changes throughout pre-treatment as well. Previous studies have shown that EEM-PARAFAC component fluorescence intensities

decrease with additional treatments (Shutova *et al.*, 2014; Baghoth *et al.*, 2011). The more comprehensive organic matter characterization available with fluorescence measurements were expected to be able to better track changes with pre-treatment; however, like turbidity and TOC, neither the EEM peak fluorescence intensities nor the EEM-PARAFAC components reliably track organic matter changes throughout pre-treatment (see SI Figure S1). Only four sets of EEM peak fluorescence intensity grouped bars in SI Figure S1 (10/20/2014, 11/18/2014, 12/4/2014, 2/10/2015) show a consistent trend in fluorescence throughout pre-treatment (in this case decreasing intensity), and none of the EEM peak fluorescence intensity measurements show statistically significant differences throughout pre-treatment (SI figure S1 and Table S5). Finally, the only differences in EEM-PARAFAC component maximum fluorescent values within pre-treatment that were statistically significant were C2 between SW and CF in Period 1, and C3 in Period 1 between both subsequent pairs (SW to CF and CF to UF), as well as overall (SW to UF). Thus, surprisingly, EEM analysis was not suitable for determining the effect of specific pre-treatment steps on TOC character or fouling potential.

4.5 Conclusions

The results of this study suggest implementation of online fluorescence monitoring of fouling potential could provide real-time information for process control. TOC, peak EEM fluorescence intensities, and EEM-PARAFAC component maximums show significantly different behavior in the two fouling periods – Period 1, characterized by frequent increases in differential pressure and need for cleaning, and Period 2, characterized by stable differential pressure and less need for cleaning. Further, classification techniques identified threshold values that should be monitored in an online detection system for multiple EEM-PARAFAC components at various

points in the treatment systems prior to membrane treatment, as well as TOC for the CF samples. All three EEM-PARAFAC components showed clear distinctions between Period 1 and Period 2, with a specific threshold value at the intake. This result is important for implementation of online detection because it means that monitoring can occur at the intake, providing operators with more time to make an operational change in the plant once a fluorescent signal of concern is identified. Although multiple parameters show potential for prediction of fouling events, none of the organic matter characterization techniques were effective in detecting differences in organic matter throughout pre-treatment that were relevant for fouling control.

Chapter 5

SUMMARY, CONCLUSIONS AND FUTURE WORK

Summary and Conclusions

This research focused on fluorescence EEM-PARAFAC components, a NOM characterization technique, and investigated its utility in two water treatment applications – disinfection by-product formation and membrane fouling – in two field studies. The literature review outlines the many different NOM characterization techniques along with their limitations, while the two research applications assess the utility of fluorescence NOM characterization for operational control. Three main conclusions are:

1. Despite years of NOM characterization research, there is still disagreement in the literature about how the various NOM characterization techniques can track treatability in water treatment systems. Various patterns exist in the literature about which NOM fractions are the most reactive in DBP formation or have the highest propensity to foul membranes. Fluorescence EEM provide an advantage over other NOM characterization techniques in that the information-dense measurements provide unique fingerprints of the NOM character. The use of fluorescence EEM in DBP formation potential and membrane fouling studies has identified its utility in addressing water treatment challenges.

2. Classification techniques and the use of fluorescence EEM-PARAFAC components as inputs can be used to create regional watershed models for general treatability concerns. Robust classification models were developed for four regulatory and DBP speciation targets – the TTHM MCL, 80% of the TTHM MCL, a BIF of 0.75, and 50% Br-THM. Fluorescence NOM

characterizations were important input parameters in all the classification trees, whereas bromide was only incorporated into the brominated THM (0.75 BIF and 50% Br-THM) trees, not the regulatory TTHM ones (TTHM MCL and 80% TTHM MCL). Finally, site by site validation tests showed compelling evidence that these models can be applied to multiple sites within a geographic region. The models provided a method for (1) addressing treatability concerns and (2) predicting finished water quality within a regional watershed.

3. Multiple organic matter characterizations – TOC, peak EEM fluorescence, and EEM-PARAFAC components – can be used to track differences in membrane fouling behavior within a full-scale treatment plant. The use of classification techniques in determining threshold values of NOM fluorescence parameters that correspond to fouling behavior provides important results necessary for future implementation of online fluorescence detection. Further, organic matter fluorescence parameters at multiple locations within the plant showed clear divisions between the two main fouling behaviors, indicating that monitoring could occur at multiple locations, including the influent, which would provide the greatest opportunity for operational changes to occur. Although indicative of fouling behavior, none of the organic matter characterization techniques were successful in tracking organic matter changes due to pre-treatment in the plant. Other techniques may be necessary for addressing general treatability concerns of the water and pre-treatments.

Overall, fluorescence EEM-PARAFAC components show promise for use in full-scale water treatment applications. They have proven to show distinct differences in intensities and some variation in signals among various water treatment outcomes – meeting or exceeding DBP target

values as well as indicating a general level of membrane fouling. This research is an important step in furthering the effort to apply EEM-PARAFAC components for operational control in water treatment systems through online fluorescence detection.

Future Work

Future work is needed in expanding knowledge of fluorescence NOM characterization for implementation of online fluorescence detectors in full-scale water treatment plants, as well as in investigating the potential to track DBP formation and fouling propensity jointly within a system. Additionally, further work is needed to bridge the gap between lab studies and implementation in full-scale systems – specifically, in performing field studies with full-scale treatment plants. Further, the full-scale systems seeking to implement online detection must identify the specific fluorescence signals that are indicative of such water treatment challenges. More field studies incorporating full-scale plants in geographically diverse regions are needed to determine if there are universal fluorescence signals for DBP formation and membrane fouling prediction. Otherwise, work should be done to develop generalized models that can be calibrated to fit region-specific NOM parameters prior to implementation. EEM-PARAFAC components also face limitations that should be addressed in future work, namely the lack of a strong fluorescence signal when the DOC concentration is too low or when the DOC is made up of simple organic structures. In these cases, EEM-PARAFAC should be coupled with another NOM characterization techniques that can accurately characterize these samples.

The first priority is performing field studies at full-scale treatment plants that implement fluorescent monitoring of raw water and water following the various pre-treatment steps. Such

studies will enable a more robust assessment of the ability of fluorescence measurements to monitor treatability. Furthermore, conducting field studies at various treatment plants will provide more information on whether these methods are universally applicable (with site-specific validation) or whether there are limitations based on location. The second priority is to then conduct field studies with online fluorescence detection. Successful detection of water quality changes that indicate downstream treatment problems with the use of fluorescence monitoring will provide ample support for implementation of fluorescence detectors for routine monitoring in treatment plants.

The combination of fluorescence detection of DBP formation and membrane fouling predictors also has implications for water reuse. With fresh water resources becoming more strained due to population growth and climate change, there are increasing efforts to produce potable water from seawater (desalination) and wastewater (reuse). Desalination and water reuse employ membranes for water treatment, and when the end product is potable water for consumers, disinfection and disinfection by-products must be considered. In these applications it is necessary to know the DBP formation potential of membrane treated water if it is for eventual human consumption. If field studies (previously outlined) show successful implementation of online fluorescence detectors, future work should include analysis of fluorescence measurements for both membrane fouling potential and DBP formation potential for a membrane treatment plant.

Appendix A

SUPPLEMENTAL INFORMATION FOR CHAPTER 3 – APPLICATION OF CLASSIFICATION TREES FOR PREDICTING DISINFECTION BY-PRODUCT FORMATION TARGETS FROM SOURCE WATER CHARACTERISTICS

EEM-PARAFAC COMPONENT DATA

TableA1: Summary of EEM-PARAFAC Component Data for 109 instances

Date	Site	C1	C2	C3	Fmax	C1/Fmax	C2/Fmax	C3/Fmax	ILR1	ILR2	ILR3
2/16/2011	A	0.053	0.028	0.018	0.099	0.534	0.282	0.184	0.007	-1.195	0.300
4/27/2011	A	0.074	0.036	0.020	0.130	0.571	0.277	0.152	0.284	-1.044	0.424
6/29/2011	A	0.111	0.072	0.052	0.236	0.473	0.307	0.220	0.251	-0.493	0.238
7/13/2011	A	0.094	0.056	0.033	0.183	0.514	0.307	0.179	0.268	-0.730	0.382
8/3/2011	A	0.103	0.073	0.035	0.210	0.488	0.346	0.166	0.289	-0.595	0.518
8/11/2011	A	0.109	0.073	0.052	0.235	0.466	0.311	0.223	0.227	-0.485	0.236
8/17/2011	A	0.079	0.043	0.025	0.147	0.538	0.293	0.169	0.233	-0.919	0.390
10/5/2011	A	0.097	0.059	0.026	0.182	0.534	0.323	0.143	0.373	-0.761	0.574
3/20/2012	A	0.052	0.030	0.012	0.094	0.556	0.316	0.128	0.128	-1.258	0.639
5/2/2012	A	0.064	0.042	0.015	0.121	0.530	0.347	0.124	0.201	-1.043	0.728
5/29/2012	A	0.063	0.035	0.015	0.114	0.556	0.308	0.136	0.209	-1.121	0.580
7/5/2012	A	0.067	0.048	0.018	0.132	0.504	0.361	0.136	0.154	-0.946	0.692
7/12/2012	A	0.066	0.048	0.018	0.132	0.497	0.368	0.135	0.138	-0.938	0.707
7/23/2012	A	0.049	0.030	0.010	0.089	0.547	0.335	0.117	0.108	-1.284	0.744
7/30/2012	A	0.095	0.056	0.032	0.182	0.519	0.308	0.173	0.287	-0.739	0.408
8/14/2012	A	0.079	0.043	0.026	0.148	0.533	0.292	0.175	0.218	-0.906	0.363
5/4/2010	B	0.051	0.025	0.014	0.091	0.563	0.280	0.157	0.074	-1.292	0.408
5/18/2010	B	0.043	0.024	0.009	0.076	0.568	0.315	0.117	0.075	-1.426	0.700
6/29/2010	B	0.057	0.030	0.019	0.106	0.536	0.286	0.178	0.058	-1.144	0.337
7/13/2010	B	0.053	0.031	0.016	0.100	0.530	0.313	0.157	0.039	-1.177	0.486
7/22/2010	B	0.072	0.051	0.024	0.147	0.488	0.348	0.165	0.111	-0.849	0.528
7/27/2010	B	0.030	0.012	0.009	0.051	0.586	0.242	0.172	-0.169	-1.733	0.240
8/17/2010	B	0.039	0.022	0.011	0.072	0.549	0.305	0.147	-0.063	-1.438	0.518
8/24/2010	B	0.060	0.046	0.020	0.126	0.475	0.369	0.156	0.020	-0.940	0.609
8/31/2010	B	0.059	0.041	0.016	0.116	0.510	0.354	0.136	0.104	-1.046	0.677
9/7/2010	B	0.053	0.033	0.017	0.102	0.515	0.319	0.165	0.000	-1.143	0.467
9/21/2010	B	0.051	0.034	0.018	0.103	0.497	0.329	0.174	-0.057	-1.114	0.450
9/29/2010	B	0.047	0.026	0.014	0.087	0.538	0.299	0.164	-0.022	-1.288	0.425
10/7/2010	B	0.050	0.030	0.013	0.093	0.538	0.324	0.138	0.058	-1.239	0.601
10/19/2010	B	0.100	0.053	0.035	0.187	0.535	0.280	0.185	0.330	-0.741	0.295

10/28/2010	B	0.108	0.076	0.035	0.219	0.491	0.349	0.160	0.328	-0.570	0.551
12/16/2010	B	0.076	0.044	0.027	0.146	0.518	0.299	0.183	0.162	-0.896	0.347
1/19/2011	B	0.041	0.017	0.010	0.068	0.595	0.255	0.149	0.031	-1.533	0.380
2/16/2011	B	0.034	0.013	0.010	0.057	0.601	0.223	0.176	-0.070	-1.668	0.164
4/27/2011	B	0.046	0.018	0.012	0.076	0.605	0.240	0.155	0.108	-1.466	0.307
5/25/2011	B	0.089	0.043	0.024	0.155	0.572	0.275	0.153	0.373	-0.923	0.418
6/23/2011	B	0.090	0.047	0.039	0.175	0.511	0.269	0.220	0.208	-0.756	0.140
6/29/2011	B	0.070	0.033	0.031	0.134	0.520	0.248	0.231	0.096	-0.962	0.050
7/6/2011	B	0.110	0.063	0.036	0.208	0.528	0.300	0.172	0.379	-0.658	0.395
7/13/2011	B	0.051	0.029	0.014	0.094	0.546	0.303	0.151	0.058	-1.242	0.493
8/17/2011	B	0.099	0.066	0.029	0.195	0.509	0.339	0.152	0.328	-0.680	0.570
8/31/2011	B	0.083	0.051	0.024	0.158	0.528	0.320	0.151	0.271	-0.854	0.531
9/7/2011	B	0.040	0.021	0.020	0.081	0.497	0.259	0.244	-0.229	-1.285	0.043
10/5/2011	B	0.068	0.031	0.017	0.116	0.585	0.270	0.145	0.277	-1.142	0.438
10/21/2011	B	0.075	0.040	0.023	0.138	0.544	0.288	0.168	0.218	-0.971	0.378
5/2/2012	B	0.042	0.025	0.010	0.077	0.543	0.328	0.129	-0.007	-1.380	0.659
5/29/2012	B	0.067	0.032	0.020	0.119	0.566	0.266	0.169	0.203	-1.104	0.322
7/23/2012	B	0.045	0.030	0.011	0.086	0.527	0.343	0.130	0.007	-1.282	0.686
7/30/2012	B	0.083	0.041	0.029	0.153	0.544	0.266	0.190	0.250	-0.898	0.240
8/14/2012	B	0.071	0.037	0.021	0.129	0.550	0.285	0.165	0.207	-1.025	0.390
5/4/2010	C	0.063	0.032	0.015	0.111	0.571	0.291	0.139	0.226	-1.157	0.522
6/29/2010	C	0.080	0.048	0.029	0.157	0.509	0.307	0.184	0.174	-0.832	0.364
7/5/2010	C	0.092	0.061	0.032	0.185	0.497	0.331	0.172	0.240	-0.700	0.464
7/22/2010	C	0.071	0.051	0.025	0.146	0.485	0.347	0.168	0.100	-0.848	0.515
7/27/2010	C	0.084	0.052	0.029	0.165	0.510	0.312	0.178	0.208	-0.795	0.398
8/17/2010	C	0.086	0.055	0.030	0.171	0.503	0.324	0.173	0.212	-0.763	0.441
8/24/2010	C	0.074	0.055	0.026	0.155	0.475	0.356	0.169	0.101	-0.792	0.527
8/31/2010	C	0.078	0.063	0.021	0.162	0.481	0.387	0.132	0.209	-0.770	0.760
9/7/2010	C	0.069	0.053	0.024	0.146	0.473	0.364	0.163	0.077	-0.831	0.567
9/21/2010	C	0.084	0.053	0.027	0.164	0.511	0.322	0.168	0.220	-0.803	0.462
9/29/2010	C	0.079	0.054	0.029	0.162	0.487	0.334	0.178	0.142	-0.778	0.444
10/7/2010	C	0.070	0.035	0.027	0.131	0.530	0.265	0.205	0.125	-0.987	0.182
10/19/2010	C	0.103	0.057	0.040	0.200	0.516	0.285	0.199	0.300	-0.670	0.253
10/28/2010	C	0.077	0.057	0.028	0.162	0.477	0.349	0.174	0.120	-0.763	0.493
11/3/2010	C	0.099	0.069	0.043	0.211	0.469	0.327	0.204	0.197	-0.564	0.334
12/16/2010	C	0.082	0.045	0.036	0.163	0.504	0.276	0.220	0.151	-0.798	0.160
1/19/2011	C	0.053	0.029	0.010	0.093	0.572	0.318	0.110	0.210	-1.287	0.753
5/19/2011	C	0.094	0.045	0.034	0.173	0.545	0.258	0.197	0.309	-0.812	0.194
5/25/2011	C	0.081	0.039	0.032	0.151	0.535	0.256	0.209	0.207	-0.893	0.141
5/4/2010	D	0.068	0.041	0.016	0.125	0.544	0.327	0.129	0.240	-1.039	0.657
5/18/2010	D	0.070	0.027	0.017	0.113	0.617	0.235	0.148	0.348	-1.198	0.328
1/19/2011	D	0.044	0.020	0.013	0.076	0.576	0.258	0.166	0.014	-1.430	0.311
2/16/2011	D	0.056	0.027	0.018	0.102	0.550	0.269	0.181	0.070	-1.192	0.279

6/17/2011	D	0.064	0.040	0.014	0.119	0.540	0.339	0.121	0.226	-1.067	0.727
6/29/2011	D	0.069	0.039	0.024	0.132	0.525	0.294	0.181	0.133	-0.976	0.344
7/6/2011	D	0.083	0.048	0.021	0.152	0.548	0.317	0.136	0.333	-0.905	0.600
8/3/2011	D	0.079	0.044	0.022	0.145	0.546	0.303	0.151	0.273	-0.938	0.490
8/11/2011	D	0.090	0.066	0.027	0.184	0.491	0.361	0.149	0.261	-0.692	0.627
8/24/2011	D	0.095	0.061	0.029	0.185	0.513	0.332	0.155	0.304	-0.723	0.537
8/31/2011	D	0.094	0.063	0.025	0.182	0.515	0.347	0.138	0.337	-0.737	0.653
9/20/2011	D	0.113	0.059	0.041	0.212	0.530	0.276	0.193	0.373	-0.648	0.252
4/17/2012	D	0.029	0.012	0.007	0.048	0.605	0.248	0.147	-0.115	-1.795	0.371
5/2/2012	D	0.049	0.032	0.010	0.092	0.535	0.351	0.113	0.107	-1.246	0.800
5/29/2012	D	0.086	0.046	0.019	0.152	0.567	0.305	0.127	0.398	-0.931	0.618
7/5/2012	D	0.060	0.040	0.017	0.116	0.515	0.341	0.144	0.099	-1.055	0.610
7/12/2012	D	0.057	0.035	0.014	0.107	0.536	0.330	0.134	0.127	-1.141	0.635
7/23/2012	D	0.068	0.046	0.016	0.130	0.523	0.355	0.123	0.223	-0.985	0.751
7/30/2012	D	0.065	0.040	0.020	0.124	0.521	0.319	0.160	0.119	-1.013	0.489
8/14/2012	D	0.088	0.053	0.028	0.169	0.518	0.316	0.167	0.256	-0.791	0.452
5/4/2010	E	0.046	0.028	0.010	0.084	0.543	0.336	0.122	0.058	-1.317	0.718
5/18/2010	E	0.066	0.039	0.021	0.126	0.519	0.312	0.169	0.109	-1.000	0.433
1/19/2011	E	0.036	0.023	0.005	0.064	0.563	0.362	0.075	0.138	-1.541	1.109
4/27/2011	E	0.064	0.033	0.015	0.112	0.570	0.297	0.133	0.237	-1.152	0.568
5/11/2011	E	0.070	0.043	0.015	0.128	0.545	0.338	0.117	0.286	-1.022	0.751
6/8/2011	E	0.071	0.046	0.017	0.134	0.529	0.341	0.130	0.234	-0.972	0.684
7/30/2012	E	0.060	0.038	0.018	0.116	0.516	0.328	0.156	0.080	-1.053	0.528
8/6/2012	E	0.081	0.045	0.027	0.153	0.531	0.293	0.176	0.225	-0.882	0.362
8/14/2012	E	0.096	0.051	0.036	0.184	0.525	0.279	0.196	0.284	-0.743	0.248
9/5/2012	E	0.075	0.050	0.024	0.149	0.506	0.333	0.161	0.169	-0.864	0.512
5/4/2010	F	0.046	0.028	0.013	0.088	0.526	0.321	0.154	-0.032	-1.265	0.521
5/18/2010	F	0.046	0.025	0.012	0.083	0.555	0.301	0.144	0.032	-1.341	0.522
1/19/2011	F	0.044	0.040	0.013	0.097	0.453	0.415	0.132	-0.110	-1.088	0.807
4/27/2011	F	0.075	0.043	0.023	0.142	0.531	0.304	0.165	0.204	-0.934	0.433
5/11/2011	F	0.069	0.042	0.017	0.128	0.537	0.327	0.136	0.217	-1.015	0.621
5/19/2011	F	0.101	0.056	0.031	0.189	0.537	0.298	0.165	0.362	-0.739	0.416
6/8/2011	F	0.080	0.055	0.020	0.155	0.514	0.355	0.131	0.270	-0.848	0.707
7/30/2012	F	0.090	0.059	0.032	0.180	0.498	0.327	0.175	0.225	-0.717	0.442
8/14/2012	F	0.096	0.059	0.037	0.191	0.501	0.307	0.192	0.244	-0.681	0.332
9/5/2012	F	0.079	0.065	0.027	0.170	0.462	0.379	0.159	0.135	-0.705	0.617

SOURCE WATER ORGANIC MATTER

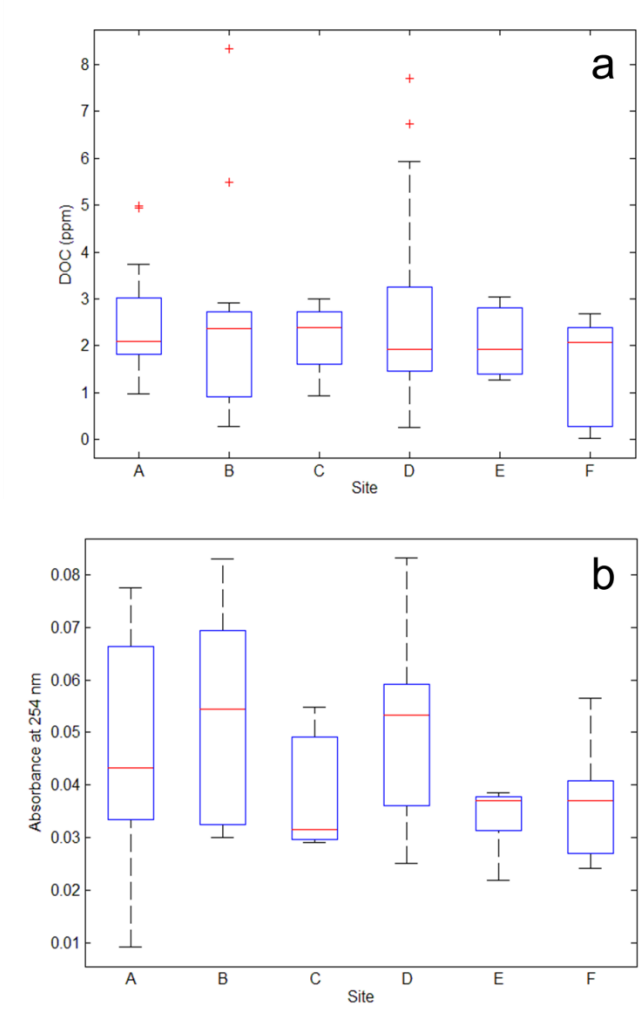

Figure A1: Boxplots of (a) DOC (ppm) concentration, and (b) UV Absorbance at 254 nm at each of the six sampling sites. Plots show median values, 75th and 25th quartiles (upper and lower ends of the box), most extreme non-outlier values (ends of whiskers), and outliers (+ signs).

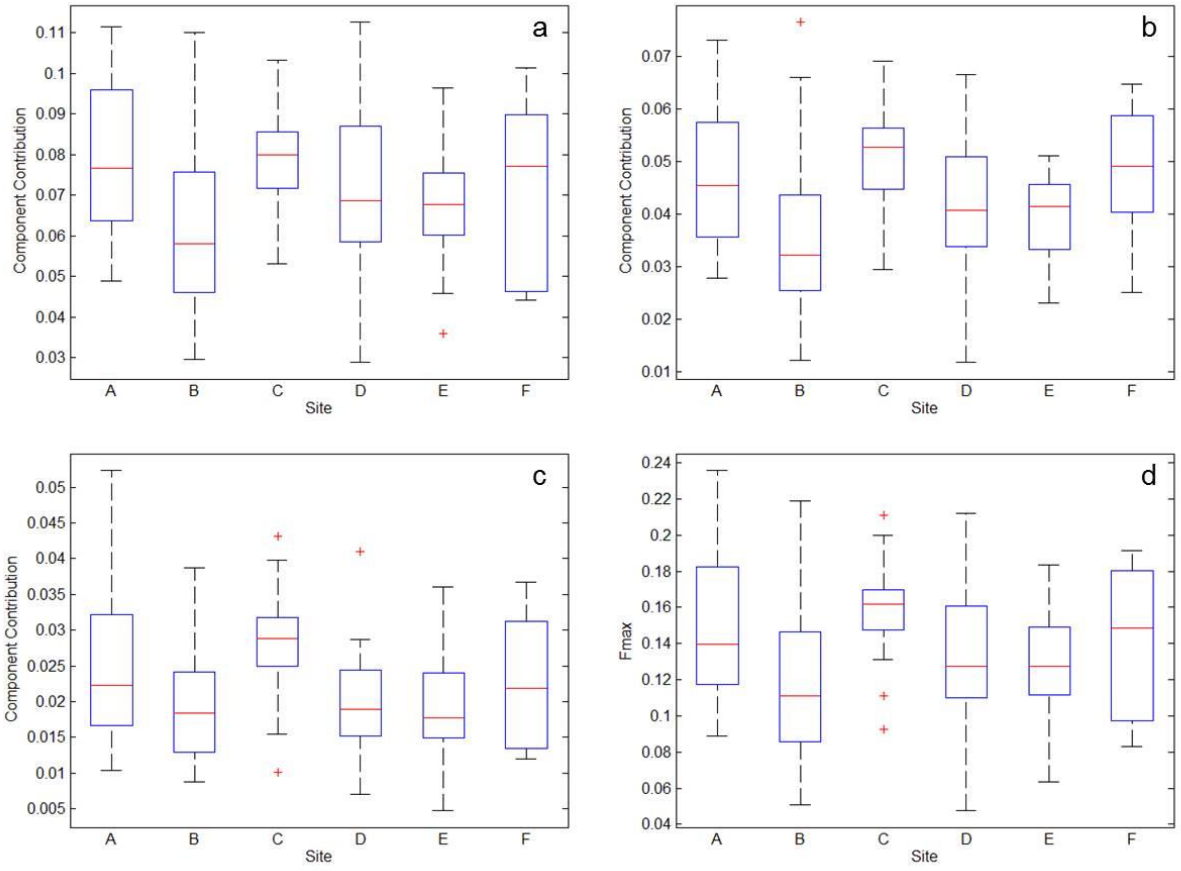

Figure A2: Boxplots of each of the individual PARAFAC Components and the total fluorescence, F_{max}, as follows: (a) C1, (b) C2, (c) C3, (d) F_{max}. Plots show median values, 75th and 25th quartiles (upper and lower ends of the box), minimum and maximum (non-outlier) values (ends of whiskers), and outliers (+ signs).

REGRESSION ANALYSIS

Results of the linear and log transformed regressions are shown in Table A2 and Table A3, respectively. PARAFAC components (C1, C2, C3), total fluorescence intensity (Fmax), and component fluorescence ratios (C1/Fmax, C2/Fmax, C3/Fmax) were all considered as possible fluorescence input variables. Multiple variations of input values were evaluated to avoid high collinearity among the fluorescence measurements (i.e. C1, C2, and C3), and the best fit model was chosen. Though the R^2 values do not indicate strong linear relationships, the residual

standard error (RSE) values indicate that most of the linear and log transformed regression models are comparable to those identified by Ged *et al.* (2015) to be best available models.

For example, the RSE for TTHM is 32.7 μg/L for the linear regression and 50.1 μg/L for the log transformed regression, which are both less than 60 μg/L, similar to the best TTHM models determined by Ged *et al.* (2015). However, the errors of 32.7 μg/L and 50.1 μg/L are still substantial compared to the TTHM MCL of 80 μg/L. Models with this level of error would not be useful for operators who need to maintain regulatory compliance. The results of the regression analysis in the present work are thus similar to prior work with regression models (Ged *et al.*, 2015; Obolensky and Singer, 2008; Westerhoff *et al.*, 2000; Kulkarni and Chellam, 2010), indicating that organic carbon characterization does not provide enough additional information to create regression models with adequate predictive power for DBP species-specific concentrations.

Table A2 provides a summary of the best linear regressions for each quantitative DBP parameter, including the adjusted R^2 and residual standard error (RSE) of the model. Both sets of component inputs (C1, C2, C3 and C1/F, C2/F, C3/F, F) were tested and the best model overall was chosen.

Table A2: Results of the Linear Regression Analyses of the source water constituents (bromide and NOM) and finished water parameters – TTHM (µg/L), $CHCl_3$ (µg/L), $CHBrCl_2$ (µg/L), $CHBr_2Cl$ (µg/L), $CHBr_3$ (µg/L), BIF, and percent brominated.

Model	Adjusted R^2	RSE
$TTHM = 908.05\ C2 + 8.87$	0.13	32.66
$CHCl_3 = -53.84Br + 322.30\ C1 + 3.12$	0.13	17.64
$CHBrCl_2 = 324.95\ C2 + 2.23$	0.07	15.90
$CHBr_2Cl = 75.02Br - 82.33UV_{254} - 74.61\frac{C1}{F_{max}}$	0.37	6.83
$CHBr_3 = 21.42Br - 19.28\ \frac{C1}{F_{max}} + 10.96$	0.34	1.97
$BIF = 3.46Br - 4.55\frac{C1}{F_{max}} - 2.45F_{max} + 3.26$	0.44	0.27
$\%Brom = 130.86Br - 251.94\ \frac{C1}{F_{max}} - 97.76F_{max} + 187.46$	0.29	16.42

Table A3 shows the results of the log transformed regression analyses, including the regression model, adjusted R^2, and residual standard error (RSE).

Table A3: Results of the Linear Log Transformed Function Analyses of the source water constituents (bromide and NOM) and finished water parameters – TTHM (µg/L), $CHCl_3$ (µg/L), $CHBrCl_2$ (µg/L), $CHBr_2Cl$ (µg/L), $CHBr_3$ (µg/L), BIF, and percent brominated.

Model	Adjusted R^2	RSE
$TTHM = 5.99(TOC)^{0.19}(F_{max})^{1.25}$	0.22	50.06
$CHCl_3 = 3.93(Br)^{-0.13}(DOC)^{0.17}(UV_{254})^{-0.30}(F_{max})^{1.29}$	0.25	29.06
$CHBrCl_2 = 10.96(TOC)^{0.72}(F_{max})^{1.42}(\frac{C3}{F_{max}})^{3.71}$	0.10	23.21
$CHBr_2Cl = -6.45(Br)^{1.57}(UV_{254})^{-3.08}$	0.12	1.37×10^5
$CHBr_3 = -13.30(Br)^{0.83}(UV_{254})^{1.49}\left(\frac{C1}{F_{max}}\right)^{-30.08}$	0.26	3.82×10^8
$BIF = 3.24(Br)^{0.19}(DOC)^{0.37}\left(\frac{C3}{F_{max}}\right)^{1.82}$	0.28	7.31
$\%Brom = 7.39(Br)^{0.12}(DOC)^{0.41}\left(\frac{C3}{F_{max}}\right)^{1.92}$	0.28	540

CLASSIFICATION TREES

Predicting TTHM in excess of 80% of the Maximum Contaminant Level (MCL)

The component classification tree (Figure 3.7a) has a sensitivity of 0.66, a specificity of 0.81, and an accuracy of 0.77. The component ratio classification tree (Figure 3.7b) has a sensitivity of 0.72, a specificity of 0.88 and an accuracy of 0.83. Sensitivity, specificity, and accuracy values were calculated using the confusion matrices in Table A4. In both cases the classification trees are fairly balanced (sensitivity similar to specificity) and show good fits of the data (high sensitivity, specificity, and accuracy values), however the component ratio tree (Figure 3.7b) shows a slightly better fit of the data set. According to the component classification tree (Figure 3.7a), instances are likely to exceed 64 μg/L TTHM when either the C2 value is high (≥ 0.04) and the DOC is high (≥ 4.0 mg/L), or when the C2 value is high (≥ 0.04), the DOC is low (< 2.9 mg/L), and the C3 value is high (≥ 0.03). In both cases, higher fluorescence intensity of the component signals are related to increased exceedance of the 64 μg/L threshold.

The component ratio classification tree (Figure 3.7b), on the other hand identifies three cases in which there is a likelihood of exceeding the threshold – (1) $C1/F_{max}$ is low (< 0.54), F_{max} is moderate (0.16 – 0.18), DOC is low (< 2.9 mg/L), and $C3/F_{max}$ is high (≥ 0.16), (2) $C1/F_{max}$ is low (< 0.54), DOC is low (< 2.9 mg/L), and F_{max} is high (≥ 0.18), and (3) $C1/F_{max}$ is low (< 0.54), F_{max} is high (≥ 0.11), and DOC is high (≥ 4.0 mg/L). In both the component and component ratio trees there are cases where lower DOC contributes to an increased likelihood of exceeding the 64 μg/L threshold. These results are somewhat counter-intuitive because NOM is a known DBP precursor. However, in all of these cases, there are at least 2 other NOM characterizations (mostly fluorescence measurements) that contribute to the increased likelihood

of exceedance outcome. The results indicate that in some cases the character of the NOM (described here by fluorescence character) may be more influential in meeting or exceeding a regulatory threshold than NOM quantity.

Predicting BIF Values in Excess of 0.75

The component classification tree has a sensitivity of 0.61, a specificity of 0.96, and an accuracy of 0.83, while the component ratio tree has a sensitivity of 0.53, a specificity of 0.94, and an accuracy of 0.80. Both trees show fairly good fits – high specificity and accuracy values and lower, but still reasonable sensitivity values. The lower sensitivity values indicate that the trees slightly under-predict exceeding the 0.75 BIF threshold. According to the component classification tree (Figure 3.8a), there is a likelihood of exceeding the 0.75 BIF threshold if the bromide concentration is high (≥ 40 μg/L), C1 is low (< 0.09), and C2 is high (≥ 0.05), or if bromide is high (≥ 40 μg/L), C2 is low (<0.05), and C1 is very low (<0.06). According to the component ratio classification tree (Figure 3.8b), there is a likelihood of exceeding the 0.75 BIF threshold if the bromide concentration is high (≥ 40 μg/L) and the $C3/F_{max}$ ratio is high (≥ 0.16).

Predicting THM Bromination in Excess of 50%

The component classification tree (Figure 3.9a) has a sensitivity of 0.76, a specificity of 0.83, and an accuracy of 0.80, and the component ratio classification tree (Figure 3.9b) has a sensitivity of 0.80, a specificity of 0.73, and an accuracy of 0.76. Both trees show fairly balanced results (similar sensitivity and specificity values), as well as overall good fits, as demonstrated by the relatively high accuracy, sensitivity, and specificity values. According to the component

classification tree (Figure 3.9a), there is a likelihood of exceeding 50% brominated THM (by mass) in three separate scenarios: namely when (1) the bromide is low (< 60 μg/L) and the UV_{254} is low (< 0.03); (2) the bromide is low (< 60 μg/L), UV_{254} is high (≥ 0.03), C2 is low (< 0.05), and C3 is high (≥ 0.02); and (3) bromide is high (≥ 60 μg/L) and C1 is low (< 0.08). The component ratio tree (Figure 3.9b) identifies only one scenario in which there is a likelihood of exceeding 50% brominated THM – low bromide (< 60 μg/L), low $C1/F_{max}$ ratio (< 0.55), and low UV_{254} (< 0.04). Overall, the classification trees performed very well, correctly classifying 76% to 83% of the 109 instances (accuracy values ranging from 0.76 to 0.83). Further, most classification trees had sensitivity and specificity values ranging from 0.53 to 0.96, indicating high true positive and true negative rates.

Confusion Matrices for all Classification Trees

Table A4: Confusion Matrices for Classification Trees for each of the four parameters – The left column shows matrices for the trees using components as inputs (C1, C2, C3) and the right column uses component ratios and total fluorescence intensity as inputs ($C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}). E row/column headings indicate "exceed", M row/column headings indicate "meet," rows show actual values (subscript "A"), and columns show predicted outcomes (subscript "P"). Each matrix shows the number of instances classified as true positive (top left), true negative (bottom right), false positive (bottom left), and false negative (top right), where positive is taken to be "exceed" and negative is taken to be "meet."

	C1, C2, C3			**$C1/F_{max}$, $C2/F_{max}$, $C3/F_{max}$, F_{max}**		
TTHM MCL		E_P	M_P		E_P	M_P
	E_A	5	15	E_A	16	4
	M_A	4	85	M_A	14	75
80% MCL		E_P	M_P		E_P	M_P
	E_A	19	10	E_A	21	8
	M_A	15	65	M_A	10	70
0.75 BIF		E_P	M_P		E_P	M_P
	E_A	23	15	E_A	20	18
	M_A	3	68	M_A	4	67
50% BrTHM		E_P	M_P		E_P	M_P
	E_A	35	11	E_A	37	9
	M_A	11	52	M_A	17	46

Appendix B

SUPPLEMENTAL INFORMATION FOR CHAPTER 4 – FLUORESCENCE CHARACTERIZATION OF NATURAL ORGANIC MATTER AND FOULING IN A FULL-SCALE REVERSE OSMOSIS MEMBRANE TREATMENT PLANT

Significance Tests for Differences between Sample Distributions: Turbidity, TOC, Conductivity

The following three tables show the results of the Wilcoxon Rank Sum Test for differences between two sample distributions. The results were obtained using R. The tables show the two sample distributions that are being compared in the left-most column (i.e. SW Period 1 and SW Period 2 are written as "SW (P1 vs P2)." When the same type of samples are being compared between the two periods, i.e. SW Period 1 and SW Period 2, the samples are not paired. However, when the samples are from the same period, but come from two different sampling locations, i.e. SW Period 1 and CF Period 1, the samples are paired. The tables also contain p-values to determine statistical significance (significant if $p < 0.05$), the Hodge-Lehmann estimator (the non-parametric difference between the two sample distributions), and the 95% Confidence Interval. Results are shown for Turbidity (Table B1), TOC (Table B2), and Conductivity (Table B3).

Table B1: Results of Wilcoxon Rank Sum Tests for Turbidity

Sample	P-value	Hodges-Lehmann estimator	CI (95%)
SW (P1 vs P2)	0.009	-4.30	[-24.40, -1.36]
CF (P1 vs P2)	0.052	-2.83	[-8.86, 2.52]
UF (P1 vs P2)	0.647	-0.11	[-0.45, 0.41]
SW v CF (P1)	0.438	-0.71	[-4.91, 0.69]
CF v UF (P1)	0.031	3.16	[0.89, 9.26]
SW v UF (P1)	0.03125	2.49	[1.21, 4.35]
SW v CF (P2)	1.0	0.05	[-1.19, 23.65]
CF v UF (P2)	0.0625	6.03	[3.17, 12.24]
SW v UF (P2)	0.0625	9.12	[4.51, 26.82]

Table B2: Results of Wilcoxon Rank Sum Tests for TOC

Sample	P-value	Hodges-Lehmann estimator	CI (95%)
SW (P1 vs P2)	0.030	0.53	[0.07, 0.95]
CF (P1 vs P2)	0.004	0.59	[0.29, 0.95]
UF (P1 vs P2)	0.017	0.41	[0.11, 0.70]
SW v CF (P1)	0.399	0.06	[-0.07, 0.17]
CF v UF (P1)	0.059	0.35	[0.30, 0.40]
SW v UF (P1)	0.031	0.34	[0.14, 0.43]
SW v CF (P2)	0.584	0.11	[-0.05, 0.46]
CF v UF (P2)	0.201	0.12	[-0.02, 0.18]
SW v UF (P2)	0.313	0.16	[-0.07, 0.61]

Table B1: Results of Wilcoxon Rank Sum Tests for Conductivity

Sample	P-value	Hodges-Lehmann estimator	CI (95%)
SW (P1 vs P2)	0.052	-146.4	[-221.7, 13.0]
CF (P1 vs P2)	0.178	-119.85	[-232.4, 27.9]
UF (P1 vs P2)	0.429	-128.2	[-194.8, 70.6]
SW v CF (P1)	0.063	-33.75	[-75.3, 7.8]
CF v UF (P1)	1.0	-5.90	[-58.1, 11.5]
SW v UF (P1)	0.063	-48.30	[-89.4, 12.4]
SW v CF (P2)	0.813	-18.25	[-82.3, 45.8]
CF v UF (P2)	1.0	1.15	[-66.8, 75.3]
SW v UF (P2)	0.438	-18.65	[-49.5, 35.9]

Fluorescence EEM-PARAFAC Results

Table B2: Summary of the Fluorescence EEM-PARAFAC Results

Sample ID	Date	Sample Location	C1	C2	C3
1	9.1.2014	SW	0.035518	0.01737	0.018311
2	9.1.2014	CF	0.032823	0.016243	0.015593
3	9.1.2014	UF	0.036563	0.016329	0.015226
4	9.1.2014	RO R1	0.110594	0.069161	0.081583
5	9.1.2014	RO R2	2.07E-15	0	1.17E-13
6	9.4.2014	SW	0.032557	0.016955	0.017077
7	9.4.2014	CF	0.032321	0.016785	0.015957
8	9.4.2014	UF	0.035424	0.017139	0.015215
9	9.4.2014	RO R1	0.103502	0.065691	0.076728
10	9.4.2014	RO R2	0	0	6.84E-15
11	9.24.2014	SW	0.056398	0.021879	0.040313
12	9.24.2014	CF	0.056633	0.019535	0.037412
13	9.24.2014	UF	0.064795	0.020113	0.031849
14	9.24.2014	RO R1	0.194008	0.093276	0.130279

15	9.24.2014	RO R2	0.0017	0.000819	0
16	10.3.2014	SW	0.064605	0.026585	0.039923
17	10.3.2014	CF	0.084327	0.025659	0.029103
18	10.3.2014	UF	0.081766	0.023571	0.027716
19	10.3.2014	RO R1	0.21217	0.104893	0.135992
20	10.3.2014	RO R2	0	1.48E-06	0
21	10.20.2014	SW	0.063718	0.027446	0.050437
22	10.20.2014	CF	0.063634	0.025638	0.045091
23	10.20.2014	UF	0.061385	0.021483	0.042403
24	10.20.2014	RO R1	0.194611	0.105746	0.147476
25	10.20.2014	RO R2	0	4.13E-05	0.000137
26	10.23.2014	SW	0.063851	0.027052	0.04922
27	10.23.2014	CF	0.067589	0.026965	0.045618
28	10.23.2014	UF	0.061751	0.022385	0.042585
29	10.23.2014	RO R1	0.195827	0.106614	0.14638
30	10.23.2014	RO R2	0	1.15E-05	4.43E-05
31	11.18.2014	SW	0.089279	0.028074	0.059362
32	11.18.2014	CF	0.075797	0.027533	0.047464
33	11.18.2014	UF	0.072295	0.024943	0.047728
34	11.18.2014	RO R1	0.2098	0.110577	0.150441
35	11.18.2014	RO R2	0.001766	0.001102	0
36	12.04.2014	SW	0.108466	0.040181	0.075455
37	12.04.2014	CF	0.103492	0.040808	0.068001
38	12.04.2014	UF	0.094774	0.034128	0.059949
39	12.04.2014	RO R1	0.223741	0.147583	0.19268
40	12.04.2014	RO R2	0	6.6E-06	0
41	1.27.2015	SW	0.100713	0.037528	0.061075
42	1.27.2015	CF	0.086064	0.032756	0.057122
43	1.27.2015	UF	0.086715	0.028975	0.052341
44	1.27.2015	RO R1	0.199999	0.125055	0.164642
45	1.27.2015	RO R2	4.99E-06	1.51E-05	1.31E-06
46	2.10.2015	SW	0.115241	0.045381	0.082141
47	2.10.2015	CF	0.115146	0.043607	0.079998
48	2.10.2015	UF	0.114311	0.039504	0.076701
49	2.10.2015	RO R1	0.258388	0.15944	0.221221
50	2.10.2015	RO R2	0.003157	0.001369	0.007446
51	5.16.2015	SW	0.139583	0.049401	0.095883
52	5.16.2015	CF	0.146698	0.050323	0.089127
53	5.16.2015	UF	0.140434	0.047517	0.083983
54	5.16.2015	RO R1	0.26843	0.180095	0.234006
55	5.16.2015	RO R2	0.011471	0.002549	0.005783

Table B5: Wilcoxon Rank Sum Tests for Peak EEM Fluorescence Intensities

Sample	P-value	Hodges-Lehmann estimator	CI (95%)
SW (P1 v P2)	0.004	-0.067	[-0.103, -0.040]
CF (P1 v P2)	0.017	-0.060	[-0.103, -0.015]
UF (P1 v P2)	0.009	-0.054	[-0.098, -0.013]
SW v CF (P1)	0.688	0.001	[-0.026, 0.004]
CF v UF (P1)	1.0	5.84E-07	[-0.010, 0.006]
SW v UF (P1)	0.688	-0.001	[-0.021, 0.005]
SW v CF (P2)	0.188	0.009	[-0.006, 0.019]
CF v UF (P2)	0.125	0.005	[-0.000, 0.012]
SW v UF (P2)	0.063	0.012	[0.002, 0.023]

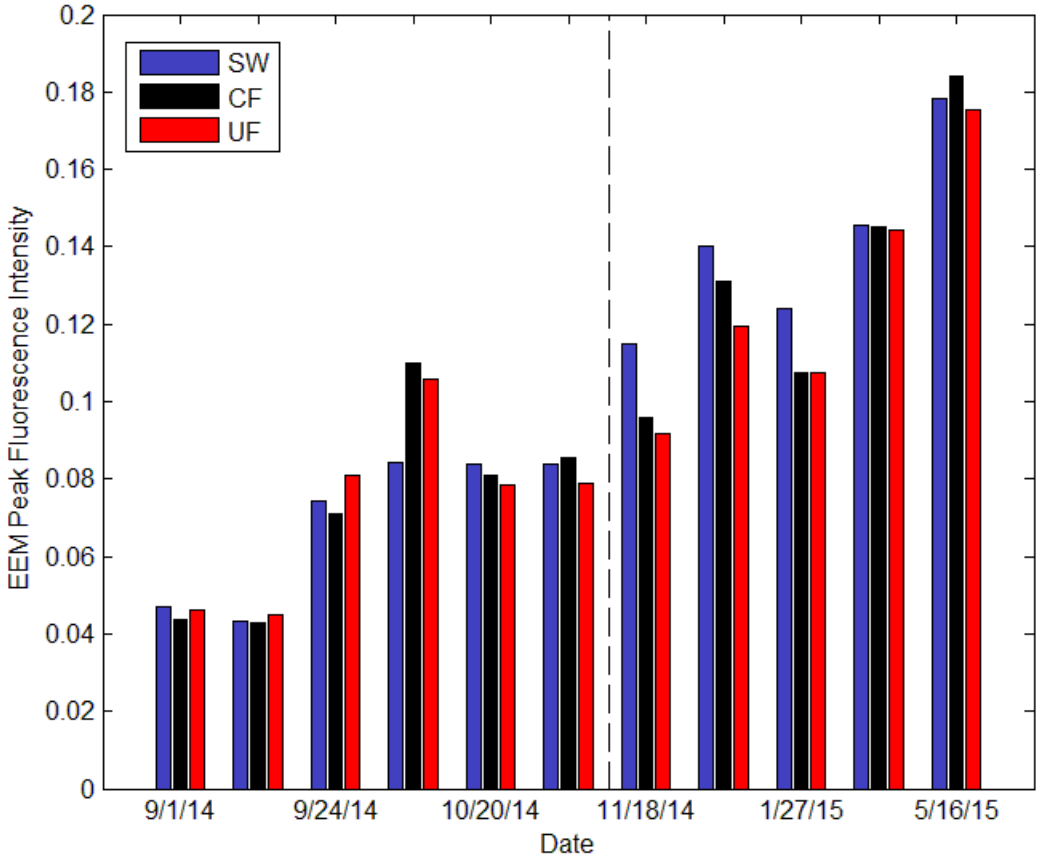

Figure B1: Plot of EEM peak fluorescence intensity for pre-membrane samples (SW, CF, UF) throughout field study. Period 1 and Period 2 are divided by vertical black dotted line.

Table B6: Wilcoxon Rank Sum Tests for EEM-PARAFAC Components

Sample	P-value	Hodges-Lehmann estimator	CI (95%)
C1 (P1 vs P2)	2.314E-08	-0.051	[-0.067, -0.033]
C2 (P1 vs P2)	5.785E-08	-0.017	[-0.022, -0.011]
C3 (P1 vs P2)	2.314E-08	-0.037	[-0.046, -0.026]
C1 P1 (SW v CF)	0.6875	-0.0005	[-0.0197,0.0027]
C1 P1 (CF v UF)	0.6875	-0.0006	[-0.008, 0.006]
C1 P1 (SW v UF)	0.3125	-0.003	[-0.017, 0.002]
C1 P2 (SW v CF)	0.3125	0.005	[-0.007, 0.015]
C1 P2 (CF v UF)	0.125	0.004	[-0.001, 0.009]
C1 P2 (SW v UF)	0.125	0.008	[-0.001, 0.0169]
C2 P1 (SW v CF)	0.031	0.001	[0.0001, 0.002]
C2 P1 (CF v UF)	0.438	0.002	[-0.001, 0.005]
C2 P1 (SW v UF)	0.063	0.003	[-0.0002, 0.006]
C2 P2 (SW v CF)	0.625	0.0006	[-0.0009, 0.005]
C2 P2 (CF v UF)	0.063	0.004	[0.003, 0.007]
C2 P2 (SW v UF)	0.063	0.005	[0.002, 0.009]
C3 P1 (SW v CF)	0.031	0.004	[0.001, 0.011]
C3 P1 (CF v UF)	0.031	0.002	[0.0004, 0.006]
C3 P1 (SW v UF)	0.031	0.007	[0.002, 0.012]
C3 P2 (SW v CF)	0.063	0.007	[0.002, 0.012]
C3 P2 (CF v UF)	0.125	0.004	[-0.000, 0.008]
C3 P2 (SW v UF)	0.063	0.010	[0.005, 0.016]

References

Abouleish, M. Y. and Wells, M. J. 2015. Trihalomethane formation potential of aquatic and terrestrial fulvic and humic acids: Sorption on activated carbon. *Sci Total Environ*, 521-522**,** 293-304.

Acero, J. L., Piriou, P. and von Gunten, U. 2005. Kinetics and mechanisms of formation of bromophenols during drinking water chlorination: assessment of taste and odor development. *Water Res.*, 39**,** 2979-93.

Airey, D., Yao, S., Wu, J., Chen, V., Fane, A. G. and Pope, J. M. 1998. An investigation of concentration polarization phenomena in membrane filtration of colloidal silical suspensions by NMR micro-imaging *Journal of Membrane Science*, 145**,** 145-158.

Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.*, 19**,** 716-723.

Al-Omari, A., Fayyad, M. and Qader, A. 2004. Modeling trihalomethane formation for jabal amman water supply in jordan. *Environ. Model Assess.*, 9**,** 245-252.

Allard, S., Tan, J., Joll, C. A. and von Gunten, U. 2015. Mechanistic Study on the Formation of Cl-/Br-/I-Trihalomethanes during Chlorination/Chloramination Combined with a Theoretical Cytotoxicity Evaluation. *Environ Sci Technol*, 49**,** 11105-14.

Amy, G. 2008. Fundamental understanding of organic matter fouling of membranes. *Desalination*, 231**,** 44-51.

Amy, G. L., Chadik, P. A. and Chowdhury, Z. K. 1987. Developing Models for Predicting Trihalomethane Formation Potential and Kinetics. *J. Am. Water Works Assoc.*, 79**,** 89-97.

Ang, W. S., Tiraferri, A., Chen, K. L. and Elimelech, M. 2011. Fouling and cleaning of RO membranes fouled by mixtures of organic foulants simulating wastewater effluent. *Journal of Membrane Science*, 376**,** 196-206.

Archer, A. D. and Singer, P. C. 2006. Effect of SUVA and enhanced coagulation on removal of TOX precursors. *J. Am Water Works Assoc.*, 98**,** 97-107.

Arora, M. L. and Trompeter, K. M. 1983. Fouling of RO Membranes in Wastewater Applications. *Desalination*, 48**,** 299-319.

Ates, N., Kitis, M. and Yetis, U. 2007. Formation of chlorination by-products in waters with low SUVA--correlations with SUVA and differential UV spectroscopy. *Water Res.*, 41**,** 4139-48.

Awad, J., van Leeuwen, J., Chow, C., Drikas, M., Smernik, R. J., Chittleborough, D. J. and Bestland, E. 2016. Characterization of Dissolved Organic Matter for Prediction of Trihalomethane Formation Potential in Surface and Sub-surface Waters. *Journal of Hazardous Materials*, http://dx.doi.org/10.1016/j.jhazmat.2016.01.030.

Babcock, D. B. and Singer, P. C. 1979. Chlorination and Coagulation of Humic and Fulvic Acids. *J. Am. Water Works Assoc.*, 71**,** 149-152.

Bae, H., Kim, S. and Kim, Y. J. 2006. Decision algorithm based on data mining for coagulant type and dosage in water treatment systems. *Wat. Sci. Tech.*, 53**,** 321-329.

Baghoth, S. A., Sharma, S. K. and Amy, G. L. 2011. Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation-emission matrices and PARAFAC. *Water Res.*, 45**,** 797-809.

Bauer, D. F. 1972. Constructing Confidence Sets Using Rank Statistics. *Journal of the American Statistical Association*, 67**,** 687-690.

Bazri, M. M., Martijn, B., Kroesbergen, J. and Mohseni, M. 2016. Impact of anionic ion exchange resins on NOM fractions: Effect on N-DBPs and C-DBPs precursors. *Chemosphere*, 144**,** 1988-95.

Becker, W., Stanford, B. and Rosenfeldt, E. 2013. Guidance on Complying With Stage 2 D/DBP Regulation. *Water Research Foundation*, Web Report #4427.

Bieroza, M., Baker, A. and Bridgeman, J. 2009. Relating freshwater organic matter fluorescence to organic carbon removal efficiency in drinking water treatment. *Science of the Total Environment*, 407**,** 1765-74.

Bolto, B., Dixon, D., Eldridge, R., King, S. and Linge, K. 2002. Removal of natural organic matter by ion exchange. *Water Research* 36**,** 5057–5065.

Bonnélye, V., Guey, L. and Del Castillo, J. 2008. UF/MF as RO pre-treatment: the real benefit. *Desalination*, 222**,** 59-65.

Boorman, G. A., Dellarco, V., Dunnick, J. K., Chapin, R. E., Hunter, S., Hauchman, F., Gardner, H., Cox, M. and Sills, R. C. 1999. Drinking Water Disinfection Byproducs: Review and Approach to Toxicity Evaluation. *Environmental Health Perspectives* 107**,** 207-217.

Brehant, A., Bonnelye, V. and Perez, M. 2002. Comparison of MF/UF pretreatment with conventiotial filtration prior to RO membranes for surface seawater desalination. *Desalination* 144**,** 353-360.

Bro, R. 1997. PARAFAC. Tutorial and applications. *Chemometr. Intell. Lab*, 38**,** 149-171.

Cabaniss, S. E. and Shuman, M. S. 1987. Synchronous Fluorescence Spectra of Natural Waters: Tracing Sources of Dissolved Organic Matter. *Marine Chemistry*, 21**,** 37-50.

Cantor, K. P., Villanueva, C. M., Silverman, D. T., Figueroa, J. D., Real, F. X., Garcia-Closas, M., Malats, N., Chanock, S., Yeager, M., Tardon, A., Garcia-Closas, R., Serra, C., Carrato, A., Castano-Vinyals, G., Samanic, C., Rothman, N. and Kogevinas, M. 2010. Polymorphisms in GSTT1, GSTZ1, and CYP2E1, disinfection by-products, and risk of bladder cancer in Spain. *Environ. Health Perspect.*, 118**,** 1545-50.

Carstea, E. M., Baker, A., Bieroza, M., Reynolds, D. M. and Bridgeman, J. 2014. Characterisation of dissolved organic matter fluorescence properties by PARAFAC analysis and thermal quenching. *Water Research*, 61**,** 152-61.

Chambers, J. M. and Hastie, T. J. 1992. *Statistical Models in S*, Wadsworth and Brooks/Cole, Pacific Grove, CA**,** 46–47.

Chang, E. E., Lin, Y. P. and Chiang, P. C. 2001. Effects of bromide on formation of THMs and HAAs. *Chemosphere*, 43**,** 1029-1034.

Chen, B. and Westerhoff, P. 2010. Predicting disinfection by-product formation potential in water. *Water Res.*, 44**,** 3755-62.

Chen, F., Peldszus, S., Peiris, R. H., Ruhl, A. S., Mehrez, R., Jekel, M., Legge, R. L. and Huck, P. M. 2014a. Pilot-scale investigation of drinking water ultrafiltration membrane fouling rates using advanced data analysis techniques. *Water Res.*, 48**,** 508-18.

Chen, P., Pan, D. and Mao, Z. 2014b. Development of a portable laser-induced fluorescence system used for in situ measurements of dissolved organic matter. *Optics & Laser Technology*, 64**,** 213-219.

Chen, W. and Weisel, C. 1998. Halogenated DBP concentrations in a distribution system. *J. Am. Water Works Assoc.*, 90**,** 151-163.

Chen, W., Westerhoff, P., Leenheer, J. A. and Booksh, K. 2003. Fluorescence Excitation-Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter. *Environ. Sci. Technol.* , 37**,** 5701-5710.

Chin, Y.-P., Alken, G. and O'Loughlin, E. 1994. Molecular Weight, Polydispersity, and Spectroscopic Properties of Aquatic Humic Substances. *Environmental Science and Technology* 28**,** 1853-1858.

Cho, J., Amy, G. and Pellegrino, J. 2000. Membrane filtration of natural organic matter: factors and mechanisms affecting rejection and flux decline with charged ultrafiltration (UF) membrane. *J. Membr. Sci.*, 164**,** 89–110.

Choi, J., Choi, W., Kim, H., Alaud-din, A., Cho, K. H., Kim, J. H., Lim, H., Lovitt, R. W. and Chang, I. S. 2014. Fluorescence imaging for biofoulants detection and monitoring of biofouled strength in reverse osmosis membrane. *Analytical Methods*, 6**,** 993.

Chowdhury, S., Champagne, P. and James McLellan, P. 2010. Investigating effects of bromide ions on trihalomethanes and developing model for predicting bromodichloromethane in drinking water. *Water Res.*, 44**,** 2349-59.

Chowdhury, S., Champagne, P., McLellan, J. 2009. Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review. *Sci. Total Environ.* , 407 4189–4206.

Chu, W., Gao, N., Krasner, S. W., Templeton, M. R. and Yin, D. 2012. Formation of halogenated C-, N-DBPs from chlor(am)ination and UV irradiation of tyrosine in drinking water. *Environ Pollut*, 161**,** 8-14.

Coble, P. G. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry*, 51**,** 325-346.

Cornelissen, E. R., Moreau, N., Siegers, W. G., Abrahamse, A. J., Rietveld, L. C., Grefte, A., Dignum, M., Amy, G. and Wessels, L. P. 2008. Selection of anionic exchange resins for removal of natural organic matter (NOM) fractions. *Water Res*, 42**,** 413-23.

Cowman, G. A. a. S., Philip C . 1996. Effect of Bromide Ion on Haloacetic Acid Speciation Resulting from Chlorination and Chloramination of Aquatic Humic Substances. *Environ. Sci. Technol.***,** 16-24.

Criquet, J., Allard, S., Salhi, E., Joll, C. A., Heitz, A. and von Gunten, U. 2012. Iodate and iodo-trihalomethane formation during chlorination of iodide-containing waters: role of bromide. *Environ Sci Technol*, 46**,** 7350-7.

Cuss, C. W. and Gueguen, C. 2014. Relationships between molecular weight and fluorescence properties for size-fractionated dissolved organic matter from fresh and aged sources. *Water Res*, 68C**,** 487-497.

Danileviciute, A., Grazuleviciene, R., Vencloviene, J., Paulauskas, A. and Nieuwenhuijsen, M. J. 2012. Exposure to drinking water trihalomethanes and their association with low birth weight and small for gestational age in genetically susceptible women. *International Journal of Environmental Research and Public Health*, 9**,** 4470-85.

Diagne, F., Malaisamy, R., Boddie, V., Holbrook, R. D., Eribo, B. and Jones, K. L. 2012. Polyelectrolyte and silver nanoparticle modification of microfiltration membranes to mitigate organic and bacterial fouling. *Environ Sci Technol*, 46**,** 4025-33.

Do, T. D., Chimka, J. R. and Fairey, J. L. 2015. Improved (and Singular) Disinfectant Protocol for Indirectly Assessing Organic Precursor Concentrations of Trihalomethanes and Dihaloacetonitriles. *Environ Sci Technol*, 49**,** 9858-65.

Durmishi, B. H., Reka, A. A., Gjuladin – Hellon, T., Ismaili, M., Srbinovski, M. and Shabani, A. 2015. Disinfection of Drinking Water and Trihalomethanes: A Review. *International Journal of Advanced Research in Chemical Science* 2**,** 45- 56.

Edzwald, J. K., Becker, W. C. and Wattier, K. L. 1985. Surrogate Parameters for Monitoring Organic Matter and THM Precursors. *J. Am. Water Works Assoc.*, 77**,** 122-132.

Elshorbagy, W., Abu-Qdais, H., Elsheamy, M. 2000. Simulation of THM species in water distribution systems. *Water Res.*, 34**,** 3431-3439.

EPA 1995. Method 551.1: Determination of Chlorination Disinfection Byproducts, Chlorinated Solvents, and Halogenated Pesticides/Herbicides in Drinking Water by Liquid-Liquid Extraction and Gas Chromatography With Electron-Capture Detection. *Revision 1.0*.

EPA 1999. Enhanced Coagulation and Enhanced Precipitative Softening Guidance Manual. *Office of Water (4607) EPA 815-R-99-012*.

EPA 2006. National Primary Drinking Water Regulations: Stage 2 Disinfectants and Disinfection Byproducts Rule. *Federal Register*, 71**,** 388-493.

EPA 2009. Method 415.3: Determination of Total Organic Carbon and Specific UV Absorbance at 254 nm in Source Water and Drinking Water.

EPA 2010. Comprehensive Disinfectants and Disinfection Byproducts Rules (Stage 1 and Stage 2): Quick Reference Guide. *Office of Water (4606M) EPA 816-F-10-080*.

Francis, R. A., Small, M. J. and VanBriesen, J. M. 2009. Multivariate distributions of disinfection by-products in chlorinated drinking water. *Water Res.*, 43**,** 3453-68.

Francis, R. A., VanBriesen, J. M. and Small, M. J. 2010. Bayesian Statistical Modeling of Disinfection Byproduct (DBP) Bromine Incorporation in the ICR Database. *Environ. Sci. Technol.*, 44**,** 1232-1239.

Frimmel, F. H. 1998 Characterization of natural organic matter as major constituents in aquatic systems. *Journal of Contaminant Hydrology* 35**,** 201–216.

Gallard, H., Pellizzari, F., Croué, J. P. and Legube, B. 2003. Rate constants of reactions of bromine with phenols in aqueous solution. *Water Res.*, 37**,** 2883-2892.

Ged, E. C. and Boyer, T. H. 2014. Effect of seawater intrusion on formation of bromine-containing trihalomethanes and haloacetic acids during chlorination. *Desalination*, 345**,** 85-93.

Ged, E. C., Chadik, P. A. and Boyer, T. H. 2015. Predictive capability of chlorination disinfection byproducts models. *J. Environ. Manage*, 149**,** 253-62.

Gloor, R. and Leidner, H. 1979. Universal Detector for Monitoring Organic Carbon in Liquid Chromatography. *Analytical Chemistry*, 51**,** 645-647.

Goldman, J. H., Rounds, S. A., Keith, M. K. and Sobieszczyk, S. 2014. Investigating Organic Matter in Fanno Creek, Oregon, Part 3 of 3: Identifying and Quantifying Sources of Organic Matter to an Urban Stream. *Journal of Hydrology*.

Gould, J. P., Fitchhorn, L. E. and Urheim, E. 1983. Formation of brominated trihalomethanes: extent and kinetics. *Water Chlorination: Environmental Impact and Health Effects*, Ann Arbor Science Publishers, Ann Arbor, MI**,** 297-310.

Gray, S. R., Dow, N., Orbell, J. D., Tran, T. and Bolto, B. A. 2011. The significance of interactions between organic compounds on low pressure membrane fouling. *Water Sci Technol*, 64**,** 632-639.

Green, S. T., Small, M. J. and Casman, E. A. 2009. Determinants of National Diarrheal Disease Burden. *Environ. Sci. Technol.*, 43**,** 993-999.

Grelot, A., Grelier, P., Vincelet, C., Brüss, U. and Grasmick, A. 2010. Fouling characterisation of a PVDF membrane. *Desalination*, 250**,** 707-711.

Guastalli, A. R., Simon, F. X., Penru, Y., de Kerchove, A., Llorens, J. and Baig, S. 2013. Comparison of DMF and UF pre-treatments for particulate material and dissolved organic matter removal in SWRO desalination. *Desalination*, 322**,** 144-150.

Gyparakis, S. and Diamadopoulos, E. 2007. Formation and Reverse Osmosis Removal of Bromate Ions during Ozonation of Groundwater in Coastal Areas. *Separation Science and Technology*, 42**,** 1465-1476.

Haag, W. R. and Holgne, J. 1983. Ozonation of Bromide-Containing Waters: Kinetics of Formation of Hypobromous Acid and Bromate. *Environ. Sci. Technol.*, 17**,** 261-267.

Handke, P. 2008. Trihalomethane speciation and the relationship to elevated total dissolved solid concentrations affecting drinking water quality at systems utilizing the Monongahela River as a primary source during the 3rd and 4th quarters of 2008. *Pennsylvania Department of Environmental Protection*.

Harrington, G. W., Chowdhury, Z. K. and Owen, D. M. 1992. Developing a Computer Model to Simulate DBP Formation During Water Treatment. *J. Am. Water Works Assoc.*, 84**,** 78-87.

Harshman, R. A. and Lundy, M. E. 1994. PARAFAC: Parallel factor analysis. *Comput. Stat. Data An.*, 18.

Harvey, R., Murphy, H. M., McBean, E. A. and Gharabaghi, B. 2015. Using Data Mining to Understand Drinking Water Advisories in Small Water Systems: a Case Study of Ontario First Nations Drinking Water Supplies. *Water Resour. Manag.*, 29**,** 5129-5139.

He, S., Yan, M. and Korshin, G. V. 2015. Spectroscopic examination of effects of iodide on the chloramination of natural organic matter. *Water Res*, 70**,** 449-57.

He, W. and Hur, J. 2015. Conservative behavior of fluorescence EEM-PARAFAC components in resin fractionation processes and its applicability for characterizing dissolved organic matter. *Water Res.*, 83**,** 217-26.

He, X. S., Xi, B. D., Li, X., Pan, H. W., An, D., Bai, S. G., Li, D. and Cui, D. Y. 2013. Fluorescence excitation-emission matrix spectra coupled with parallel factor and regional integration analysis to characterize organic matter humification. *Chemosphere*, 93**,** 2208-15.

Helsel, D. R. 1990. Less than obvious - statistical treatment of data below the detection limit. *Environ Sci Technol*, 24**,** 1766-1774.

Her, N., Amy, G., McKnight, D., Sohn, J. and Yoon, Y. 2003. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. *Water Research*, 37**,** 4295-4303.

Herzberg, M. and Elimelech, M. 2007. Biofouling of reverse osmosis membranes: Role of biofilm-enhanced osmotic pressure. *Journal of Membrane Science*, 295**,** 11-20.

Hoek, E. M. V. and Elimelech, M. 2003. Cake-Enhanced Concentration Polarization: A New Fouling Mechanism for Salt-Rejecting Membranes. *Environmental Science and Technology*, 37**,** 5581-5588.

Hoek, E. M. V., Kim, A. S. and Elimelech, M. 2002. Influence of Crossflow Membrane Filter Geometry and Shear Rate on Colloidal Fouling in Reverse Osmosis and Nanofiltration Separations. *Environmental Engineering Science*, 19**,** 357-372.

Hong, S. and Elimelech, M. 1997. Chemical and physical aspects of natural organic matter (NOM) fouling of nanofiltration membranes. *Journal of Membrane Science* 132 159-181.

Howe, K. J. and Clark, M. M. 2002. Fouling of microfiltration and ultrafiltration membranes by natural waters. *Environ Sci Technol*, 36**,** 3571–3576.

Hsu, S. and Singer, P. C. 2010. Removal of bromide and natural organic matter by anion exchange. *Water Res*, 44**,** 2133-40.

Hua, B., Veum, K., Koirala, A., Jones, J., Clevenger, T. and Deng, B. 2006a. Fluorescence fingerprints to monitor total trihalomethanes and N-nitrosodimethylamine formation potentials in water. *Environmental Chemistry Letters*, 5**,** 73-77.

Hua, G. and Reckhow, D. A. 2007a. Characterization of Disinfection Byproduct Precursors Based on Hydrophobicity and Molecular Size. *Environ. Sci. Technol.*, 41**,** 3309-3315.

Hua, G. and Reckhow, D. A. 2007b. Comparison of disinfection byproduct formation from chlorine and alternative disinfectants. *Water Res*, 41**,** 1667-78.

Hua, G. and Reckhow, D. A. 2012. Evaluation of bromine substitution factors of DBPs during chlorination and chloramination. *Water Res*, 46**,** 4208-16.

Hua, G., Reckhow, D. A. and Abusallout, I. 2015. Correlation between SUVA and DBP formation during chlorination and chloramination of NOM fractions from different sources. *Chemosphere*, 130**,** 82-9.

Hua, G., Reckhow, D. A. and Kim, J. 2006b. Effect of Bromide and Iodide Ions on the Formation and Speciation of Disinfection Byproducts during Chlorination. *Environ. Sci. Technol.*, 40**,** 3050-3056.

Hur, J., Shin, J., Kang, M. and Cho, J. 2014. Tracking variations in fluorescent-dissolved organic matter in an aerobic submerged membrane bioreactor using excitation-emission matrix spectra combined with parallel factor analysis. *Bioprocess Biosyst Eng*.

Ivancev-Tumbas, I. 2014. The fate and importance of organics in drinking water treatment: a review. *Environ Sci Pollut Res Int*, 21**,** 11794-810.

Jacob Daniel Hosen, McDonough, O. T., Febria, C. M. and Palmer, M. A. 2014. Dissolved Organic Matter Quality and Bioavailability Changes across an Urbanization Gradient in Headwater Streams. *Environ. Sci. Technol.* .

Johnstone, D. W., Sanchez, N. P. and Miller, C. M. 2009. Parallel Factor Analysis of Excitation–Emission Matrices to Assess Drinking Water Disinfection Byproduct Formation During a Peak Formation Period. *Environ Eng Sci*, 26**,** 1551 - 1559.

Junaidi, M. U., Leo, C. P., Kamal, S. N. and Ahmad, A. L. 2013. Fouling mitigation in humic acid ultrafiltration using polysulfone/SAPO-34 mixed matrix membrane. *Water Sci Technol*, 67**,** 2102-9.

Jutaporn, P., Singer, P. C., Cory, R. M. and Coronell, O. 2016. Minimization of short-term low-pressure membrane fouling using a magnetic ion exchange (MIEX) resin. *Water Res*, 98**,** 225-234.

Karnik, B. S., Davies, S. H., Baumann, M. J. and Masten, S. J. 2005. The effects of combined ozonation and filtration on disinfection by-product formation. *Water Res*, 39**,** 2839-50.

Kawamoto, T. and Makihata, N. 2004. Distribution of Bromine/Chlorine-Containing Disinfection By-Products in Tap Water from Different Water Sources in the Hyogo Prefecture. *J. Health Sci.*, 50**,** 235-247.

Kennedy, M. D., Chun, H. K., Yangal, V. A. Q., Heijman, B. G. J. and 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**,** 73-83.

Kennedy, M. D., Kamanyi, J., Heijman, B. G. J. and Amy, G. 2008. Colloidal organic matter fouling of UF membranes: role of NOM composition & size. *Desalination*, 220**,** 200–213.

King, W. D. and Marrett, L. D. 1996. Case-control study of bladder cancer and chlorination by-products in treated water (Ontario, Canada). *Cancer Cause Control*, 7**,** 596–604.

Kitis, M., Karanfil, T., Kilduff, J. E. and Wigton, A. 2001. The reactivity of natural organic matter to disinfection by-products formation and its relation to specific ultraviolet absorbance. *Wat. Sci. Tech.*, 43**,** 9-16.

Kitis, M., Karanfil, T., Wigton, A. and Kilduff, J. E. 2002. Probing reactivity of dissolved organic matter for disinfection by-product formation using XAD-8 resin adsorption and ultrafiltration fractionation. *Water Res.*, 36**,** 3834–3848.

Korak, J. A., Dotson, A. D., Summers, R. S. and Rosario-Ortiz, F. L. 2013. Critical analysis of commonly used fluorescence metrics to characterize dissolved organic matter. *Water Res*, 49C**,** 327-338.

Korn, C., Andrews, R. C. and Escobar, M. D. 2002. Development of chlorine dioxide-related by-product models for drinking water treatment. *Water Res.*, 36**,** 330–342.

Korshin, G. V., Li, C.-W. and Benjamin, M. M. 1997. Monitoring the Properties of Natural Organic Matter Through UV Spectroscopy: A Consistent Theory. *Water Research*, 31**,** 1787-1795.

Kristiana, I., Joll, C. and Heitz, A. 2011. Powdered activated carbon coupled with enhanced coagulation for natural organic matter removal and disinfection by-product control: application in a Western Australian water treatment plant. *Chemosphere*, 83**,** 661-7.

Kulkarni, P. and Chellam, S. 2010. Disinfection by-product formation following chlorination of drinking water: artificial neural network models and changes in speciation with treatment. *Sci. Total Environ.* , 408**,** 4202-10.

Kumar, S., Forand, S., Babcock, G., Richter, W., Hart, T. and Hwang, S. A. 2014. Total trihalomethanes in public drinking water supply and birth outcomes: a cross-sectional study. *Maternal and Child Health Journal*, 18**,** 996-1006.

Lavonen, E. E., Kothawala, D. N., Tranvik, L. J., Gonsior, M., Schmitt-Kopplin, P. and Kohler, S. J. 2015. Tracking changes in the optical properties and molecular composition of dissolved organic matter during drinking water production. *Water Res*, 85**,** 286-94.

Lawaetz, A. J. and Stedmon, C. A. 2009. Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water. *Appl Spectrosc*, 63**,** 936-940.

Lee, S. and Elimelech, M. 2006. Relating Organic Fouling of Reverse Osmosis Membranes to Intermolecular Adhesion Forces. *Environ Sci Technol*, 40**,** 980-987.

Li, A., Zhao, X., Mao, R., Liu, H. and Qu, J. 2014a. Characterization of dissolved organic matter from surface waters with low to high dissolved organic carbon and the related disinfection byproduct formation potential. *J. Hazard. Mater.*, 271**,** 228-35.

Li, K., Huang, T., Qu, F., Du, X., Ding, A., Li, G. and Liang, H. 2016. Performance of adsorption pretreatment in mitigating humic acid fouling of ultrafiltration membrane under environmentally relevant ionic conditions. *Desalination*, 377**,** 91-98.

Li, W. T., Chen, S. Y., Xu, Z. X., Li, Y., Shuang, C. D. and Li, A. M. 2014b. Characterization of dissolved organic matter in municipal wastewater using fluorescence PARAFAC analysis and chromatography multi-excitation/emission scan: a comparative study. *Environ Sci Technol*, 48**,** 2603-9.

Li, W. T., Xu, Z. X., Li, A. M., Wu, W., Zhou, Q. and Wang, J. N. 2013. HPLC/HPSEC-FLD with multi-excitation/emission scan for EEM interpretation and dissolved organic matter analysis. *Water Research*, 47**,** 1246-56.

Li, Z., Clark, R. M., Buchberger, S. G. and Jeffrey Yang, Y. 2014c. Evaluation of Climate Change Impact on Drinking Water Treatment Plant Operation. *J. Environ. Eng.*, 140**,** A4014005.

Liang, L. and Singer, P. 2003. Factors Influencing the Formation and Relative Distribution of Haloacetic Acids and Trihalomethanes in Drinking Water. *Environ. Sci. Technol.*, 37**,** 2920-2928.

Liu, X., Zhang, Y., Shi, K., Zhu, G., Xu, H. and Zhu, M. 2014. Absorption and fluorescence properties of chromophoric dissolved organic matter: implications for the monitoring of water quality in a large subtropical reservoir. *Environ Sci Pollut Res Int*.

Lochmuller, C. H. and Saavedra, S. S. 1986. Conformational Changes in a Soil Fulvic Acid Measured by Time-Dependent Fluorescence Depolarization. *Anal. Chem.*, 58**,** 1978-1981.

Lorain, O., Hersant, B., Persin, F., Grasmick, A., Brunard, N. and Espenan, J. M. 2007. Ultrafiltration membrane pre-treatment benefits for reverse osmosis process in seawater desalting. Quantification in terms of capital investment cost and operating cost reduction. *Desalination*, 203**,** 277-285.

Louie, S. M., Tilton, R. D. and Lowry, G. V. 2013. Effects of molecular weight distribution and chemical properties of natural organic matter on gold nanoparticle aggregation. *Environ Sci Technol*, 47**,** 4245-54.

Lu, J., Zhang, T., Ma, J. and Chen, Z. 2009. Evaluation of disinfection by-products formation during chlorination and chloramination of dissolved natural organic matter fractions isolated from a filtered river water. *J Hazard Mater*, 162**,** 140-5.

Ma, D., Peng, B., Zhang, Y., Gao, B., Wang, Y., Yue, Q. and Li, Q. 2014. Influences of dissolved organic matter characteristics on trihalomethanes formation during chlorine disinfection of membrane bioreactor effluents. *Bioresour Technol*, 165**,** 81-7.

Mao, Y., Wang, X., Yang, H., Wang, H. and Xie, Y. F. 2014. Effects of ozonation on disinfection byproduct formation and speciation during subsequent chlorination. *Chemosphere*, 117**,** 515-20.

Martínez, C., Gómez, V., Pocurull, E. and Borrull, F. 2015. Characterization of organic fouling in reverse osmosis membranes by headspace solid phase microextraction and gas chromatography-mass spectrometry. *Water Sci Technol*, 71**,** 13-21.

Matilainen, A., Gjessing, E. T., Lahtinen, T., Hed, L., Bhatnagar, A. and Sillanpaa, M. 2011. An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. *Chemosphere*, 83**,** 1431-42.

Mayer, B. K., Daugherty, E. and Abbaszadegan, M. 2015. Evaluation of the relationship between bulk organic precursors and disinfection byproduct formation for advanced oxidation processes. *Chemosphere*, 121**,** 39-46.

McKnight, D. M., Boyer, E. W., Westerhoff, P. K., Doran, P. T., Kulbe, T. and Andersen, D. T. 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography*, 46**,** 38-48.

Miller, J. W. and Uden, P. C. 1983. Characterization of nonvolatile aqueous chlorination products of humic substances. *Environ. Sci. Technol.*, 17**,** 150-157.

Miyoshi, T., Nagai, Y., Aizawa, T., Kimura, K. and Watanabe, Y. 2015. Proteins causing membrane fouling in membrane bioreactors. *Water Sci Technol*, 72**,** 844-849.

Montesinos, I. and Gallego, M. 2013. Speciation of common volatile halogenated disinfection by-products in tap water under different oxidising agents. *J. Chromatogr.*, 1310**,** 113-20.

Moslemi, M. D., Simon H.; Masten, Susan J. 2012. Empirical modeling of bromate formation during drinking water treatment using hybrid ozonation membrane filtration. *Desalination*.

Muellner, M. G., Wagner, E. D., McCalla, K., Richardson, S. D., Woo, Y.-T. and Plewa, M. J. 2007. Haloacetonitriles vs. Regulated Haloacetic Acids: Are Nitrogen-Containing DBPs More Toxic? *Environ Sci Technol*, 41**,** 645-651.

Murphy, H. M., Bhatti, M. A., Harvey, R. and McBean, E. A. 2016. Using decision trees to predict drinking water advisories in small water systems. *J. Am. Water Works Assoc.*, 108.

Murphy, K. R., Stedmon, C. A., Graeber, D. and Bro, R. 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Analytical Methods*, 5**,** 6557.

Myat, D. T., Stewart, M. B., Mergen, M., Zhao, O., Orbell, J. D. and Gray, S. 2014. Experimental and computational investigations of the interactions between model organic compounds and subsequent membrane fouling. *Water Res*, 48**,** 108-18.

Najm, I. N., Patania, N. L., Jacangelo, J. G. and Krasner, S. W. 1994. Evaluating surrogates for disinfection by-products. *J. Am. Water Works Assoc.*, 86.

Nam, J.-W., Hing, S.-H., Park, J.-Y., Park, H.-S., Kim, H.-S. and Jang, A. 2013. Evaluation of chemical cleaning efficiency of organic-fouled SWRO membrane by analyzing filtration resistance. *Desalination and Water Treatment*, 51**,** 6172-6178.

Navalon, S., Alvaro, M. and Garcia, H. 2008. Carbohydrates as trihalomethanes precursors. Influence of pH and the presence of Cl(-) and Br(-) on trihalomethane formation potential. *Water Res.*, 42**,** 3990-4000.

Ng, H. K. T. and Balakrishnan, N. 2004. Wilcoxon-Type Rank-Sum Precedence Tests. *Aust. N. Z. J. Stat.* , 46**,** 631–648.

Nissinen, T. K., Miettinen, I. T., Martikainen, P. J. and Vartiainen, T. 2001. Molecular Size Distribution of Natural Organic Matter in Raw and Drinking Waters. *Chemosphere*, 45**,** 865-873.

Nokes, C., Fenton, E., Randall, J. 1999. Modeling the formation of brominated trihalomethanes in chlorinated drinking waters. *Water Res.*, 33**,** 3557-3568.

Obolensky, A. and Singer, P. C. 2005. Halogen Substitution Patterns among Disinfection Byproducts in the Information Collection Rule Database. *Environ. Sci. Technol.*, 39**,** 2719-2730.

Obolensky, A. and Singer, P. C. 2008. Development and Interpretation of Disinfection Byproduct Formation Models Using the Information Collection Rule Database. *Environ. Sci. Technol.*, 42**,** 5654–5660.

Oliver, B. G. and Lawrence, J. 1979. Haloforms in Drinking Water: A Study of Precursors and Precursor Removal. *J. Am Water Works Assoc.*, 71**,** 161-163.

Owen, D. M., Amy, G. L., Chowdhury, Z. K., Paode, R., McCoy, G. and Viscosil, K. 1995. NOM Characterization and Treatability. *J. Am Water Works Assoc.*, 87**,** 46-63.

Peiris, R. H., Budman, H., Moresoli, C. and Legge, R. L. 2010a. Understanding fouling behaviour of ultrafiltration membrane processes and natural water using principal component analysis of fluorescence excitation-emission matrices. *Journal of Membrane Science*, 357**,** 62-72.

Peiris, R. H., Halle, C., Budman, H., Moresoli, C., Peldszus, S., Huck, P. M. and Legge, R. L. 2010b. Identifying fouling events in a membrane-based drinking water treatment process

using principal component analysis of fluorescence excitation-emission matrices. *Water Res*, 44**,** 185-94.
Peiris, R. H., Jaklewicz, M., Budman, H., Legge, R. L. and Moresoli, C. 2013. Assessing the role of feed water constituents in irreversible membrane fouling of pilot-scale ultrafiltration drinking water treatment systems. *Water Res*, 47**,** 3364-74.
Peleato, N. M. and Andrews, R. C. 2015. Comparison of three-dimensional fluorescence analysis methods for predicting formation of trihalomethanes and haloacetic acids. *Journal of Environmental Sciences*, 27**,** 159-167.
Peleato, N. M., McKie, M., Taylor-Edmonds, L., Andrews, S. A., Legge, R. L. and Andrews, R. C. 2016. Fluorescence spectroscopy for monitoring reduction of natural organic matter and halogenated furanone precursors by biofiltration. *Chemosphere*, 153**,** 155-61.
Pickhardt, W. P., Oemler, A. N. and John Mitchell, J. 1955. Determination of Total Carbon in Organic Materials by a Wet-Dry Combustion Method. *Analytical Chemistry*, 27 1784-1788.
Pifer, A. D. and Fairey, J. L. 2012. Improving on SUVA 254 using fluorescence-PARAFAC analysis and asymmetric flow-field flow fractionation for assessing disinfection byproduct formation and control. *Water Res.*, 46**,** 2927-36.
Pifer, A. D. and Fairey, J. L. 2014. Suitability of Organic Matter Surrogates to Predict Trihalomethane Formation in Drinking Water Sources. *Environ. Eng. Sci.*, 31**,** 117-126.
Pifer, A. D., Miskin, D. R., Cousins, S. L. and Fairey, J. L. 2011. Coupling asymmetric flow-field flow fractionation and fluorescence parallel factor analysis reveals stratification of dissolved organic matter in a drinking water reservoir. *J. Chromatogr.*, 1218**,** 4167-78.
Pisarenko, A. N., Stanford, B. D., Snyder, S. A., Rivera, S. B. and Boal, A. K. 2013. Investigation of the use of Chlorine Based Advanced Oxidation in Surface Water: Oxidation of Natural OrganicMatter and Formation of Disinfection Byproducts. *Journal of Advanced Oxidation Technology*, 16**,** 137-150.
Plewa, M. J., Kargalioglu, Y., Vankerk, D., Minear, R. A. and Wagner, E. D. 2002. Mammalian cell cytotoxicity and genotoxicity analysis of drinking water disinfection by-products. *Environ. Mol. Mutagen.*, 40**,** 134-42.
Plewa, M. J., Wagner, E. D., Richardson, S. D., Thruston, A. D. J., Woo, Y.-T. and McKague, B. A. 2004. Chemical and Biological Characterization of Newly Discovered Iodoacid Drinking Water Disinfection Byproducts. *Environmental Science and Technology*, 38**,** 4713-4722.
Pramanik, B. K., Roddick, F. A. and Fan, L. 2016. Long-term operation of biological activated carbon pre-treatment for microfiltration of secondary effluent: Correlation between the organic foulants and fouling potential. *Water Res*, 90**,** 405-14.
Rathburn, R. E. 1996a. Bromine Incorporation Factors for Triahlomethane Formation for the Mississippi, Missouri, and Ohio Rivers. *Sci Total Environ*, 192**,** 111-118.
Rathburn, R. E. 1996b. Speciation of trihalomethane mixtures for the Mississippi, Missouri, and Ohio Rivers. *Sci Total Environ*, 180**,** 125-135.
Rausa, R., Mazzolari, E. and Calemma, V. 1991. Determination of molecular size distributions of humic acids by high-performance size-exclusion chromatography. *Journal of Chromatography*, 541**,** 419-429.
RCoreTeam 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*, URL https://www.R-project.org/.

Reckhow, D. A., Singer, P. C. and Malcolm, R. L. 1990. Chlorination of Humic Materials: Byproduct Formation and Chemical Interpretations. *Environ. Sci. Technol.*, 24**,** 1655-1664.

Regli, S., Chen, J., Messner, M., Elovitz, M. S., Letkiewicz, F., Pegram, R., Pepping, T. J., Richardson, S. and Wright, J. M. 2015. Estimating Potential Increased Bladder Cancer Risk Due to Increased Bromide Concentrations in Sources of Disinfected Drinking Waters. *Environ Sci Technol*.

Richardson, S., Thruston, A., Caughran, T., Chen, P., Collette, T., Schenck, K., Lykins, B., Rav-Acha, C. and Glezer, V. 2000. Identification of new drinking water disinfection by-products from ozone, chlorine dioxide, chloramine, and chlorine. *Water, air, and soil pollution*, 123**,** 95 - 102.

Richardson, S. D., Plewa, M. J., Wagner, E. D., Schoeny, R. and Demarini, D. M. 2007. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. *Mutat Res*, 636**,** 178-242.

Richardson, S. D., Thruston, A. D., Collette, T. W., Patterson, K. S., Lykins, B. W., Majetich, G. and Zhang, Y. 1994. Multispectral Identification of Chlorine Dioxide Disinfection Byproducts in Drinking Water. *Environ. Sci. Technol.* , 28**,** 592-599.

Richardson, S. D., Thruston, J., Alfred D. , Rav-Acha, C., Groisman, L., Popilevsky, I., Juraev, O., Glezer, V., McKague, A. B., Plewa, M. J. and Wagner, E. D. 2003. Tribromopyrrole, Brominated Acids, and Other Disinfection Byproducts Produced by Disinfection of Drinking Water Rich in Bromide. *Environ. Sci. Technol.*, 37**,** 3782-3793.

Roberson, J. A., Cromwell III, J. E., Krasner, S. W. and McGuire, M. J. 1995. The D/DBP Rule: where did the numbers come from? *J. Am. Water Works Assoc.*, 87**,** 46-57.

Roccaro, P., Chang, H. S., Vagliasindi, F. G. and Korshin, G. V. 2008. Differential absorbance study of effects of temperature on chlorine consumption and formation of disinfection by-products in chlorinated water. *Water Res*, 42**,** 1879-88.

Roccaro, P. and Vagliasindi, F. G. A. 2010. Monitoring emerging chlorination by-products in drinking water using UV-absorbance and fluorescence indexes. *Desalination and Water Treatment*, 23**,** 118-122.

Roccaro, P., Vagliasindi, F. G. A. and Korshin, G. V. 2009. Changes in NOM Fluorescence Caused by Chlorination and their Associations with Disinfection By-Products Formation. *Environ Sci Technol*, 43**,** 724-729.

Roccaro, P., Yan, M. and Korshin, G. V. 2015. Use of log-transformed absorbance spectra for online monitoring of the reactivity of natural organic matter. *Water Res*, 84**,** 136-43.

Rodriguez, M., Serodes, J., Levallois, P. and Proulx, F. 2007. Chlorinated disinfection by-products in drinking water according to source, treatment, season, and distribution location. *Journal of Environmental Engineering and Science*, 6**,** 355-365.

Rodriguez, M. J., Serodes, J. B. and Levallois, P. 2004. Behavior of trihalomethanes and haloacetic acids in a drinking water distribution system. *Water Res*, 38**,** 4367-82.

Rook, J. J. 1976. Haloforms in Drinking Water. *J. Am Water Works Assoc.*, 68**,** 168-172.

Rukapan, W., Khananthai, B., Srisukphun, T., Chiemchaisri, W. and Chiemchaisri, C. 2015. Comparison of reverse osmosis membrane fouling characteristics in full-scale leachate treatment systems with chemical coagulation and microfiltration pre-treatments. *Water Sci Technol*, 71**,** 580-587.

Sadiq, R. and Rodriguez, M. J. 2004. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. *Sci Total Environ*, 321**,** 21-46.

Sakai, H., Tokuhara, S., Murakami, M., Kosaka, K., Oguma, K. and Takizawa, S. 2015. Comparison of chlorination and chloramination in carbonaceous and nitrogenous disinfection byproduct formation potentials with prolonged contact time. *Water Res*, 88**,** 661-670.

Sanchez, N. P., Skeriotis, A. T. and Miller, C. M. 2013. Assessment of dissolved organic matter fluorescence PARAFAC components before and after coagulation-filtration in a full scale water treatment plant. *Water Research*, 47**,** 1679-90.

Saravia, F., Zwiener, C. and Frimmel, F. H. 2006. Interactions between membrane surface, dissolved organic substances and ions in submerged membrane filtration. *Desalination*, 192**,** 280-287.

Schäfer, A. I., Fane, A. G. and Waite, T. D. 2000. Fouling effects on rejection in the membrane filtration of natural waters. *Desalination*, 131**,** 215-224.

Shan, L., Fan, H., Guo, H., Ji, S. and Zhang, G. 2016. Natural organic matter fouling behaviors on superwetting nanofiltration membranes. *Water Res*, 93**,** 121-32.

Shao, S., Liang, H., Qu, F., Yu, H., Li, K. and Li, G. 2014. Fluorescent natural organic matter fractions responsible for ultrafiltration membrane fouling: Identification by adsorption pretreatment coupled with parallel factor analysis of excitation–emission matrices. *Journal of Membrane Science*, 464**,** 33-42.

Shutova, Y., Baker, A., Bridgeman, J. and Henderson, R. K. 2014. Spectroscopic characterisation of dissolved organic matter changes in drinking water treatment: From PARAFAC analysis to online monitoring wavelengths. *Water Res*, 54**,** 159-69.

Siedel, C. J., McGuire, M. J., Summers, R. S. and Via, S. 2005. Have utilities switched to chloramines? *J. Am Water Works Assoc.*, 97**,** 87-97.

Sierra, M. M. D. S., Donard, O. F. X., Lamotte, M., Belin, C. and Ewald, M. 1994. Fluorescence spectroscopy of coastal and marine waters. *Mar. Chem.* , 47**,** 127-144.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics*, 21**,** 3940-1.

Singer, P., Weinberg, H., Brophy, K., Liang, L., Roberts, M., Grisstede, I., Krasner, S., Baribeau, H., Arora, H. and Najm, I. 2002. Relative dominance of haloacetic acids and trihalomethanes in treated drinking water. *Denver, CO: American Water Works Association Research Foundation.*

Sohn, J., Amy, G., Cho, J., Lee, Y. and Yoon, Y. 2004. Disinfectant decay and disinfection by-products formation model development: chlorination and ozonation by-products. *Water Res.*, 38**,** 2461-78.

Sohn, J., Amy, G. and Yoon, Y. 2006. Bromide Ion Incorporation Into Brominated Disinfection By-Products. *Water Air Soil Poll*, 174**,** 265-277.

Song, L. and Elimelech, M. 1995. Theory of Concentration Polarization in Crossflow Filtration. *J. Chem. Soc., Faraday Trans.*, 91**,** 3389-3398.

States, S., Cyprych, G., Stoner, M., Wydra, F., Kuchta, J., Monnell, J. and Casson, L. 2013. Brominated THMs in Drinking Water: A Possible Link to Marcellus Shale and Other Wastewaters. *J. Am. Water Works Assoc.*, 105**,** E432-E448.

Stedmon, C. A. and Bro, R. 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnol. Oceanogr.: Methods*, 6**,** 572–579.

Stedmon, C. A. and Markager, S. 2005. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol. Oceanogr.*, 50**,** 686-697.

Stedmon, C. A., Markager, S. and Bro, R. 2003a. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* , 82**,** 239-254.

Stedmon, C. A., Markager, S. and Bro, R. 2003b. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine chemistry*, 82**,** 239-254.

Stedmon, C. A., Seredynska-Sobecka, B., Boe-Hansen, R., Le Tallec, N., Waul, C. K. and Arvin, E. 2011. A potential approach for monitoring drinking water quality from groundwater systems using organic matter fluorescence as an early warning for contamination events. *Water Res*, 45**,** 6030-8.

Tang, C. Y., Kwon, Y.-N. and Leckie, J. O. 2007. Characterization of Humic Acid Fouled Reverse Osmosis and Nanofiltration Membranes by Transmission Electron Microscopy and Streaming Potential Measurements. *Environ Sci Technol*, 41**,** 942-949.

Tang, F., Hu, H. Y., Sun, L. J., Sun, Y. X., Shi, N. and Crittenden, J. C. 2016. Fouling characteristics of reverse osmosis membranes at different positions of a full-scale plant for municipal wastewater reclamation. *Water Res*, 90**,** 329-36.

Tian, C., Liu, R., Guo, T., Liu, H., Luo, Q. and Qu, J. 2013. Chlorination and chloramination of high-bromide natural water: DBPs species transformation. *Sep. Purif. Technol.*, 102**,** 86-93.

Tiraferri, A. and Elimelech, M. 2012. Direct quantification of negatively charged functional groups on membrane surfaces. *Journal of Membrane Science*, 389**,** 499-508.

Traina, S., Novak, J. and Smeck, N. E. 1990. An Ultraviolet Absorbance Method of Estimating the Percent Aromatic Carbon Content of Humic Acids. *Journal of Environmental Quality*, 19**,** 151-153.

Tran, N. H., Ngo, H. H., Urase, T. and Gin, K. Y. 2015. A critical review on characterization strategies of organic matter for wastewater and water treatment processes. *Bioresour Technol*, 193**,** 523-33.

Trueman, B. F., MacIsaac, S. A., Stoddart, A. K. and Gagnon, G. A. 2016. Prediction of disinfection by-product formation in drinking water via fluorescence spectroscopy. *Environ. Sci.: Water Res. Technol.*, 2**,** 383-389.

Uyak, V., Yavuz, S., Toroz, I., Ozaydin, S. and Genceli, E. A. 2007. Disinfection by-products precursors removal by enhanced coagulation and PAC adsorption. *Desalination*, 216**,** 334-344.

Vial, D. and Doussau, G. 2002. The use of microfiltration membranes for seawater pre-treatment prior to reverse osmosis membranes. *Desalination* 153**,** 141-147.

Villanueva, C. M., Cantor, K. P., Cordier, S., Jaakkola, J. J. K., King, W. D., Lynch, C. F., Porru, S. and Kogevinas, M. 2004. Disinfection Byproducts and Bladder Cancer. *Epidemiology*, 15**,** 357-367.

Vuorio, E., Vahala, R., Rintala, J. and Laukkanen, R. 1998. The Evaluation of Drinking Water Treatment Performed with HPSEC. *Environment International*, 24**,** 617-623.

Wang, D. S., Zhao, Y. M., Yan, M. Q. and Chow, C. W. K. 2013. Removal of DBP precursors in micro-polluted source waters: A comparative study on the enhanced coagulation behavior. *Separation and Purification Technology*, 118**,** 271-278.

Wang, Y., Wilson, J. M. and VanBriesen, J. M. 2015. The effect of sampling strategies on assessment of water quality criteria attainment. *J. Environ. Manage*, 154**,** 33-39.

Watson, K., Farre, M. J., Birt, J., McGree, J. and Knight, N. 2015. Predictive models for water sources with high susceptibility for bromine-containing disinfection by-product formation: implications for water treatment. *Environmental Science Pollution Research International*, 22**,** 1963-78.

Weaver, J. W., Xu, J. and Mravik, S. C. 2015. Scenario Analysis of the Impact on Drinking Water Intakes from Bromide in the Discharge of Treated Oil and Gas Wastewater. *J. Environ. Eng.*, **,** DOI: 10.1061/(ASCE)EE.1943-7870.0000968.

Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R. and Mopper, K. 2003. Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon. *Environmental Science and Technology*, 37**,** 4702-4708.

Weiss, J. W., Schindler, S. C., Freud, S., Herzner, J. A., Hoek, K. F., Wright, B. A., Reckhow, D. A. and Becker, W. C. 2013. Minimizing raw water NOM concentration through optimized source water selection. *J. Am. Water Works Assoc.*, 105**,** E596-E608.

Westerhoff, P., Debroux, J., Amy, G. L., Gatel, D., Mary, V. and Cavard, J. 2000. Applying DBP models to full-scale plants. *American Water Works Association*, 92**,** 89-102.

Wilson, J. M. 2013. Challenges for Drinking Water Plants from Energy Extraction Activities. *PhD Dissertation*, Dept. of Civil & Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA.

Wilson, J. M. and Van Briesen, J. M. 2013. Source water changes and energy extraction activities in the Monongahela River, 2009-2012. *Environ. Sci. Technol.*, 47**,** 12575-82.

Wong, S., Hanna, J. V., King, S., Carroll, T. J., Eldridge, R. J., Dixon, D. R., Bolto, B. A., Hesse, S., Abbt-Braun, G. and Frimmel, F. H. 2002. Fractionation of Natural Organic Matter in Drinking Water and Characterization by 13C Cross-Polarization Magic-Angle Spinning NMR Spectroscopy and Size Exclusion Chromatography. *Environ Sci Technol*, 36**,** 3497-3503.

Yamamura, H., Okimoto, K., Kimura, K. and Watanabe, Y. 2014. Hydrophilic fraction of natural organic matter causing irreversible fouling of microfiltration and ultrafiltration membranes. *Water Res.*, 54**,** 123-36.

Yang, L., Hur, J., Lee, S., Chang, S. W. and Shin, H. S. 2015a. Dynamics of dissolved organic matter during four storm events in two forest streams: source, export, and implications for harmful disinfection byproduct formation. *Environmental Science and Pollution Research International*, 22**,** 9173-83.

Yang, L., Kim, D., Uzun, H., Karanfil, T. and Hur, J. 2015b. Assessing trihalomethanes (THMs) and N-nitrosodimethylamine (NDMA) formation potentials in drinking water treatment plants using fluorescence spectroscopy and parallel factor analysis. *Chemosphere*, 121**,** 84-91.

Yu, H., Qu, F., Liang, H., Han, Z.-s., Ma, J., Shao, S., Chang, H. and Li, G. 2014. Understanding ultrafiltration membrane fouling by extracellular organic matter of Microcystis aeruginosa using fluorescence excitation–emission matrix coupled with parallel factor analysis. *Desalination*, 337**,** 67-75.

Zhang, X., Fan, L. and Roddick, F. A. 2014. Feedwater coagulation to mitigate the fouling of a ceramic MF membrane caused by soluble algal organic matter. *Separation and Purification Technology*, 133**,** 221-226.

Zhang, Y., Zhao, X., Zhang, X. and Peng, S. 2015. A review of different drinking water treatments for natural organic matter removal. *Wat Sci Tech*, 15**,** 442-455.

Zhao, Y., Song, L. and Ong, S. L. 2010. Fouling behavior and foulant characteristics of reverse osmosis membranes for treated secondary effluent reclamation. *Journal of Membrane Science,* 349**,** 65-74.

Zhu, X. and Elimelech, M. 1997. Colloidal Fouling of Reverse Osmosis Membranes: Measurements and Fouling Mechanisms. *Environmental Science and Technology,* 31**,** 3654-3662.

Zodrow, K., Brunet, L., Mahendra, S., Li, D., Zhang, A., Li, Q. and Alvarez, P. J. 2009. Polysulfone ultrafiltration membranes impregnated with silver nanoparticles show improved biofouling resistance and virus removal. *Water Res,* 43**,** 715-23.