INTERPRETING SOURCES AND ENDOCRINE ACTIVE COMPONENTS OF TRACE ORGANIC CONTAMINANT MIXTURES IN MINNESOTA LAKES

by

Meaghan E. Guyader

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Civil and Environmental Engineering).

Golden, Colorado

Date ______________________________

Signed: ______________________________
Meaghan E. Guyader

Signed: ______________________________
Dr. Christopher P. Higgins
Thesis Advisor

Golden, Colorado

Date ______________________________

Signed: ______________________________
Dr. Terri S. Hogue
Professor and Department Head
Department of Civil and Environmental Engineering

ABSTRACT

On-site wastewater treatment systems (OWTSs) are a suspected source of widespread trace organic contaminant (TOrC) occurrence in Minnesota lakes. TOrCs are a diverse set of synthetic and natural chemicals regularly used as cleaning agents, personal care products, medicinal substances, herbicides and pesticides, and foods or flavorings. Wastewater streams are known to concentrate TOrC discharges to the environment, particularly accumulating these chemicals at outfalls from centralized wastewater treatment plants. Fish inhabiting these effluent dominated environments are also known to display intersex qualities. Concurrent evidence of this phenomenon, known as endocrine disruption, in Minnesota lake fish drives hypotheses that OWTSs, the primary form of wastewater treatment in shoreline residences, may contribute to TOrC occurrence and the endocrine activity in these water bodies. The causative agents specific to fish in this region remain poorly understood.

The objective of this dissertation was to investigate OWTSs as sources of TOrCs in Minnesota lakes, and TOrCs as potential causative agents for endocrine disruption in resident fish. Three research efforts were executed to investigate these topics: examining chemical and biological signatures of OWTS proximity in Minnesota lakes (Chapter 2), prioritizing potential causative agents of endocrine disruption using liquid chromatography tandem mass spectrometry (LC-HRMS) (Chapter 3), and conducting a suspect search of OWTS-associated LC-HRMS features at an adjacent Minnesota lake (Chapter 4). In Chapter 2, traditional targeted aqueous analyses indicated higher concentrations of TOrCs at locations more proximal to residences with OWTSs. Residential proximity also corresponded to feminization of male sunfish. The particular contaminants detected at these locations are considered weak indicators of wastewater presence in environmental compartments and inactive with current metrics for endocrine activity. Unexpected features unique to sites with pronounced endocrine disruption were identified in Chapter 3; however, these components were still considered endocrine inactive by current toxicology metrics. This suggests that temporal resolution in sampling was too low for this chronic-toxicity endpoint, or current regulatory efforts underestimate the effects of these contaminants as environmental mixtures. Finally, Chapter 5 indicated that LC-HRMS analysis and passive sampling allowed for identification of a broader suite of OWTS-associated compounds in adjacent lake water, but that higher spatial resolution was required to refine lake-specific OWTS-compound interest lists. The

results from this dissertation encourage further investigation of residential inputs of TOrCs to Minnesota lakes, particularly prioritizing locations with endocrine disruption, so that local regulatory agencies may effectively manage these highly valued state resources.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have infinite gratitude for those that supported me during this fantastic time of growth. Firstly, thank you to my family. To my mother, Jean Guyader, thank you for continually offering me your light no matter my circumstances. You are the reason I value tenacity, compassion, and adaptability in my life. Thank you to my father, Michael Guyader, for teaching me to value my independence and assert myself. To my brother, Matthew Guyader, I will always appreciate your sense of humor and thank you for helping me avoid dwelling in my struggles. And to my extended family, thank you for always expressing your love and sharing your laughter.

Secondly, many thanks to my ever-expanding community of friends. Thank you to all the Jersey folk, Nora Polasky, Juliet Vinegra, Sandra Pavleska, Matthew Golish, Alan Chu, Stephen Hilger, and Lauren Decina, who offered so much encouragement when I first cooked up this fixation of moving across the country and pursuing this degree. You all keep me grounded in times of panic and continue to elevate the ways in which I celebrate life. Thank you to all of my Colorado roommates who have put up with my growing pains and helped me establish a sense of home in a new location, particularly Charley Parker, Mark Webb, and Shay Skinner. Thank you to Maxine Morris, Aniela Burant, Karl Oetjen, Daniel van Hoomisen, Carrie McDonough, Kate Spangler, Michael Greco, and all Chicken Hut affiliates for offering your friendship in a way that both comforted and challenged my sense of self during this peculiar age. Thank you to Gabriela Ferrat and Kate Kelsey who counseled me while I navigated a new path toward mental and physical health and harmony. Thank you to the beautiful members of the Ladies of the Lake housing collective that helped me solidify my decision to continue my growth in the Denver metro area, especially my good friend Michaela Miller. And thank you to the broader Golden and Denver metro area community, for allowing me to experience a way to organize people around ideals, dreams, and things many of us transplants were previously taught were impossibilities.

Finally, thank you to my colleagues and mentors who engaged my technical curiosities and advanced my professional prospects. Thank you to my very first mentor in graduate school, Bridget Ulrich, my advisor, Chris Higgins, the Higgins research group, my committee, and the staff from USGS National Water Quality Laboratory, for helping me refine my understanding of water quality analysis and data interpretation. And many thanks to members of the HASS and HE departments who nourished my interest in confronting the boundaries of environmental research.

This work is dedicated to Hurley T.

CHAPTER 1
INTRODUCTION

Monitoring efforts conducted over the past decade indicate widespread trace organic contaminant (TOrC) occurrence in lakes throughout Minnesota (Baker et al., 2014; Ferrey et al., 2015; Writer et al., 2010). TOrCs are a diverse class of environmental pollutants that encompass a wide variety of chemical structures, including pharmaceuticals, herbicides, pesticides, hormones, steroids, personal care products, cleaning agents, and food preservatives (Lapworth et al., 2012). With the regular production, usage, and disposal of these chemicals by humans, TOrCs can enter the environment from a number of point and non-point sources. The Minnesota Pollution Control Agency (MPCA) started monitoring Minnesota surface waters for TOrCs in 2010 in compliance with the EPA's National Aquatic Resource Survey (Engelking and Kovacevic, 2016). A randomized study of fifty lakes reported TOrC detections across the state's surface waters at ng/L to µg/L concentrations (Ferrey et al., 2015). Detections in both populated and remote regions, particularly regions distant from point sources such as wastewater treatment effluent outfalls, drives a general consensus that these contaminants originate from non-point sources (Ferrey et al., 2015; Writer et al., 2010). These water bodies are highly valued and integrated into many aspects of Minnesota residents' lifestyles. As a result, there are a number of potential TOrC sources to consider from recreational activities such as fishing and boating to anthropogenic activities at temporary and permanent shoreline residences (Baker et al., 2014). In addition, the hydrology of lakes allow for diffuse TOrC transport through multiple pathways, including atmospheric deposition, surface water runoff, and subsurface transport in shallow groundwater (Ferrey et al., 2015).

Concurrent evidence of endocrine disruption in Minnesota lake fish raises concern for the presence of these contaminants (Baker et al., 2014; Writer et al., 2010). Sampling efforts indicate elevated blood vitellogenin (VTG) concentrations in male fish inhabiting the lakes. This egg yolk protein, typically produced by females, is considered a reliable bioassay for exposure to estrogen agonist mixtures (Hansen et al., 1998; Sumpter and Jobling, 1995). Identifying the particular endocrine active agents in these mixtures is inherently difficult because endocrine disruption is associated with chronic toxicity (Han et al., 2010). Even though environmental TOrC concentrations are not likely to exceed laboratory determined acute toxicity limits, such as a

TOrC's lethal concentration to kill 50% of an exposed population (LC50), prolonged chemical exposure or exposure during critical times of development lead to observed effects on biota (Farré et al., 2008). Spatial and temporal heterogeneity in TOrC occurrence over a fish's lifetime also makes it difficult to align these biological observations with chemical occurrence. In addition, interactions of co-contaminants in environmental TOrC mixtures may encourage additive or synergistic endocrine disrupting effects (Barber et al., 2012; Burkhardt-holm and Burkhardt-holm, 2015; Lapworth et al., 2012; Schwindt et al., 2014; Writer et al., 2010). Classic studies of fish feminization in the United Kingdom, indicate that hormonal steroids, such as 17β-estradiol and 17α-ethinylestradiol, can act as the dominant causative agents of endocrine disruption (Sumpter and Johnson, 2008a); however, studies from other regions in the world demonstrate alternative agents, such as plant-based TOrCs(Sun et al., 2008; Xue et al., 2015). While the severity of endocrine disrupting effects upon ecological health is still debated (Burkhardt-holm and Burkhardt-holm, 2015), a study in Canada demonstrated that prolonged dosing of a lake in Canada with 17α-ethinylestradiol could collapse an entire fish population (Kidd et al., 2007). The U.S. Environmental Protection Agency (EPA) is prioritizing endocrine disrupting chemicals for risk assessment and potential regulation through the Endocrine Disruptor Screening Program (Borgert et al., 2011; U.S. EPA, 2012). Nevertheless, the causative agents for endocrine disruption in Minnesota lakes remain unknown (Baker et al., 2014).

On-site wastewater treatment systems (OWTSs) are one likely source of TOrCs in this region. Many communities of shoreline residences lack the infrastructure necessary for municipal sewage. Wastewater is then typically treated by some form of on-site treatment, such as septic systems, whose influent is a combination of residential wastewater streams, which settle in a subsurface holding tank whose effluent is dispersed throughout the local subsurface, also called a soil treatment unit (STU) (Standley et al., 2008; West, 2008). While OWTSs are specifically designed to remove traditional contaminants, such as nutrients and suspended solids, they also regularly receive TOrC loadings. Studies of TOrC removal by OWTSs indicate mixed efficacy depending on the contaminants' ability to sorb to regional soils or biodegrade (Carrara et al., 2008b; Godfrey et al., 2007; Schaider et al., 2016; Subedi et al., 2015; Teerlink et al., 2012b). More polar TOrCs are typically more mobile and travel with shallow groundwater in the region, while non-polar TOrCs sorb to solids and particulate matter in the holding tank or STU (Conn et al., 2010; Swartz et al., 2006). In addition, the varied recalcitrance of TOrCs in different redox conditions leads to variable attenuation. Multiple studies have demonstrated the hindered transport of many TOrC parent compounds under aerobic subsurface conditions (Banzhaf et al., 2012; Schaider et al., 2014; Standley et al., 2008; Teerlink et al., 2012b). Yet, several exceptions exist, such as sulfamethoxazole, which is better removed under anaerobic conditions. Furthermore, sucralose's resistance to aerobic degradation and aqueous solubility justify the use of this common-use artificial sweetener as a wastewater tracer (Banzhaf et al., 2012; Schaider et al., 2016; Teerlink et al., 2012b). Improper maintenance, variable shallow groundwater conditions reducing the depth of unsaturation in the STU, and high septic density can also exacerbate threats to the quality of down-gradient environments, including those from TOrC discharges (Carrara et al., 2008a; Stanford and Weinberg, 2010; Swartz et al., 2006; Van Cuyk, 2001; West, 2008). Finally, from a biological perspective, TOrC discharges from centralized wastewater treatment plants are also associated with endocrine disruption (Kolpin et al., 2002; Sumpter and Johnson, 2008b; Vajda et al., 2008). Considering these factors, OWTS contribution to TOrC occurrence and fish feminization in Minnesota lakes requires further investigation.

Recent advancements in environmental analysis encourage robust characterization of TOrCs in different environmental matrices. Traditional liquid chromatography tandem triple quadrupole (LC-MS/MS) allows for separation of chemical analytes within an environmental matrix and identification using distinct pairs of molecular weight precursor ions with their fragmented product ions (Dresen et al., 2010). These reproducible transitions are established using chemical standards also used for regular instrument calibration and quantitation in environmental samples (Schymanski et al., 2015). Isotopically labeled standards are also incorporated into this analysis to account for losses during sample preparation or interferences during instrument analysis that can affect the accuracy of quantitation (Schymanski et al., 2014). However, even with extensive targeted analyte selection, this type of analysis is prone to omitting chemical occurrence data valuable for addressing research objectives, particularly for unexpected chemical constituents (Hernández et al., 2005; Krauss et al., 2010; Schymanski et al., 2014). Liquid chromatography tandem high-resolution mass spectrometry (LC-HRMS) is now a viable alternative to LC-MS/MS that allows for the detection and identification of unique chemical features without the use of chemical standards (Colby et al., 2017; Hug et al., 2014). Instruments such as quadrupole time of flight (QToF/MS) and Orbitrap mass spectrometers are capable of resolving unique, co-eluting chemical components with the same nominal mass. Acquired data can then be extracted for a broad suite of suspected or unknown chemical features. Suspect LC-HRMS screening typically uses existing knowledge of a studied system, such as regional chemical usage, suspected sources, expected fate and transport processes, and even anticipated bioactive agents to define chemicals of interest (Brack et al., 2016; Du et al., 2017; Gago-Ferrero et al., 2018; Leendert et al., 2015; McEachran et al., 2017). Unexpected chemical monitoring, also known as non-target screening, usually follows some type of statistical prioritization or pattern finding workflow to focus identification and trends analysis on important components in an environmental matrix (Blum et al., 2017; Peter et al., 2018; Schollée et al., 2015). Nevertheless, LC-HRMS is still subject to biasing of results due to sample preparation methods, column chromatography, matrix interferences, and a chemical's ability to ionize (Hernández et al., 2005; Krauss et al., 2010; Schymanski et al., 2014). The transition to suspect and non-target analysis could be useful in addressing several of the research questions regarding TOrC occurrence in Minnesota lakes.

In addition to sample analysis, tools for field monitoring have also progressed to accommodate the transient nature of environmental TOrC mixtures. Without an appropriate mode and frequency, traditional grab samples cannot accurately represent environmental loadings to these systems or exposure conditions for chronic toxicity phenomena (Ort et al., 2010; Teerlink et al., 2012b). Time or flow weighted composite samples and passive samplers are now integrated into field studies to understand concentrations of TOrCs relevant on daily or weekly to monthly time scales, respectively (Chambers et al., 2007; Moschet et al., 2015; Ruff et al., 2015). Composite samples capture more temporally representative mass loadings within a sample volume (Chambers et al., 2007). Passive samplers uptake environmental contaminants onto a sorbent material of interest during their deployment (Liscio et al., 2009; McDonough et al., 2014; Togola and Budzinski, 2007; Vetter et al., 2018). These devices allow for longer accumulation time of TOrCs and promote ease of sampling by eliminating the need for many grab sampling materials. Contemporary studies of these tools are currently concerned with methods to standardize their usage in field monitoring. Depending on research objectives, all sample types are relevant for accomplishing environmental research objectives.

The objective of this dissertation was to investigate OWTSs as sources of TOrCs in Minnesota lakes as well as an analysis of TOrCs as potential causative agents for endocrine disruption in resident fish. This was accomplished through three studies: examining chemical and biological signatures of OWTS proximity in five Minnesota lakes (Chapter 2, prioritizing potential causative agents of endocrine disruption using LC-HRMS (Chapter 3), and conducting a suspect search of OWTS-associated LC-HRMS features at an adjacent Minnesota Lake (Chapter 4).

1.1 Objectives and Hypothesis

This section details the research objectives of the three studies, which make up this dissertation and the hypotheses tested during their execution. The first and second objectives are outcomes of a field sampling effort conducted in the Summer of 2016, while the third objective was supported by a field sampling effort from September of 2017. The hypotheses of the first objective were tested using targeted aqueous analysis samples, while the second and third objectives used an LC-HRMS analytical approach.

1.1.1 Objective 1: Examining chemical and biological signatures of OWTS proximity in five Minnesota lakes

The first objective was to evaluate the aqueous concentrations of TOrCs and corresponding bioassay results of fish from locations proximal and distal from shoreline residences in five Minnesota Lakes. The following hypothesis was tested:

Hypothesis 1: Elevated TOrC concentrations and positive endocrine disruption assays will occur at lake locations proximal relative to those distal from shoreline residences

Most prior studies relied on a whole-lake aqueous sampling approach, relying on a single sampling location and assuming it represented TOrC occurrence throughout the lake (Ferrey et al., 2015; Writer et al., 2010). This approach fails to assess intralake spatial and temporal heterogeneity in TOrC concentrations, which could address research questions regarding TOrC sources and lakewater estrogenic activity. This hypothesis was tested by conducting a spatial analysis of aqueous TOrC concentrations in Minnesota lakes supported by targeted quantitation for a suite of common wastewater indicators, known endocrine disruptors, and frequently detected chemicals from previous monitoring efforts. Aqueous samples were taken from spawning habitats of bluegill sunfish where adult males were also captured to conduct *in vitro* assays of endocrine disruption and assess holistic fish health.

1.1.2. Objective 2: Prioritizing potential causative agents of endocrine disruption using LC-HRMS

The second objective was to conduct LC-HRMS analysis on a subset of samples from Objective 1 to provide a more robust characterization of potential causative agents of observed endocrine disruption. The following hypothesis was tested:

Hypothesis 2: Sites with positive VTG assays will contain unique LC-HRMS features associated with estrogen agonist potency

Notable detections from the targeted analysis in Objective 1 could not explain pronounced intralake differences in results from VTG assays(Guyader et al., 2018a). LC-HRMS analysis allows for a more robust characterization of environmental matrices by examining unknown chemical components. Identifying these components is labor intensive, particularly when considering the extensive number of unique chemical features detected within a sample. For this reason, statistical prioritization of features in accordance with research objectives is now widely used as an integral part of LC-HRMS data processing. This hypothesis was tested by executing a t-test on LC-HRMS feature intensities. Toxicological endpoints were exploited to bin samples as either endocrine active or reference. Identification of features with higher intensities in the endocrine active group was then attempted using spectral libraries. Principal component variable grouping assessed if prioritized features were site-specific or shared by all locations associated with endocrine activity. Finally, the estrogenic potency of high confidence feature identifications was evaluated using contemporary database-driven methods curated by the U.S. EPA.

1.1.3 Objective 3: Delineate OWTS-associated LC-HRMS features at an adjacent Minnesota lake

The third objective was to conduct a suspect screening of OWTS-derived TOrCs in adjacent lake water. The following hypothesis was tested:

Hypothesis 3: A lake location closer to an OWTS will contain higher intensities of TOrCs present in a shoreline residence's septic tank and drain field

TOrC occurrence in a single family residence's septic tank is highly variable (Teerlink et al., 2012). While the chemicals used by a single household are generally finite, patterns of usage and discrete discharges to the wastewater stream are irregular. As a result, septic characterization is challenged by considerations of sampling bias. To test this hypothesis, passive samplers were deployed for 2, 4, and 8 days in the septic tank and drain field of a Minnesota shoreline residence's OWTS. The features identified with high confidence that occurred across all field replicates defined an interest list of monoisotopic masses for locally relevant, OWTS-derived TOrCs. A suspect search of these chemicals was then executed on extracts from passive samplers deployed at two lake locations: one adjacent to the OWTS and a reference lake distant from any shoreline residences. Multivariate differences in feature responses at lake locations was then assessed using principal component analysis. In addition, passive sampler characterization of the OWTS was assessed by evaluating temporal trends in interest list feature intensity, and comparing the interest list to features prioritized by transient aqueous samples concurrently collected from the septic tank and drain field.

1.2 Dissertation Organization

This dissertation is organized into five chapters. This chapter (Chapter 1) explains the broader motivation for the hypotheses tested and associated research objectives that make up this dissertation. The main body of the dissertation (Chapters 2 - 4) describes greater detail of the motivation, experimental methods, results, and conclusions or the three research studies executed to address the three research objectives. Each of the main body chapters were modified from manuscripts that were either published, submitted for peer review, or in preparation for submission to a peer reviewed journal. Supporting information for these chapters is provided in the Appendices. The following describes each of the chapters within the main body of the dissertation:

- Chapter 2, "Trace Organic Contaminant (TOrC) Mixtures in Minnesota Littoral Zones: Effects of On-site Wastewater Treatment Proximity and Biological Impact", by Meaghan E. Guyader (primary researcher and author), Les D. Warren (M.S. graduate from St. Cloud State University, collected field samples and executed bioassay analysis), Emily Green (B.S. graduate from Colorado School of Mines, assisted with aqueous sample preparation and analysis), Riley Bertram (B.S. Student of St. Cloud State University, assisted with field sampling and biological laboratory analysis), Andrew P. Proudian (Ph.D. Candidate at the

Colorado School of Mines, contributed parametric spatial analysis of aqueous results), Richard L. Kiesling (hydrologist at the Unite States Geological Survey, supervised field sampling), Heiko L. Schoenfuss (Professor at St. Cloud State University, principal investigator supervising biological analyses), and Christopher Higgins (Associate Professor at the Colorado School of Mines, principal investigator of chemical analyses and corresponding author), was published in Science of the Total Environment (Guyader et al., 2018). This paper addresses Objective 1 and Hypothesis 1. Science of the Total Environment automatically grants copyright approval to reuse published manuscripts in dissertations. Approval for republication was confirmed by all co-authors.

- Chapter 3, "Suspect Analysis of On-site Wastewater Treatment Trace Organic Contaminants in Passive Samplers Deployed in an Adjacent Minnesota Lake", by Meaghan Guyader (primary researcher and author), Les D. Warren (M.S. graduate of St. Cloud State University, collected samples and executed biological analyses), Craig Butt (Product Application Specialist at Sciex, supported LC-HRMS feature identification efforts), Gordana Ivosev (Research Scientist at Sciex, critiqued statistical LC-HRMS feature prioritization methods), Richard L. Kiesling (Hydrologist at the U.S. geological survey, supervised field sampling), Heiko L. Schoenfuss (Professor at St. Clous State Unversity, principal investigator supervising biological analyses), and Christopher P. Higgins (Associate Professor at the Colorado School of Mines, principal investigator supervising chemical analyses and corresponding author), was submitted to Science of the Total Environment. This paper address Objective 2 and Hypothesis 2. Science of the Total Environment automatically grants copyright approval to reuse published manuscripts in dissertations. Approval for republication was confirmed by all co-authors.

- Chapter 4, "Suspect Analysis of On-site Wastewater Treatment Trace Organic Contaminants in Passive Samplers Deployed in an Adjacent Minnesota Lake", by Meaghan Guyader (primary researcher and author), Emily Green (B.S. graduate from Colorado School of Mines, assisted with aqueous sample preparation and analysis), Richard L. Kiesling (hydrologist at the Unite States Geological Survey, supervised field sampling), Heiko L. Schoenfuss (Professor at St. Cloud State University, principal investigator), and

Christopher P. Higgins (Associate Professor at the Colorado School of Mines, principal investigator supervising chemical analyses and corresponding author), is currently in preparation for submittal to Science of the Total Environment. This paper address Objective 3 and Hypothesis 3. Science of the Total Environment automatically grants copyright approval to reuse published manuscripts in dissertations. Approval for republication was confirmed by all co-authors.

CHAPTER 2

TRACE ORGANIC CONTAMINANT (TORC) MIXTURES IN MINNESOTA LITTORAL ZONES: EFFECTS OF ON-SITE WASTEWATER TREATMENT SYSTEM (OWTS) PROXIMITY AND BIOLOGICAL IMPACTS

Modified from an article published in *Science of the Total Environment*[1]

Guyader, Meaghan E.[2,3]; Warren, Les D.[4]; Green, Emily[5]; Bertram, Riley[4]; Proudian, Andrew P.[6]; Kiesling, Richard L.[7]; Schoenfuss, Heiko L.[4]; Higgins, Christopher P.[2,8]

2.1 Abstract

On-site wastewater treatment systems (OWTSs) are an international wastewater management strategy for rural and semi-rural communities without access to centralized sewage treatment. These systems are a suspected source of trace organic contaminants (TOrCs) that may be responsible for endocrine disrupting effects to resident fish species in Minnesota Lakes. This study assessed localized porewater concentrations of TOrCs in near-shore environments across five Minnesota Lakes. Sampling sites were designated as either likely (HOME) or unlikely (REF) to receive OWTS discharges based on their proximity to shoreline households. Sampling sites also served as sunfish spawning habitats concurrently studied for biologic impacts to resident adult males. Two-group hypothesis tests demonstrated significantly ($p = 0.02$) higher total TOrC concentrations in HOME (Mean = 841 ng/L) versus REF (Mean = 222 ng/L) sites. HOME sites also contained a wider suite of TOrC detections relative to REF sites. The distance to the nearest household (most proximal distance; MPD) negatively correlated ($r = -0.62$) with total TOrC concentrations. However, 2,4-D and DEET were major contributors to these total concentrations, suggesting that anthropogenic influence from households may not be exclusively attributed to OWTS discharges. Further, TOrC presence and elevated nitrogen concentrations in REF site porewater suggest additional, non-household TOrC discharges to these lakes. Significantly higher blood concentrations of vitellogenin ($p = 0.03$) and 11-ketotestosterone ($p = 0.01$) were observed

[1]Reprinted with permission from Science of the Total Environment
[2] Civil and Environmental Engineering Department, Colorado School of Mines
[3]Primary researcher and author
[4]Biology, Department, St. Cloud State University, St. Cloud, MN, U.S.
[5]Chemical Engineering Department, Colorado School of Mines
[6]Applied Physics Department, Colorado School of Mines
[7]Minnesota U.S. Geological Survey
[8]Corresponding author

in adult male sunfish captured from HOME versus REF sites. Comparisons between chemical and biological data indicate enhanced bioactive effects of co-contaminants. The findings from this study demonstrate multiple diffuse transport pathways contribute to the presence of biologically active TOrC mixtures in Minnesota Lakes, and mitigation efforts should consider minimizing residential inputs of chemicals associated with both outdoor and OWTS activity.

2.2 Introduction

Trace organic contaminants (TOrCs) represent many emerging contaminants prioritized in current environmental monitoring efforts. TOrCs encompass pharmaceuticals, herbicides, pesticides, hormones, steroids, personal care products, cleaning agents, and food preservatives detected at low level (ng/L) concentrations throughout the environment (Fatta-Kassinos et al., 2011; Kolpin et al., 2002; Lapworth et al., 2012). Many of these chemicals are not currently regulated despite the association of several TOrCs with adverse biologic impacts, particularly endocrine disrupting effects (Ortiz de García et al., 2014). Endocrine disrupting chemicals mimic or inhibit normal androgen or estrogen receptor function, resulting in abnormal virilization or feminization of affected species, respectively (Diamanti-Kandarakis et al., 2009; Söffker et al., 2015). The endocrine disruption capabilities of individual TOrCs are defined using laboratory exposure experiments through biochemical, histological, and behavioral endpoints (Blair et al., 2000; Elliott et al., 2014; Han et al., 2010; Oropesa et al., 2016). Still, the endocrine activity of environmental TOrC mixtures are poorly understood, particularly in light of the likely co-occurrence of unknown TOrCs (with unknown biological activities) and the potential for co-contaminants to enhance biological impacts (McCarty and Borgert, 2006).

Minnesota littoral zones are a prime field environment for studying diffuse sources of TOrCs and subsequent effects to aquatic life. Previous assessments of nutrient loadings in the United States (US) have led to the consensus that diffuse sources, such as agricultural runoff, groundwater infiltration, and atmospheric deposition, are responsible for most water quality degradation (U.S. Environmental Protection Agency, 1996). The lack of discrete inputs within proximity of surveyed Minnesota water resources indicate diffuse sources are also responsible for widespread TOrC occurrence in these waters (Erickson et al., 2014; Ferrey et al., 2015; Writer et al., 2010). In addition, many lakes in the state (90% of those surveyed by Writer et al.) also contain adult male fish with elevated vitellogenin concentrations, a biomarker of fish feminization (Writer et al., 2010). The spatial and seasonal heterogeneity of TOrC presence, both across lakes and within the same lake, impedes alignment of current chemical and biological observations (Baker et al., 2014). More strategic sampling methods, specifically sampling porewater in littoral zones (near-shore environments with depth < 5 m) during the spring and summer months should enable better characterization of biologically active TOrC mixtures that affect Minnesota fish species. Littoral zones serve as spawning habitats for fish species, such as the commonly studied bluegill sunfish *lepomis macrochirus*. Spawning season is a critical time of TOrC exposure for larvae and the adult male sunfish that guard them (Becker, 1983). On-site wastewater treatment systems (OWTSs) are one of the proposed diffuse sources affecting the health of these fish (Baker et al., 2014; Writer et al., 2010). Analysis of sediment porewater in these locations advantageously provides insight into the TOrC concentrations of inflowing, potentially OWTS-impacted shallow groundwater and relevant exposure concentrations to fish interacting with lake sediments while spawning.

OWTSs are a documented diffuse source of wastewater-derived contaminants in groundwater, drinking water wells, and surface waters around the world (Gago-Ferrero et al., 2017; Godfrey et al., 2007; Phillips et al., 2015; Schaider et al., 2014; Subedi et al., 2015). This method of treatment typically serves rural and semi-rural populations without access to centralized sewage, around 25% of the population in the United States and 20% of Minnesotans (U.S. Environmental Protection Agency, 2014; West, 2008). Removal of nutrients, suspended solids, and pathogens is achieved by percolating pre-treated wastewater through unsaturated native soils (Crites and Tchobanoglous, 1998; Stanford et al., 2010). TOrCs readily sorbed or biotransformed in these subsurface conditions are also effectively attenuated, even though their removal is not considered in OWTS design (Conn et al., 2010, 2006; Teerlink et al., 2012b). Nevertheless, certain TOrCs remain recalcitrant to modern wastewater treatment technologies (Du et al., 2014; Wode et al., 2015). For this reason, several TOrCs, such as carbamazepine, are now designated as environmental domestic wastewater indicators (Kahle et al., 2009). Furthermore, the ability of OWTSs to effectively treat heterogeneous inputs of TOrCs at their small sewershed scale is highly variable (Teerlink et al., 2012a). Out-of-compliance systems, attributed to either improper installation or maintenance, allow insufficiently treated wastewater to reach the water table and enter shallow groundwater flow paths (Bremer and Harter, 2012; Yates, 1985). In addition, cesspool and leach pit OWTS designs have a decreased ability to remove TOrCs compared to a conventional two-stage system (Schaider et al., 2017). Transport through the subsurface poses an environmental health risk, and can lead to diarrhea outbreaks in children consuming water from OWTS-impacted drinking wells (Borchardt et al., 2003). An estimated 21% of OWTSs in Minnesota are operated out of compliance (Robinson and Schultz, 2015), but subsurface clay moraines with low hydraulic conductivity may also compromise soil driven treatment in the region (Engelking and Kovacevic, 2016). While advanced treatment options, such as aerated biofilters, are available, current OWTS regulation does not require their implementation (Jantrania and Gross, 2006). In light of the potential for OWTSs to serve as sources of TOrCs to Minnesota waters, it seems prudent to evaluate their occurrence in relation to OWTSs and their potential biological impacts before additional steps are taken to reduce these potential impacts.

The objectives of this study were to characterize targeted TOrC mixtures in littoral zones affected by discharges from OWTSs, and evaluate potential associations between these TOrCs and biological impacts to adult fish. We hypothesized that locations more proximal to shoreline households would have more TOrC detections at higher concentrations and these locations would contain sunfish with elevated biomarkers of endocrine disruption. The following research questions were addressed: (1) what TOrC mixtures are present at near shore environments in Minnesota Lakes, (2) are there significant compositional differences between sites likely impacted by OWTSs versus those which likely are not, (3) how do localized environmental TOrC mixtures affect biological responses in fish species? To address these questions, targeted aqueous analysis of porewater grab samples from spawning habitats in five Minnesota Lakes were compared to endpoints of biologic impact in captured adult male sunfish.

2.3 Materials and Methods

2.3.1 Site Selection

Five lakes were selected for this study: Cedar Lake (Wright County, MN), Franklin Lake (Otter Tail County, MN), Lake Mary (Wright County, MN), Pearl Lake (Stearns County, MN), and Sullivan Lake (Wright County, MN) (Figure 2.1). Lakes were chosen based on the following criteria: influence of groundwater, presence of suitable bluegill nesting habitats, shoreline development greater than 30 percent, and the use of OWTSs for wastewater treatment of domestic wastewater (regardless of OWTS functionality). The lakes are from the North Central Hardwood Forest Ecoregion from the EPA's EcoRegion III classifications (U.S. Environmental Protection Agency, 2013). Lakes were surrounded by residences, agricultural croplands, and several municipal buildings including churches and a summer camp. Each lake is also associated with recreational activities, including a public access boat ramp and regular stocking for recreational fishing (Minnesota Department of Natural Resources, 2017). Likelihood of groundwater influx was examined using historical water tables, groundwater flow, and stable isotope data collected from U.S. Geological Survey (USGS) Hydrologic Atlases and Minnesota Department of Natural Resources County Geologic Atlases (Adams, 2016; Lujan Jr. and Peck, 1992). A preliminary survey in the summer of 2015 verified groundwater influx at the candidate lakes. Areas with persistent temperature and specific conductance differences between sediment porewater and overlying surface water were deemed as “gaining” due to groundwater input and selected as sites for the study. Sediment porewater was pumped to the surface with a mini piezometer to monitor basic water quality measurements, such as temperature and specific conductance for verification of the influence of groundwater. The use of bed sediment temperatures as an indicator of groundwater inputs is relatively accurate in determining the influence of groundwater on surface water in riverine systems (Conant Jr., 2004) and lakes (Constantz et al., 2007; Jones, 2006).

Four active sunfish spawning habitats per lake were chosen according to their proximity to shoreline households: two sites which likely received discharges from household OWTSs (HOME) and two reference sites (REF) likely unaffected by OWTS inputs. This study design assumes that the littoral zone in each lake sampled was not well mixed and discrete inputs would result in localized chemical and biological signatures of impact from OWTSs. Explicitly, REF sites are used as a control group in this study rather than reference lakes. As undeveloped lakes are anomalous with respect to land use characteristics in the Upper Midwest where lakeshore properties are highly desirable, the use of reference lakes would not have strengthened the experimental design of the study. HOME and REF were distinguished in the field as sites with or without a household in eyesight of the spawning habitat, respectively.

Site distinctions for two-group analysis were re-examined after sampling using aerial imagery analysis. OWTS locations are extremely difficult to obtain; even with access to their public records, designs are usually detailed by hand-drawn representations with minimal geographic information. Therefore, household locations were used as a proxy. Household locations were recorded based on Google Maps imagery (Maps, 2017). Latitude and longitude direct decimal (DD) coordinates were recorded as the center of visible households or the center of household plots covered by trees after verifying addresses with Google Map streetview. Sampling location coordinates were obtained in the field using a global positioning system (GPS). The distance between each sampling location and all shoreline households within 100m of lakeshore were determined using the Euclidean distance technique (Gower, 1982). The nearest household at each sampling location was determined by calculating the minimum of the set of distances attributed to each sampling location. The minimum distance at each site is herein referred to as the most proximal distance (MPD).

2.3.2 Sample Collection

Aqueous Samples. Porewater grab samples were collected between the months of May and July 2016. Samples were collected by pumping porewater through piezometers to the surface and accumulating water in appropriate sampling vessels. Piezometers were driven into the sediment until the screened terminal end of the probe reached saturated conditions determined to contain inflowing groundwater (~0.5-1 m depth), as verified by the methods described in the site selection section. Once placed, a peristaltic pump (Geotech Environmental Supply, Denver, CO, U.S.) was connected to the piezometer and porewater was pumped (~35 mL/min) for collection at the surface. This pumping rate was chosen so that the rate of groundwater replenishment at the sampling point was not exceeded. Pumping equipment was rinsed thoroughly with filtered water (Omni Water; Thermo Fisher Scientific, Walthman, MA, U.S.) before collection at each site to prevent cross contamination. Samples intended for TOrC analysis were collected in 1L amber glass bottles pre-cleaned by scrubbing with liquinox, rinsing with DI water, and triple rinsing with reagent grade methanol. Samples for inorganic analysis were collected in 50mL polypropylene tubes. All samples were stored and shipped at 4°C to the Colorado School of Mines (CSM; Golden, CO, U.S.) for further analysis.

Biological Samples. Resident sunfish samples were collected concurrently with the aqueous grab samples. Male fish were collected directly off spawning beds by rod and reel (permitted by Minnesota Department of Natural Resources). During the spawning season, male sunfish will continuously defend their small nest site (approximately 0.5 m in diameter) making them unable to forage for food and readily accepting of baited hooks. This behavior ensures that only nest defending males, who have likely been site-bound for days or weeks, were captured as non-nest holding sneaker males are able to feed between forays into the spawning grounds. A total of 124 male fish were collected from OWTS-influenced sites and 116 from reference sites. On a lake basis, 98 sunfish were collected from Sullivan Lake, 83 from Lake Mary, 35 from Cedar Lake, of 2016 disrupted spawning activity (i.e. n_{bio} = 8 rather than 10). and 24 from Lake Franklin. Fish were not collected at Pearl Lake because early ice-off in the spring. Males captured were immediately euthanized using a buffered MS-222 solution approved by the St. Cloud State University Institutional Animal Care and Use Committee (IACUC # 8-77). A whole blood sample from the caudal vein was taken using a 22-gauge needle, stored on ice, and transferred to St. Cloud State University (SCSU; St. Cloud, MN, U.S.) for centrifugation. Fish carcasses were placed on ice and transferred to SCSU for dissection and further analysis.

2.3.3 Sample Analysis

Chemical analysis. Basic porewater characteristics at each site, including temperature, pH, specific conductance, and dissolved oxygen (DO) concentrations were collected in the field using a Yellow Springs Instrument (YSI; Yellow Springs, OH, U.S.) probe. Surface water at each site was also measured for DO concentrations, specific conductance, and temperature to confirm adequate environmental conditions to support fish spawning.

Samples received at CSM intended for TOrC analysis were filtered using Whatman GF/F filters and spiked with surrogate standards (20 ng each of 48 unique stable isotope standards; Sigma Aldrich, St. Louis, MO, U.S.) within 48 hours of sampling. Samples were then enriched using solid phase extraction (SPE) within two weeks of spiking with surrogate. SPE was executed using an AutoTrace 280 as follows: 6cc 500 mg Oasis HLB cartridges were pre-conditioned with 5 mL methyl tert-butyl ether (MTBE), followed by 5 mL methanol, and then 5 mL HPLC water. The 1L sample was then loaded onto the cartridge. The cartridge was washed with 10 mL HPLC water, and then dried for 60 minutes using nitrogen gas. Cartridges were eluted with 5mL methanol followed by 5 mL of 90/10 Methanol/MTBE (% v/v). The extract was blown down to 500 µL using an N-Evap system, then reconstituted in 2 mL of methanol. Methanol extracts were diluted 10:1 in ultra-pure water (Thermo Fisher Scientific, Walthman MA, U.S) and analyzed using a AB Sciex 3200 QTRAP liquid chromatograph tandem mass spectrometer (LC-MS/MS). Sample data was acquired with two 1 mL injections: one positive electrospray ionization method (ESI +) and one negative electrospray ionization (ESI -) method. Targeted analytes were selected to represent common indicators of domestic wastewater and TOrCs with known endocrine disrupting capabilities (Table A.1). For ESI + runs, a Phenomenex Luna C18 column (Phenomenex, Torrance, CA, U.S) was used with water and methanol eluents buffered with 10mM ammonium formate and 1 mL formic acid (Thermo Fisher Scientific, Walthman MA, U.S). The ESI – method used a Phenomenex Gemini C18 column (Phenomenex, Torrance, CA, U.S) with water and

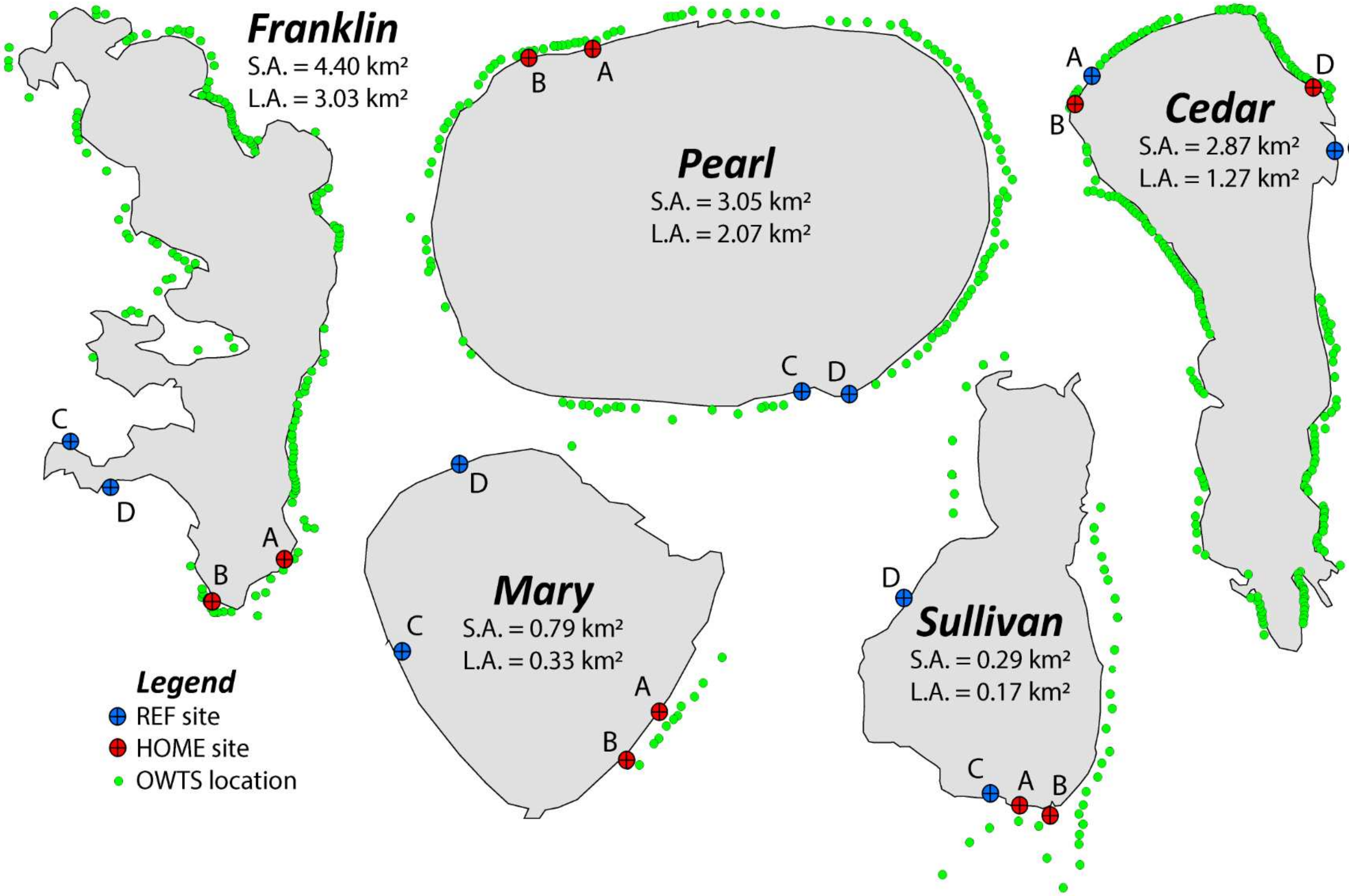

Figure 2.1 Summary of sampling locations within each of the five lakes samples. Lake characteristics are provided under each lake's name, where S.A. = surface area, L.A. = littoral area

methanol eluents buffered with 5mM ammonium fluoride (Thermo Fisher Scientific, Walthman, MA, U.S).

Inorganic samples were prepped and preserved at CSM in accordance with their intended instrument analysis. All inorganic samples were filtered using 0.4 µm syringe filter within 48 hours of sampling. Major nutrients were analyzed using ion chromatography (Dionex Thermo Fisher ICS-900) using unacidifed aliquots of the filtered sample. Total nitrogen and total organic carbon were assessed with sample acidified using hydrochloric acid on a Schimadzu TOC system (Shimadzu TOCV-TNM-LCSH). Trace metals were quantified using inductively coupled plasma atomic emission electroscopy (ICP-AES) with filtered aliquots acidified with nitric acid.

Biological Samples. Glucose concentrations in whole blood samples were measured using a TRUEbalance Blood Glucose Monitor (Moore Medical, Farmington, CT, U.S.). Blood samples were then centrifuged (8,000 x g) for 12 min at 4°C, plasma was pipetted into separate vials and frozen at -80°C until analysis. Plasma vitellogenin concentration were determined through an enzyme linked immunosorbent assay (ELISA) using purified sunfish vitellogenin and sunfish validated vitellogenin antibodies. The protocol followed parameters as used in Schultz et al. 2013 (Schultz et al., 2013). 11-ketotestosterone concentrations were determined using an ELISA (Cayman Chemical Company, Kit #582751) following the manufacturer's guidelines. Wet weight of fish carcasses was determined upon return to SCSU (within 6 hrs of fish capture), and liver and gonad were excised and weighed. From these values, body condition factor (weight/(total length)3 x 100,000), hepatosomatic index (liver weight/ mass fish x 100), and gonadal somatic index (gonad weight/ mass fish x 100) were calculated (Bolger and Connolly, 1989; Fulton, 1904).

2.3.4 Quality Assurance and Quality Control

Reported aqueous chemistry data were subject to various field and laboratory quality control measures. All grab samples were collected in triplicate along with field blanks and equipment blanks collected at each lake. Field blanks were collected at each lake by passing purified water (OmniWater, Thermo Fisher Scientific, Walthman, MA, U.S.) through the piezometer pumping set up. Laboratory blanks and instrument blanks were also included to ensure contamination introduced during sample handling could not skew reported results. Raw data from targeted LC-MS/MS data acquisition was initially processed using Sciex's MultiQuant software. Quantitation limits for each analyte were set as the concentration of the lowest calibration standard included in a valid calibration curve. Valid calibration curves needed to include at least four points, exclude points with possible instrument or laboratory contamination (defined as containing analyte peak area greater than three times analyte peak area in the blank run before the calibration standards), and have a (1/x)- or (1/x2)-fit trendline with a Pearson's r value greater than 0.99. Analyte signals influenced by field contamination were redacted by setting reporting limits above the quantitation limit. Reporting limits were set as the value of the average field and equipment blank concentrations plus three times the standard deviation. For analytes with only one field blank or equipment blank with quantifiable contamination, this value was set as three times the field blank concentration (Table A.2). Average concentrations were reported if at least two out of three replicates had quantified results. Sites with detections in zero or only one of the three replicates were reported as below the quantitation limit. Below reporting limit and below quantitation limit values were handled as censored values during statistical analysis according to Helsel 2012 (Helsel, 2012). Censored data analysis methods are explained in the Statistical Analysis section. Instrument performance was verified with regular calibration standards and calibration verification standards. TOrC results were also constrained by matrix spike and surrogate recoveries (Table A.1). Furthermore, analytes were redacted if their matrix spike recoveries were outside the 70-130% range or if surrogate recoveries were less than 10%. Thirty-seven of the TOrCs analyzed met all QA/QC requirements and were used for statistical comparisons between HOME and REF sites.

Biological data were also subject to quality assurance and quality control procedures. For plasma vitellogenin measurements, all samples were analyzed at three dilutions (1:50, 1:250, and 1:1,000). An eight-point standard curve was then used to reference absorbance readings of samples. Four replicate samples were added to each plate and replicate samples were added across plates. All samples were randomized across plates.

2.3.5 Statistical Analysis

Non-parametric statistical comparisons of biological and chemical data were executed as two-group sample hypothesis tests between HOME (nbio = 8, nchem = 10) and REF (nbio = 8, nchem = 10) sites. Rejection of the null hypothesis was considered valid for p-values less than 0.05. Tests on aqueous chemistry data were conducted using the "NADA" package in RStudio to ensure proper handling of the numerous left-censored data in both organic and inorganic targeted datasets (Lee, 2017; Rstudio Team, 2015). Average concentrations of all inorganic and organic analytes, as well as the sum of the average targeted TOrC concentrations, referred to hereafter as total TOrC concentration, in HOME and REF groups were compared using the cendiff() function, a Mann-Whitney-Wilcoxon test of the empirical cumulative distribution functions within each group. This non-parametric statistical test does not assume a normal distribution of values within each group, which is appropriate for comparing concentrations across lakes from different geographical regions, unique OWTS owners, and different resultant baseline TOrCs in the respective lake systems. The larger number of biological samples collected allowed for HOME and REF two group tests to be executed at both inter- and intra-lake levels. Intra-lake comparisons were assumed to have a normal distribution in biological endpoints. Mean comparisons from biological data were conducted with Tukey's honest significant difference (HSD) test and two-sided t-tests.

A bivariate analysis of the transformed variables ΔMPD and $\Delta TOrC$ provide a parametric assessment of OWTS proximity on TOrC concentrations. The variables are defined as:

$$\Delta MPD = MPD_{site} - MPD_{median} \qquad \textit{(Equation 2.1)}$$

$$\Delta TOrC = TOrC_{site} - TOrC_{median} \ldots \ldots \textit{(Equation 2.2)}$$

where MPD_{site} is the MPD value specific to a sampling location and MPD_{median} is the calculated median MPD value attributed to the four sampling locations at each lake. Similarly, $TOrC_{site}$ is the total TOrC concentration specific to a sampling location and $TOrC_{median}$ is the calculated median total TOrC concentration attributed to the four sampling locations at each lake. These transformed variables were used to allow better comparison across lakes. Explicitly, MPD values were modified to ΔMPD to better compare HOME/REF site selection across lakes with varied surface areas, and TOrC values were modified to $\Delta TOrC$ to better compare across lakes with varied background concentrations.

2.4 Results and Discussion

2.4.1 Site Distinction Analysis

MPDs were significantly ($p = 0.005$) lower at HOME sites relative to REF sites (Figure A.1). This corroborates the two-group distinctions used for non-parametric hypothesis testing of the chemical and biological results. Importantly, the MPD metric of proximity does not consider pronounced chemical and biological effects in regions that could be impacted by OWTS leachate from more than one system. Assessing impact from density, such as the approach used in Bremer and Harter et al., was also considered (Bremer and Harter, 2012); however, obtaining the exact location and compliance status of all relevant OWTSs would require a level of cooperation from homeowners unattainable at this time. Analysis of OWTS density in this study was therefore, determined to be too speculative and outside the scope of this research effort.

2.4.2 TOrC Detections and Non-Parametric Two Group Comparisons

Fifteen of the thirty-seven reported TOrCs were detected in at least one of the twenty sites sampled in this study (Table A.3). Pharmaceuticals, such as carbamazepine, are a preferred indicator of wastewater presence in an environmental matrix (Subedi et al., 2015). While detections of particular pharmaceuticals were not widespread enough to generate meaningful statistics comparing HOME and REF sites, it is noted that the pharmaceuticals carbamazepine, dilantin, and ibuprofen were only detected at HOME sites (Tables A.3 and A.4). Interestingly, the synthetic estrogen 17α-ethinylestradiol (EE2) was only detected in two REF sites. EE2 is a known endocrine disruptor that was shown to collapse a fish population during a lake dosing study in Canada (Kidd et al., 2007). EE2 is used as both a synthetic birth control hormone and a livestock hormone to improve productivity and treat livestock diseases (Gadd et al., 2010). The REF sites with EE2 detections, SUL_D at 25 ng/L and CED_A at 5 ng/L are suspected to receive shallow groundwater carrying agricultural runoff from fields that use hormone-fed livestock manure as fertilizer (Zaharin Aris et al., 2014). These concentrations match the upper end of previously observed EE2 concentrations in surface waters (Zaharin Aris et al., 2014). In addition, bed sediments usually have higher EE2 concentrations attributed to the contaminants hydrophobic and persistent properties (Zaharin Aris et al., 2014). Therefore, the observed concentrations of EE2 in porewater sampled during this study are consistent with those expected in bed sediments (Zaharin Aris et al., 2014). The EE2 chemical identity was made using LC-MS/MS analysis only in this study, but confidence could be improved by examining the possibility of isobaric chemicals in the sampled matrices using LC-HRMS.

Other known or suspected endocrine active compounds that were commonly detected include the cosmetic preservative, methylparaben, and estrone, which were detected in 40% and 20% of the sites sampled, respectively (Bergman et al., 2012). The steroidal hormones androstenedione and testosterone, as well as 4-tert-octylphenol (used in the manufacturing of anionic surfactants, such as detergents or, less commonly, as an emulsifier in personal care products such as insect repellents) were also detected, but less frequently (detection frequency < 10%). Some TOrCs typically detected in environments down-gradient of OWTSs, such as sulfamethoxazole, were not detected in this study. These non-detections are attributed to the heterogeneity of inputs to wastewater treatment systems at such a small sewershed scale (Teerlink et al., 2012a), as well as the variability of subsurface conditions that affect TOrC removal in soil treatment units. Explicitly, these TOrCs may simply not have been used at the households within proximity to sampling locations, or soil regions were anaerobic and suitable for sulfamethoxazole attenuation (Massmann et al., 2008).

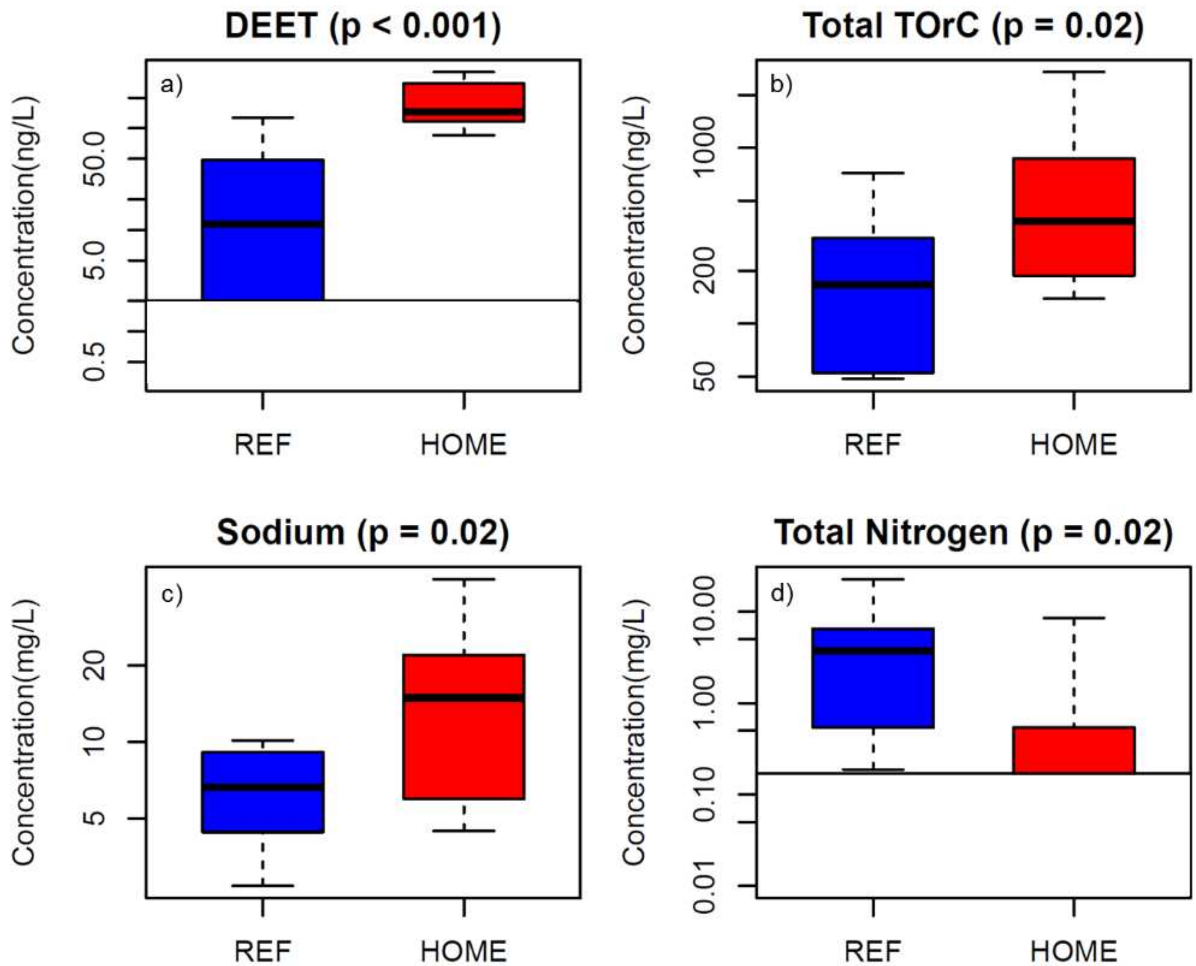

Figure 2.2 Two-group comparison of average site concentrations of a) DEET, b) total TOrC, c) sodium, and d) total nitrogen across all five lakes. Boxplots display minimum, first quartile, median, third quartile, and maximum values specific to each group. First quartiles below the detection limits of DEET and total nitrogen, 2 ng/L and 0.17 mg/L respectively, are omitted.

The most frequently detected analyte (detection frequency = 85%) was N,N-diethyltoluamide, more commonly referred to as DEET. DEET is neither a persistent or bioaccumulative organic pollutant, with a half-life in the order of days to weeks as well as acute and chronic effect concentrations orders of magnitude above observed environmental concentration (Weeks et al., 2012). This insect repellent ingredient has been previously detected in lakes across the state of Minnesota where lake recreation and mosquitoes are very common in the summer months (Ferrey et al., 2015; Writer et al., 2010). This study is the first to note significantly higher concentrations of DEET in Minnesota lake sites more proximal to households across all lakes sampled (Figure 2.2a). DEET could enter household wastewater streams through bathing or clothes washing which could then enter OWTS discharges. Gago-Fererro et al. also noted seasonally high concentrations of DEET in surface waters adjacent to OWTSs at concentrations an order of magnitude lower than those observed in our study (average summer sampling concentration of 13 ng/L) (Gago-Ferrero et al., 2017). However, designating it as a wastewater indicator in these lake systems is inappropriate (Tran et al., 2014). The associated outdoor usage of DEET suggests that this contaminant may also be entering lake systems from general anthropogenic activity in and around lakes.

The total TOrC concentrations at HOME sites were significantly ($p = 0.02$) higher compared to those at REF sites (Figure 2.2b). A more detailed presentation of the measured analyte concentrations at each lake site is provided in Figure 2.3. HOME sites contained a wider suite of targeted analytes (14 out of the 37 reported) as compared to REF sites (7 out of the 37 reported). Figure 2.3 also shows the consistently higher concentrations of specific TOrCs detected at the HOME versus REF sites. Lake Mary's HOME sites, MAR_A and MAR_B, had the highest measured total TOrC concentration across all sites assessed in this study. This is mainly attributed to the measured concentrations of the herbicide 2,4-D at both of the lake's HOME sites (mean of 2200 ng/L). 2,4-D is commonly used for outdoor home gardening applications suggesting that, similar to DEET, household activity other than OWTSs are impacting littoral porewater of these lakes (Mnif et al., 2011). 2,4-D is also one of the few measured TOrCs monitored by the EPA with a maximum contaminant level (MCL) in drinking water of 50 µg/L as a result of its association with blood, kidney, and liver toxicity (EPA, 1998).

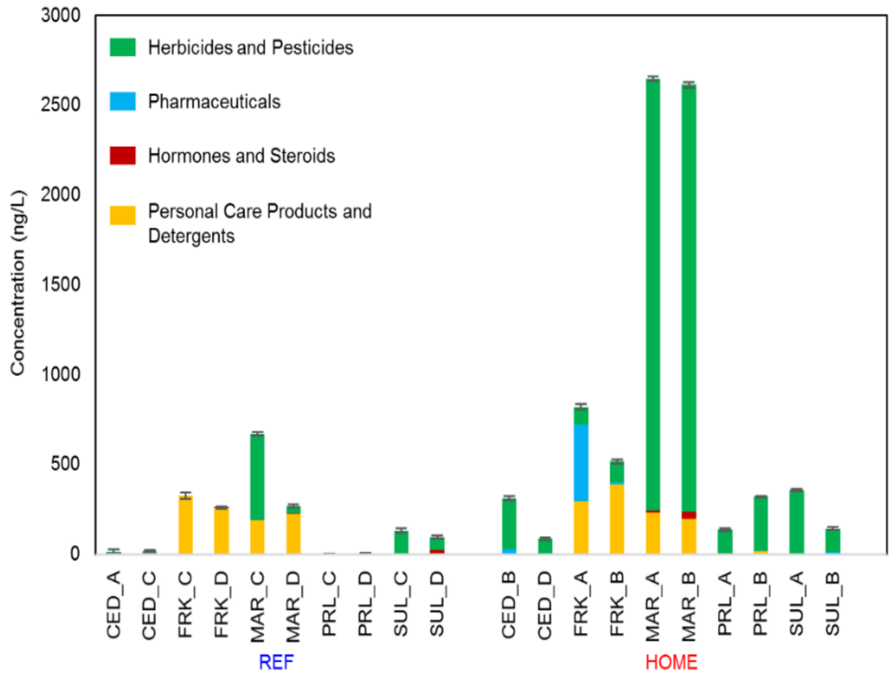

Figure 2.3 Average total TOrC concentrations measured at each site. Detected analytes are color coded by compound use. Error bars represent +/- average standard deviation across all TOrCs measured at each site.

Non-household diffuse sources are expected to contribute to the “background” presence of TOrCs, such as DEET and oxybenzone, observed at many of the REF sites (Figure 2.3). The only sites DEET was not detected at were the REF sites PRL_C, FRK_C, and FRK_D. As suggested before, agricultural operations could contribute to TOrC occurrence, particularly herbicides, pesticides, and feedlot hormones, in the REF sites sampled. Agricultural fields surround all of the lakes, and the boundaries of these operations are in closer proximity to lake shoreline unoccupied by household lots. Recreational activities, such as boating, may also act as a non-point source of TOrCs in lake locations distant from households. Each lake sampled is a stocked fishery with household and public boat ramp access points (Minnesota Department of Natural Resources, 2017). Most lakeshore households have their own docks and boats (verified with aerial imagery; Google Maps accessed March 2017) (Maps, 2017). Oxybenzone, common in sunscreens, and DEET, common in insect repellents, are both TOrCs integrated into personal care products associated with these lake recreational activities. The transport processes resulting in their presence in groundwater-impacted lake porewaters for these contaminants is not immediately clear. Ferrey et al. proposed atmospheric deposition as a diffuse TOrC transport mechanism to Minnesota Lakes (Ferrey et al., 2015). DEET has been reported to be widely present in atmospheric samples (Balducci et al., 2012; Cheng and Lehmann, 1985). However, the significantly higher concentrations of DEET in sediment porewater near households in this study suggests long range aerial transport is unlikely. TOrCs introduced at the lake surface may enter into shallow groundwater after application near the water's surface. Contaminants may then settle with suspended solids in the lake and accumulate in the sediment where they may then partition into sediment porewater and reenter littoral zones with the influx of shallow groundwater (Winter et al., 1999).

2.4.3 Patterns in Inorganic Non-point Source Indicators

Basic water quality assessments of the sampled porewater confirm inflowing groundwater had "young" or shallow flowpaths (Table A.5). The consistently low (<5 mg/L) dissolved oxygen (DO) measurements of the porewater are typical for groundwater (Peterson and Risberg, 2009). There were no significant ($p > 0.05$) differences between porewater DO concentrations at HOME versus REF sites, but the maximum observed values were present in the REF sites. In addition, conductivity values across all sites' porewater samplers were at the lower end of the typical groundwater values (typical range 50-50,000 µS/cm, (Sanders, 1998)), suggesting influence from shorter groundwater flowpaths (Erickson et al., 2014). Lake Mary notably had the lowest porewater DO concentrations. Previous studies have noted that anoxic regions lead to longer range transport of TOrCs, as these pollutants are generally more effectively attenuated through aerobic degradation (Carrara et al., 2008; Phillips et al., 2015). We speculate anoxic regions are present in the subsurface surrounding Lake Mary and contribute to the higher detected TOrC concentrations in Lake Mary littoral sediment porewaters; however, a more thorough characterization of groundwater flow paths and redox conditions at this lake are required to test this hypothesis.

Nutrient and trace metal data were also compared to assess chemical differences between HOME and REF sites (Table A.6 and Table A.7). Previous studies have shown total nitrogen concentrations to positively correlate with TOrC occurrence, particularly TOrCs derived from OWTSs (Del Rosario et al., 2014; Phillips et al., 2015; Schaider et al., 2017). Surprisingly, there were significantly ($p = 0.02$) higher total nitrogen concentrations at the REF sites, further indicating the presence of an additional non-point source in the region (Figure 2.2d). This was not reflected by significant differences ($p > 0.05$) in the measured concentrations of nitrite or nitrate.

Unmeasured ammonia or organic nitrogen are suspected to be the dominant nitrogen species at these locations, with the exception of Pearl Lake's REF site C which had nitrate as the dominant nitrogen species. Pearl Lake's REF site C also had the highest total nitrogen concentration measured across all sampling locations. Each of the lakes sampled have agricultural fields surrounding them, particularly in non-residential parts of the shoreline, that could contribute to observed elevated nitrogen concentrations at REF versus HOME sites. There was no significant difference in total organic carbon between HOME and REF sites ($p > 0.05$). Many of the other nutrients analyzed were below detection limits. Chloride and bromide ratios can be used to assess sources of groundwater (Katz et al., 2011), but consistent censoring of bromide concentrations hindered this calculation. Trace metal analysis only showed significantly higher ($p = 0.019$) sodium concentrations in the HOME sites (Figure 2.2a). Sodium salts are common in detergents and other common household products as well as softening systems, which could explain this significant difference in HOME and REF concentrations. The wastewater tracer boron, also common in household products (Woods, 1994), showed no significant ($p > 0.05$) difference in concentrations between HOME and REF sites.

2.4.4 MPD Effect on Total TOrC Concentration

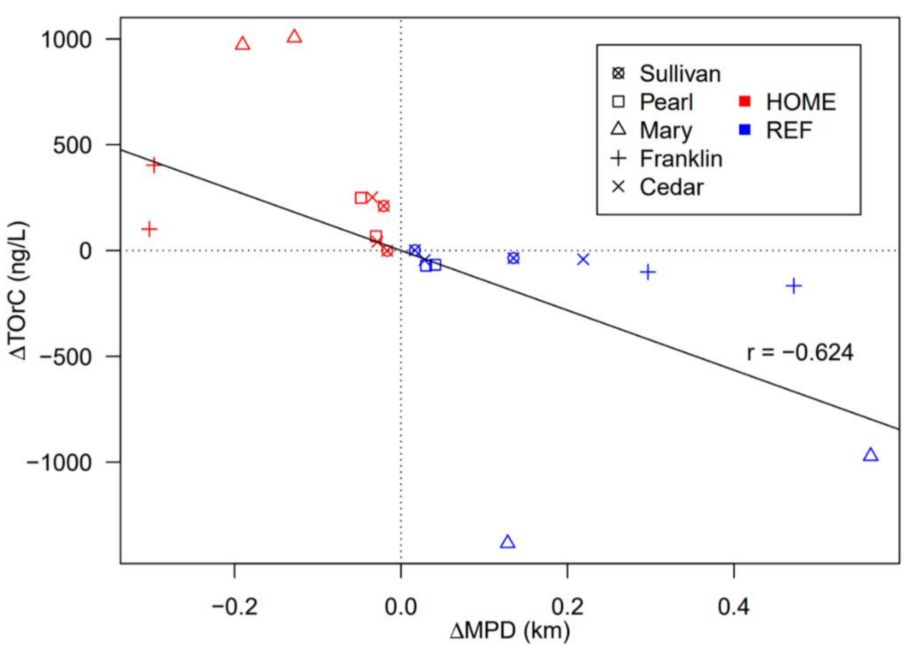

Figure 2.4 Bivariate analysis of residential proximity's effect on total TOrC concentration in littoral site porewater. Variables ΔMPD and ΔTOrC are defined in Equations 1a and 1b, respectively. Trendline: $y = a*x$, where $a = -1413.0 \pm 427.8$ ($p = 0.00374$); $r^2 = 00.3313$

Results from the parametric analysis are displayed in Figure 2.4. We anticipated HOME sites would have more positive $\Delta TOrC$ and negative ΔMPD values, in agreement with the hypothesis that sites more proximal to household OWTSs would have higher total TOrC concentrations. ΔMPD negatively ($r = -0.62$, $p < 0.001$) correlated with $\Delta TOrC$, supporting this hypothesis. Certain lakes showed more pronounced differences between HOME and REF sites than others. Specifically, Lake Mary's HOME and REF total TOrC concentrations showed the greatest difference, attributed to the high concentrations (~2200 ng/L) of 2,4-D at the lake's HOME sites. Sullivan, Cedar and Pearl lakes clustered around the origin of the plot, demonstrating poorer distinction between HOME and REF sites at these lakes. As expected, these sites with similar MPDs had less differentiation in total measured TOrC concentrations. The high concentrations measured at Lake Mary may be attributed to the lake's small surface area, along with many households at the south and eastern shorelines where the HOME site samples were collected. This spatial arrangement of households and OWTSs creates the potential for multiple wastewater streams to impact the adjacent littoral environments with minimal effects from dilution and attenuation. In addition, the DO readings for this porewater were very low (~ 1 mg/L), which has been associated with longer range transport of untransformed TOrC species in OWTS plumes (Carrara et al., 2008).

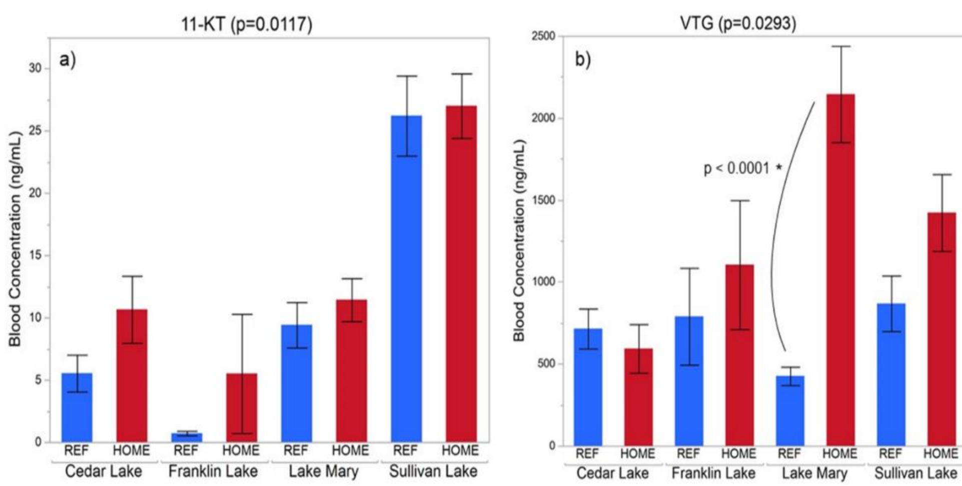

Figure 2.5 Average blood concentrations of a) 11-Ketotestostrone and b) vitellogenin in fish captured at HOME and REF within each lake. Error bars represent the standard deviation of values from the average concentration at each lake. p-values above the plots are comparisons of concentrations in fish captured from HOME and REF sites across all lakes. The in-plot p-value is the two-group comparison of HOME and REF captured fish from Lake Mary only.

2.4.5 Biological Data Two Group Hypothesis Tests

Field measurements of the lake surface water affirmed that the littoral environments could support healthy spawning at the sites sampled. DO levels were still hospitable for aquatic life (> 6 mg/L). Further, all observed aqueous temperatures were suitable for spawning with the exception of Pearl Lake (average = 12 °C), which experienced an early ice off. No sunfish were captured or analyzed from Pearl Lake; therefore, only chemical data are reported from this lake.

Two-group comparison of biological data demonstrated more pronounced biological impacts in HOME site spawning male sunfish with respect to vitellogenin and 11-KT concentrations (Figure 2.5). Increased concentrations of vitellogenin in male fish is a Tier 1 indicator of estrogen agonism and elevated concentrations of 11-ketotestosterone in male fish is a Tier 3 indicator of steroidogenesis according to the EPA's Endocrine Disruptor Screening Program (Borgert et al., 2014). All other biological endpoints measured, i.e. gonadal somatic index, glucose concentrations, and body condition factor showed no significant ($p > 0.05$) difference for captured fish at HOME and REF. Vitellogenin concentrations were significantly ($p = 0.0108$) higher in HOME versus REF sites, driven by the results from Lake Mary ($p < 0.0001$). Higher vitellogenin concentrations in HOME versus REF fish, particularly those inhabiting Lake Mary, suggest these males are being exposed to mixtures of TOrCs with estrogenic activity. This finding is particularly interesting when considering Lake Mary's HOME sites had the most detections with the highest concentrations. We speculate the notably high concentrations of 2,4-D may be enhancing estrogen agonist effects of known endocrine active co-contaminants measured at these sites. Even though 2,4-D is not itself consider an endocrine disruptor by the EPA's EDSP for the 21st century (EDSP21) dashboard (U.S. Environmental Protection Agency, 2017), a study by Kim et al. demonstrated 2,4-D and its transformation product 2,4-dichlorophenol (DCP) could enhance the androgenic effects of 5-dihydroxytestosterone (Kim et al., 2005). Steroidogenesis effects were also more pronounced at HOME sites as demonstrated by significantly higher concentrations of 11-KT in blood samples ($p = 0.0117$) when compared across all lakes. Cedar and Franklin Lake showed the most pronounced intralake site differences in 11-KT concentrations (Figure 2.5a). Interestingly, all fish sampled from Sullivan Lake, from both HOME and REF sites, had significantly higher ($p = 0.003$, two way ANOVA) 11-KT concentrations than fish sampled from other lakes. Previously reported 11-KT blood concentrations for adult male sunfish averaged at 13.8 ng/L (Knapp and Neff, 2007), suggesting Sullivan Lake's fish were anomalously high and

Franklin Lake's REF site was anomalously low. Chemical analyses did not indicate the presence of potent endocrine disruptors at Sullivan Lake, suggesting that the targeted analysis in this study may not sufficiently describe the localized TOrC mixtures present at the Sullivan Lake sites that may be impacting observed 11-KT concentrations.

2.5 Conclusions and Implications

The findings of this study suggest that TOrC occurrence in sunfish spawning habitats of Minnesota littoral environments are affected by groundwater inflows. Lakeshore households increase concentrations and detection frequency at adjacent lake locations. Minimizing TOrC loadings from households requires consideration of heterogeneous outdoor activity and domestic wastewater chemical compositions at shoreline residential locations. Hydrologic processes, such as stormwater infiltration, are suspected to increase the mobility of TOrCs in the subsurface and encourage transport from residences to littoral zones. Inputs from agricultural operations and recreational activity separate from residential locations must also be considered during mitigation efforts, particularly as this study suggests they may be sources of potent endocrine disruptors. Resultant endocrine disrupting effects are only partially justified by the TOrCs detected, suggesting total concentrations reported are merely a proxy for all components of biologically active mixtures in these environments. Non-targeted analysis with high resolution mass spectrometry could better resolve components of environmental TOrC mixtures that contribute to pronounced biologic activity (Schymanski et al., 2015).

2.6 Acknowledgements

Funding for the present study was provided by the Minnesota Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (Grant M.L. 2015, Chp. 76, Sec. 2, Subd. 04c). Several homeowners along the study lakes provided access to their property and facilitated our studies. In addition, we thank the Minnesota United States Geological Survey (USGS) for providing the equipment used for field measurements and sample collection. The expertise and assistance of Dr. Simon Roberts, Dr. Edward Furlong, Dr. Aniela Burant, and Karl Oetjen also helped greatly in the analytical chemistry techniques used in this study. Students in the Aquatic Toxicology Laboratory at St. Cloud State University assisted with dissections and histological preparations.

CHAPTER 3

PRIORITIZING POTENTIAL ENDOCRINE ACTIVE HIGH RESOLUTION MASS SPECTROMETRY FEATURES IN MINNESOTA LAKEWATER

A paper submitted to *Science of the Total Environment*[1]

Guyader, Meaghan E.[2,3]; Warren, Les D.[4]; Green, Emily[5]; Butt, Craig[6]; Ivosev, Gordana[7]; Kiesling, Richard L.[8]; Schoenfuss, Heiko L.[4]; Higgins, Christopher P.[2,9]

3.1 Abstract

Liquid chromatography high-resolution mass spectrometry (LC-HRMS) shows great potential for identifying causative agents of adverse biological effects within complex environmental chemical mixtures. This study exploited spatial differences observed in the biological activity of Minnesota lake porewater samples to conduct statistical prioritization of LC-HRMS features. Information-dependent mass spectra data acquisition of grab and temporally integrated lake porewater sample extracts was executed using a liquid chromatography quadrupole time-of-flight mass spectrometer (LC-QToF-MS) operated in both positive (ESI +) and negative (ESI -) electrospray ionization modes. Blood vitellogenin concentrations of resident adult male sunfish were used as a biological endpoint to bin sites from the same lake as either endocrine active (ACT; 2 sites) or reference (REF; 2 sites). Features at significantly higher intensities (p-value < 0.05, t-value > t-critical, log-fold change > 0.1; equal variance t-test of log2 transformed data) in ACT sites were then compiled into a suspect search list for feature identification. Adducts and isotopes of prioritized features were eliminated using pattern recognizing algorithms using mass, retention time, and intensity. Feature identities were reported according to established confidence metrics using spectral libraries and elemental composition algorithms. This LC-HRMS approach identified a number of features omitted by targeted analysis with higher relative abundances in ACT sites, including plant essential oils, fatty acids, and mycotoxins. Multivariate analysis

---

[1]Printed with permission from Science of the Total Environment

[2] Civil and Environmental Engineering Department, Colorado School of Mines
[3]Primary researcher and author
[4]Biology, Department, St. Cloud State University, St. Cloud, MN, U.S.
[5]Chemical Engineering Department, Colorado School of Mines
[6]Sciex; Farmingham, M.A.
[7]Sciex, Toronto, Canada
[8]Minnesota U.S. Geological Survey
[9]Corresponding author

determined whether features were either present at both sites (AB) or unique to individual ACT sites (A or B). Detection frequency across datasets indicated bias in feature prioritization influenced by the chosen sampling method and sample acquisition mode. The majority of features prioritized by this workflow remain tentatively identified or unidentified masses of interest, reflective of current limitations in shared spectral libraries for soft ionization analyses. Strategies similar to this workflow have the potential to reduce bias in database-driven toxicological prioritization frameworks.

3.2 Introduction

The continuous development of novel chemicals, their mass production and distribution in everyday household products, combined with the rapid advancement of analytical tools capable of measuring these chemicals in the environment continue to broaden the scope of contaminants of emerging concern. The United States Environmental Protection Agency (EPA) has committed to a risk-based approach to resource management by establishing environmental regulations determined by outcomes from prioritized chemical risk assessments (EPA, 2014). The scope of chemicals subject to EPA regulation are currently limited to expert opinions of the hazard, exposure, persistence, and bioaccumulation of known pollutants (EPA, 2014). However, this approach omits unexpected chemicals, such as unknown product impurities or environmental transformation products that may still act as causative agents for observed adverse effects in exposed organisms. To remedy these gaps in understanding, the environmental research community will likely need to integrate methods of prioritizing and identifying unknown chemicals into regular monitoring efforts when attempting to define concerning exposure scenarios.

The EPA is continually developing its regulatory framework to prioritize and regulate endocrine active chemicals (EACs) through their Endocrine Disruptor Screening Program (EDSP). Concern for EACs started with Sumpter and colleague's accidental discovery of intersex fish in wastewater effluent-dominated streams of England in 1978 (Sumpter and Johnson, 2008a). Expanded research efforts in the 1990s determined steroid estrogens to be the dominant causative agents for this region, raising concern for these contaminants throughout the environmental research community (Desbrow et al., 1998; Jobling et al., 1998; Routledge and Sumpter, 1996). Since then, global research of EAC occurrence suggest that non-steroid estrogens and a suite of unidentified chemicals can also act as agents in estrogenically active environmental matrices

(Jobling et al., 1998; Kawanishi et al., 2004; Sun et al., 2008). There is also concern for the effects of these chemicals as environmental mixtures, particularly when considering that steroid estrogens with the same mode of action are known to have additive effects at levels below their individual no-observed adverse effect level (NOAEL) (Brian et al., 2007, 2005; Thorpe et al., 2003). Considering the concern for EAC effects on humans and these gaps in knowledge, the EPA developed the EDSP in 1990 as a two-tiered approach to structure weight-of-evidence risk assessments and identify EACs for further study. Assays from the initial, Tier 1 assessments of chemicals, such as blood vitellogenin (VTG) concentrations in adult male fish, are designed to assess evidence for exposure to EACs (Borgert et al., 2014; Hansen et al., 1998; Sumpter and Jobling, 1995). Subsequent Tier 2 assays provide a more comprehensive risk analyses of chemicals identified in Tier 1 to develop evidence for an EAC's mode of action. In 2014, the EPA responded to criticisms to make the prioritization more comprehensive by including chemical database information from high throughput and *in silico* bioassays conducted by other risk assessment programs. The program now prioritizes chemicals with high persistence, bioaccumulation, and toxicity (PBT) characteristics under the Endocrine Disrupting Program in the 21st Century Program (EDSP21) (Borgert et al., 2014; U.S. Environmental Protection Agency, 2017). Strategies for inclusion of unknown chemicals and considerations of chemical mixtures are only starting to be incorporated into this process (McEachran et al., 2017; Schymanski and Williams, 2017).

Liquid chromatography high-resolution mass spectrometry (LC-HRMS) is gaining attention as a potentially viable tool for the identification of trace organic contaminants (TOrCs), particularly ionogenic TOrCs, not yet identified or observable by traditional targeted environmental monitoring strategies (Krauss et al., 2010). Quadrupole time-of-flight (QToF) and orbitrap mass analyzers provide the ability to discern between co-eluting chemicals with the same nominal mass while also acquiring diagnostic fragments that may be used to reveal a chemical's identity. While sample preparation techniques and analytical separation methods can still act as sources of bias in these analyses, HRMS allows for a more complete characterization of complex environmental matrices. Deconvolution strategies, i.e. suspect and nontarget data processing workflows, for the large data sets (up to about a gigabyte per sample) generated during data acquisition are necessary and currently lack conformity in the environmental research community (Hug et al., 2014a; Nürenberg et al., 2015; Wode et al., 2015). Suspect-screening methods involve the use of prior information to limit searches to a list of likely unknowns in a given sample matrix

(Gago-Ferrero et al., 2018). This strategy is limited by the quality of the knowledge record, allowing omissions in chemical reporting or unexpected chemicals sources to bias the understanding of environmental chemical compositions (Oetjen et al., 2018). Strategic nontarget screening methods are advantageous for their ability to conduct supervised or unsupervised feature prioritization aligned with research objectives, such as spatial and temporal trends in feature intensity, pattern finding amongst precursor masses, and matrix fractionation in combination with effect-directed analysis (Al-Salhi et al., 2012; Schollée et al., 2015; Schymanski et al., 2015a). With the added time efficiency of LC-HRMS data reduction, feature identification is then achieved by comparisons to library spectra or with the aid of *in silico* fragmentation prediction software. The quality of feature matches has also been integrated into the data reduction portion of multiple studies of nontarget workflows (Al-Salhi et al., 2012; Brack et al., 2016; Du et al., 2017a).

Prioritization of EACs and their sources is desired as part of an ongoing investigation of the widespread TOrC occurrence and fish feminization throughout lakes in Minnesota (USA) (Guyader et al., 2018a; refer to Chapter 2). TOrCs encompass a diverse set of chemical structures with a range of physicochemical properties associated with both natural and anthropogenic sources (Lapworth et al., 2012). Some of these chemical structures are similar to biosynthesized estrogens, raising concern for their potential as EACs (Vajda et al., 2008). Several studies reported diffuse source detection patterns of TOrCs in surface waters and groundwater throughout Minnesota (Baker et al., 2014; Ferrey et al., 2015; Writer et al., 2010). Proposed diffuse sources include atmospheric deposition, landfill leachate, roadway runoff, and on-site wastewater treatment systems (OWTSs). Studies of TOrC discharges from centralized wastewater treatment facilities drive hypotheses that OWTSs implemented at individual lakeshore residences likely contribute to TOrC presence and fish exposure EAC mixtures (Barber et al., 2012; Conn et al., 2010b). Previous research indicated elevated TOrC concentrations and endocrine activity at locations closer to individual lakeshore residences, which presumably had active OWTSs (Guyader et al., 2018a; refer to Chpater 2). However, the diffuse nature and anticipated temporal variability of this hypothesized source make targeted and suspect screening of EACs in these mixtures very difficult. Baker and colleagues noted misalignment of chemical and toxicological data from their targeted analysis of EAC mixtures in Minnesota lakes (Baker et al., 2014). Nontarget screening of these waters could provide more robust characterization of TOrC sources and resultant effects to aquatic life in these lakes.

The objective of this study was to strategically prioritize and identify components of aqueous TOrC mixtures at sunfish spawning habitats that may contribute to endocrine disruption of resident fish. We hypothesized that unique HRMS features would occur in higher abundances in aqueous samples from sites associated with estrogenic activity (ACT) relative to sites with control-type bioassays (REF). The following research questions were addressed: 1) What nontarget HRMS features are present in water samples from locations associated with endocrine disruption in resident fish?; 2) How are the TOrC mixtures similar and/or dissimilar to locations without biological activity?; and 3) Have any of the unique features identified at biologically active sites been linked to endocrine disruption in literature studies and are they currently prioritized by EDSP21? To address these questions, porewater samples from four sunfish spawning habitats within the same lake were analyzed using LC-QToF-MS. Sites were binned according to their designations from a previous study (Guyader et al., 2018a; refer to Chapter 2) based on proximity to shoreline residences, which also corresponded to higher targeted TOrC concentrations in site porewater and elevated blood VTG concentrations in resident male sunfish. Features were prioritized according to a univariate (equal variance t-test of log2 transformed data) analysis accompanied by multivariate statistical methods (Principal Component Variable Grouping). Adduct and isotope redundancies were then eliminated prior to feature identification using spectral libraries and molecular formula search algorithms. Confidence in feature identification was reported in alignment with conventions established by Schymanksi and colleagues (Schymanski et al., 2014). The potential EAC capabilities of each identified (Schymanksi et al., 2014 designation of Tier 2 or greater) feature was then assessed through evaluation of the relevant scientific literature and online tools (EDSP21 Dashboard, https://actor.epa.gov/edsp21/).

3.3 Materials and Methods

3.3.1 Site Selection

Samples used for this work were part of a larger sampling effort previously reported (Guyader et al., 2018, schematic in Figure 3.1a). In brief, five lakes were sampled in the summer months of 2016 to evaluate the biological and chemical effects of residential OWTSs on Minnesota littoral regions. Lake porewater and male bluegill sunfish were collected from four spawning habitats within each lake: two proximal to and two distal from residences with OWTSs. Targeted analysis of grab porewater samples (33 residential wastewater indicators and known or suspected EACs) indicated increased total TOrC concentrations at locations closer to residences with

OWTSs, with notable intralake differences in concentration at Lake Mary. The insect repellent N,N-diethyltoluamide (DEET) and insecticide 2,4-dichlorophenoxyacetic acid (2,4-D), neither of which are known EACs, were two notable detections with higher concentrations at the sites proximal to shoreline residences in Lake Mary. Male bluegill sunfish captured from these locations also demonstrated blood VTG concentrations significantly ($p < 0.05$) higher than fish sampled from the distal locations. For these reasons, Lake Mary's sampling locations were determined to fit the research objectives of this study.

3.3.2 Sample Selection

Both grab samples and temporally integrated samples from the four Lake Mary sampling locations were included in the present study. Grab sample collection was described previously (May and June of 2016) (Guyader et al., 2018a; refer to Chapter 2). Temporally integrated samples were collected in a second round of sampling during the months of July and August 2016. A mini-piezometer set up was used to pump porewater at ~35 mL/min collected over a 12-hour period into a series of 1 L baked methanol rinsed amber glass bottles (~ 25 total/ sampling location). Filled bottles were stored on ice during the duration of the sampling period. After each 12-hour sampling period, the 1L amber glass bottles were transported back to St. Cloud State University's Aquatic Toxicology Laboratory (St. Cloud, MN, U.S.A.) on ice and then combined in a deionized water rinsed churn which combined the sample into a temporally integrated sample mixture. Triplicate 1L samples of the temporally integrated samples were then collected in 1L baked amber glass bottles rinsed with methanol and shipped overnight on ice to the Colorado School of Mines (Golden, CO, U.S.)

3.3.3 Sample Analysis

Aqueous sample preparation methods were previously described (Guyader et al., 2018a, refer to Chapter 2). Briefly, within 48 hours of receiving samples both grab and temporally integrated 1L samples were filtered (Whatman GC-F filters, Maidston, U.K., spiked with 20 ng each stable isotope standards (same as those listed in Table A.1; Sigma-Aldrich, St. Louis, MO, U.S.) then enriched using solid phase extraction (SPE; Oasis HLB cartridges, Waters, Milford, MA, U.S.; eluting solvents through Thermo Fisher Scientific, Waltham, MA, U.S.). Extracts (2 mL in methanol) were stored at 4°C and until analysis.

Preceding analysis, sample extracts were diluted ten-fold in ultra-pure water (Optima Water, Thermo Fisher Scientific, Waltham, MA, U.S.). Information dependent acquisition (IDA) experiments with dynamic exclusion were conducted on a Sciex TripleTOF 5600 (Sciex, Framingham, MA, U.S.) with two separate 500 µL sample injections: one using positive (ESI +) and one using negative electrospray ionization (ESI -). Samples acquired in ESI + mode were separated using a C18 for hydrophobic compounds column (Luna, 3um C18(2) 100A 50x 4.6mm, fully porous silica, Phenomenex, Torrance, CA, U.S.) with an eluent gradient (0.5 mL/min) of ultra-pure water buffered with 0.1% formic acid and ultra-pure methanol (Thermo Fisher Scientific, Waltham, MA, U.S.) buffered with 10mM ammonium formate. Samples acquired in ESI – mode were separated using a C18 for basic compound separation column (Gemini, 3um C18 11A 50x4.6mm fully porous organo-silica, Phenomenex, Torrance, CA, U.S.) with an eluent gradient (0.5 mL/min) of ultra-pure water and methanol both buffered with 5mM ammonium fluoride (Thermo Fisher Scientific, Waltham, MA, U.S.). For both methods, the LC gradient started with 10% organic phase, ramping to 95% organic phase at 7 minutes, holding at 95% organic until 12 minutes, then returning to 10% organic and equilibrating until 18 minutes. The TripleTOF was operated with a minimum intensity of 1000 cps, dynamic exclusion, and no more than 12 candidates per cycle.

3.3.4 Data Processing Workflow

Strategic data deconvolution for the prioritization and identification of features that may contribute to estrogen agonism was executed according to the workflow in Figure 3.1b. Data were analyzed in four separate analysis sets according to sample matrix and instrument ionization mode: grab samples acquired in ESI + mode (*grab pos*), grab samples acquired in ESI – mode (*grab neg*), temporally integrated samples acquired in ESI + mode (*temp int pos*), and temporally integrated samples acquired in ESI – mode (*temp int neg*). Features were defined, prioritized, and identified using Sciex PeakView v 2.2, Sciex MarkerView v 1.3.1, Sciex MasterView v 1.1, and RStudio v 1.0.136 supported by R v 3.4.4.

Peak picking, feature alignment, and initial feature prioritization across all environmental samples within a dataset was conducted in MarkerView. Data handling considerations including the inclusion of blanks in the peak picking and alignment process are further described in the Quality Control and Quality Assurance section of the Materials and Methods. Algorithm parameters for the Enhance (Sciex, Framingham, MA, U.S) peak picking and feature alignment algorithm were informed by total ion chromatogram and extracted ion chromatogram reviewed in PeakView. Peaks were picked for all datasets between 2.5 minutes and 12 minutes of the LC gradient, with a minimum LC peak width of 12 seconds, a subtraction offset of 15 scans, a subtraction multiplication factor of 1.2, a noise threshold of 1000 counts per second (cps), a minimum spectral peak width of 5 ppm, and minimum RT peak width of 5 scans. For ESI + gran and temporally integrated datasets, a retention time tolerance of 0.1 minutes and mass tolerance of 5 ppm was used to align features across samples. For ESI- datasets, 0.2 min and 10 ppm were used for the feature alignment tolerance. The algorithm was set to exclude features identified as naturally occurring C13 isotopes and spiked isotope standards. Features intensities were first transformed using most-likely-ratio (MLR) normalization (Lambert et al., 2013) to create a matrix of features responses (normalized intensities across field replicates). MLR normalization relies on a tiered normalization technique by first comparing the quality of shared features in samples more similar to each other, then comparing features in samples of decreasing similarity to each other before creating a final scale factor that considers most likely signal trends across all samples in a set. This method of normalization was preferred over spiked isotope standard normalization to prevent biasing of data to targeted chemical responses, particularly the spiked isotope standards.

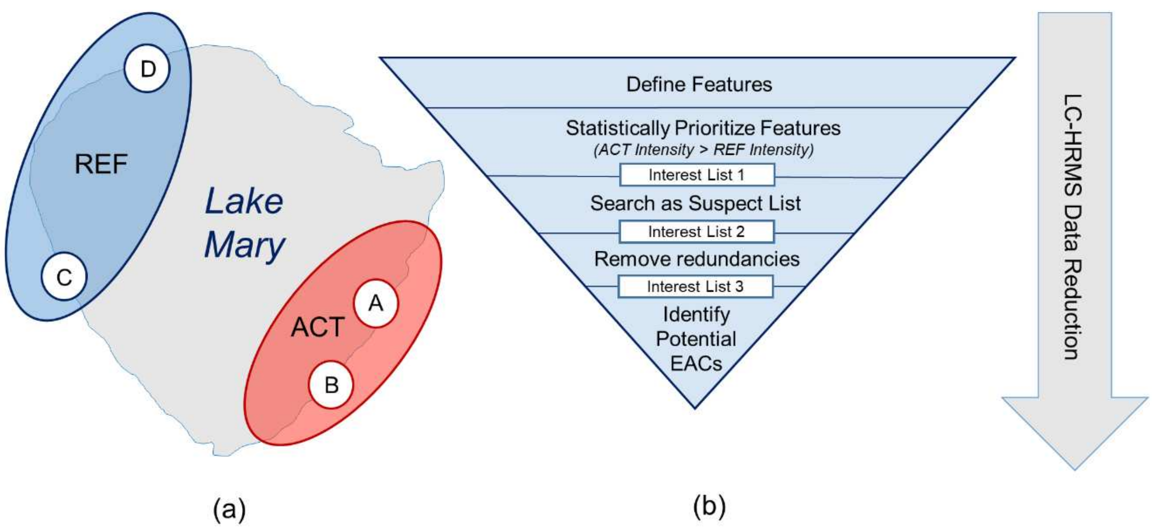

Figure 3.1 a) A schematic of sampling locations used to execute b) the workflow to identify potentially relevant biologically active chemicals. Univariate and multivariate analyses statistically prioritized liquid chromatography tandem high resolution mass spectrometry (LC-HRMS) features with higher intensities at endocrine active (ACT) versus reference (REF) sites.

Prior to statistical prioritization, peak areas were transformed so that all feature responses in a sample had a Gaussian distribution. After exporting the MLR normalized data matrix from MarkerView, responses were $\log_2$ transformed to achieve Gaussian distribution (zero response bin excluded). MarkerView automatically sets non-detect features within a sample at a response of zero. This substitution is known to skew statistical analyses in small data sets with laboratory-defined detection limits (Helsel, 2012). However, handling of censored data in LC-HRMS "big data" sets according to these recommended methods is still unprecedented, largely because there is no reliable way for unbiased prediction of detection limits for unknown unknowns in a broad compound class such as TOrCs. Log transformed matrices were then imported back into MarkerView and the statistical prioritization proceeded.

While several approaches were explored, the final method of feature prioritization relied on a t-test followed by principal component analysis (PCA) with the loadings plot deconvoluted using principal component variable grouping (PCVG). The t-test was executed in MarkerView assuming equal variance to identify features with significantly (degrees of freedom; df = 11, t-value > t-critical, $p < 0.05$, log-fold change > 0.1) higher responses in ACT versus REF group samples within each dataset separately (visualization of t-test results in Figure B.2. The df value chosen was determined by considering each field replicate as a unique sample, rather than each site as a unique sample. This preference allowed the possibility of more false positives to be retained at this point of the workflow, which was preferred considering the low intensities anticipated in the lake porewater matrix with likely influence from a diluted mixture of one or more diffuse sources and likely matrix suppression from dissolved organic matter. Features that did not meet this criterion were eliminated from the workflow and the remaining features were used to generate PCA scores and loadings plots. Similarly trending features were then grouped using a PCVG algorithm to designate them as either ubiquitous at all ACT sites (AB), or unique to individual ACT sites (A or B) (Ivosev et al., 2008). PCVG input parameters were as follows: using peak information from the first four principle components, an angle delta of 40 degrees, a minimum distance from the origin of 0.02, automatic group assignment according to magnitude, and only starting a new group at a minimum distance of 0.05 from the origin. An alternative approach of reducing features by an iterative PCVG analysis was also attempted, but found to be a time-intensive process that led to very similar results of the t-test followed by PCVG (additional details on this alternate approach are provided in the SI). Briefly, final normalized feature matrices were used to construct PCA scores and loadings plots. PCVG groups attributed to spurious trends were eliminated and PCA scores and loadings plots were regenerated. This process continued until loading plot groups could no longer be eliminated and a t-test of the remaining features was conducted to eliminate any features remaining which did not meet univariate limits of significance. Final prioritized feature lists for each dataset were compiled into a suspect search list for further investigation called “Interest List 1”.

The Interest List 1 feature list was further reduced after examination in MasterView. Input parameters for the peak-picking, library search, and formula-finding algorithms in MasterView are displayed in Figure B.1. The same peak picking algorithm is integrated in both software; however, the input parameters are slightly different; i.e. relying on a minimum number of counts and noise signal in MarkerView and a minimum feature intensity in MasterView. Selection of a control sample for blank filtering was also enabled in MasterView. The field blank corresponding to each dataset was designated as the control sample to remove any features from Interest List 1 whose signal did not exceed five times the raw intensities in the control sample. While features associated with blank contamination are unlikely to carry over to this stage of the workflow, this measure was included to verify the quality of features retained beyond this point. Features which met quality standards with consistent occurrence across all three field replicates for a given sites were retained and compiled into “Interest List 2” that noted their verified site designation (A,B, or AB) and average intensity for redundancy deconvolution in RStudio (Rstudio Team, 2015).

Isotope and adduct redundancies in the Interest List 2 feature list were identified using *nontarget* (Loos, 2016) and associated packages processed using R within the RStudio environment. Separate peak lists of feature mass, intensity and retention time for each field site, sample matrix, and acquisition mode were used as source files. For *grab po*s and *temp int pos* data sets, M+H, M+NH4, M+Na, and M+K adducts were identified using the adducts.search() function with a retention time tolerance of 0.2 minutes and an m/z tolerance of 5 ppm. For the M-H, M+Na-2H, and M-H20-H adducts in the *grab neg* and *temp int neg* data sets the same function and tolerance parameters were used. The *nontarget* algorithm identifies adduct pairs using the exact mass defect associated with each adduct relative to a chemical’s monoisotopic mass (M). All adducts except the M+H or M-H adducts were removed from the final report list of features. Non-C13 isotopes missed by MarkerView’s initial peak picking algorithm, specifically N15, O17, Cl37 and, Br81, were also identified using the pattern.search() function of *nontarget* using the same retention time tolerance parameters as the adduct search. All identified isotopes were eliminated from final report list. After verifying alignment of isotopes and adducts shared by A and B site peak lists, the final report list for each data set was compiled from the retained features at each site into a suspect search list, “Interest List 3”, for identification in MasterView.

Within MasterView, features in Interest List 3 were identified by conducting spectral library and formula searches using molecular weight ion, isotope intensity ratio, and MSMS spectral data. The libraries searched included the 2017 National Institute of Standards and Technology (NIST) library, the Sciex All-in-One library, and several custom libraries generated at the Colorado School of Mines (Golden, CO, U.S.). The libraries encompass a total of 164,204 library entries (redundancies likely) that encompass a broad suite of TOrC and other contaminants of emerging concern, including antibiotics, natural products such as myctoxins and triterpenoids and flavinoids, forensic associated chemicals such as human synthesized chemicals including metabolites of illicit drugs, and per- and polyfluorinated alkyl substances (PFASs). The library match function in MasterView assigned a library match percentage value assessing the similarity of acquired data to a library entry (Colby et al., 2017). The library search assumes fragments carry a single positive charge. The calculated library match percentage, along with manual inspection of feature data, was used to determine the quality of library matches. The quality of feature identifications were reported according to the scale established in Schymanski et al. (2014) summarized as relevant to this study in Figure B.3 (Schymanski et al., 2014a). Features with confirmed instrument-specific chromatography and diagnostic spectra using standards were reported as Level 1 IDs. Features with a 70% or greater library spectral match, verified by manual inspection to have a suitable number of diagnostic fragments (greater than 2), were reported as Level 2 IDs. Features with lower library spectral match percentages were manually inspected. If these features matched more than two but not all diagnostic fragments in the library spectra, then they were assigned a Level 3 ID designation. Features with a 70% or greater FormulaFinder match, which is only calculated using isotope MS data and, therefore, does not require MS/MS spectra, were reported as Level 4 IDs. Identification of features with poor or absent library spectral matches were assigned a Level 5 IDs, including those without acquired MSMS spectra with no or <70% FormulaFinder matches. Characteristic fragmentation for Level 1, 2, 3 and 5 IDs is provided in the Results and Discussion section. Fragmentation prediction and formula prediction algorithms were not executed at this time.

3.3.5 Investigating Estrogen Disrupting Potential of Identified Features

An EDSP21 Dashboard search was conducted using the Chemical Abstracts Services number (CAS #) of the Level 1, 2, and 3 IDs (see Table 3.1 for Levels 1 and 2). Identified features without available CAS numbers were searched for by their chemical name. The tool aggregates data from high throughput *in vivo* assays*, in vitro* exposure studies*,* and *in silico* predictions to evaluate the endocrine disrupting potential of over 1,800 chemicals. Since the bioassay integrated in this study is considered a strong indicator of exposure to estrogen agonist mixtures, interpretation of EDSP21 data focused on reported estrogenic potency.

3.3.6 Prioritized Feature Occurrence in Reference Lake

Interest List 3 features were also used to execute a suspect search of these potential EACs at a reference lake, Cedar Lake. Cedar Lake was chosen as a reference lake because bioassays from sunfish captured at all sampling locations had average blood vitellegonin concentrations similar to the REF sites of Lake Mary (Guyader et al., 2018a; refer to Chpater 2). This task was intended to make the distinction of prioritized features of chemicals potentially associated with endocrine activity more robust. We hypothesized that Interest List 3 features present at any sites at the reference lake, particularly sites proximal to shoreline residences, were more likely associated with anthropogenic diffuse source inputs than toxicological effects observed in resident fish. Grab and temporally integrated samples were collected during the same field sampling campaigns as the samples collected at Lake Mary. Cedar Lake samples were analyzed during the same analytical runs as the Lake Mary samples. The suspect searches for each corresponding set of samples, i.e. sample matrix and instrument acquisition mode, was executed using the appropriate Interest List 3 in MasterView. Algorithm search parameters were the same as those used for Lake Mary samples (Figure B.1). Retained features for the Cedar Lake report list were required to meet the same QA/QC requirements as final report features from Lake Mary, i.e. presence in all replicates and same confidence in their identity as Lake Mary.

3.3.7 Quality Control

The novelty of nontarget analysis, particularly its current lack of implementation by regulatory agencies, has prevented the establishment of streamlined quality control and quality assurance procedures. This study used several common targeted and LC-HRMS QA/QC procedures documented in the literature. The sample preparation methods for grab samples were deemed suitable by a surrogate recovery analysis executed using targeted liquid chromatography tandem triple quadrupole mass spectrometry (results for grab samples reported in Guyader et al 2018). Temporally integrated samples showed less suitable performance for feature quantitation of the broad suite of spiked standards; however, feature identification was still possible. This highlights the difference in QA/QC needs for nontarget feature prioritization versus targeted quantitation-confirmed identity. Mass resolving power of the instrument for both precursor and product ions was confirmed regularly throughout instrument runs. Briefly, a mass calibration solution was introduced to the source every four samples and measured ions were required to fall within 5 ppm error of theoretical values. Mass accuracy and instrument sensitivity was also confirmed in post processing by conducting suspect searches of spiked isotope standards in samples and standards regularly injected throughout sample runs. In addition to validating instrument acquisition performance, the quality of the peak picking algorithm and selected input parameters was evaluated through the data alignment feature of MarkerView. Each sample set was first evaluated with the appropriate field blank (according to corresponding sample matrix and instrument acquisition ionization mode) samples considered in the peak picking and alignment process. Provided that the blank samples suitably clustered together on the Pareto scaled principal component analysis (PCA) diagram, then the workflow would proceed with environmental samples only. MasterView examination of data required that all replicates at a site contained similar TOF scans for a feature to be included in the final report list. Regarding identification efforts, information dependent acquisition (IDA) can occasionally fail to trigger an MS/MS for precursor ions that meet intensity requirements. As long as a feature could be identified according to the MasterView peak picking algorithm in the TOF scan data of all field replicates and one of the replicates in a sample triggered an MS/MS spectrum, then that was designated as the diagnostic spectra for that feature. Sample replicates with multiple MS/MS triggers were also verified for diagnostic spectra alignment.

3.4 Results and Discussion

3.4.1 Peak Picking and Feature Alignment

Evaluation of total ion chromatograms (TICs) and extracted ion chromatograms (XICs) informed input parameters for the peak picking algorithm. Ion (m/z) tolerances are typical relative to previous studies (Schollée et al., 2015; Schymanski et al., 2015b); however, the retention time (RT) tolerance was set smaller than these studies which aligned with the preference for false positives as opposed to false negatives at this stage of the workflow. TICs from the *grab pos* dataset indicated chromatographic instability after 12 minutes, corresponding to the end of the run's organic hold. For this reason, a maximum retention time of 12 min was imposed on the *grab pos* and *temp int pos* data sets. Spiked isotope standards were verified to not elute after this 12 minute maximum retention time. The resultant number of total defined features for *grab pos*, *grab neg*, *temp int pos*, and *temp int neg* datasets is discussed in the Interest List 1 Data Reduction section.

3.4.2 Statistical Prioritization of LC-HRMS

Results from the statistical prioritization component of the workflow are exemplified using the *grab pos* dataset in Figure 3.2. After conducting the t-test of log transformed responses, PCA scores (Figure 3.2a) and loadings (Figure 3.2b) plots visualized multivariate trends that differentiated samples from one another. Interestingly, while separation between REF and ACT site samples was expected in the scores plot (Figure 3.2a, separation on PC1 which represented 48.1% of variance in the t-test pre-treated data), REF sites clustered closer to one another along PC2 (25.1% of variance in the t-test pre-treated data) space relative to the ACT sites. We interpret this serves as a visualization of unique TOrC discharges from the individual households most proximal to ACT sampling locations with both shared (occupying similar space relative to PC1) and unique (separation in space relative to PC2) feature compositions. These unique clustering patterns for intralake samples with low intensity features is encouraging for future studies to implement non-target sample acquisition and strategic data prioritization with higher spatial resolution sampling for better understanding diffuse sources of TOrCs. Environmental studies implementing PCA analysis for data interpretation usually detected the greatest scores plot separation between samples acquired from radically different sampling locations (Schreiber et al., 2010). Our results suggest that, provided more spatial and temporal resolution in environmental samples, nontarget screening has great potential in understanding the fate and transport of TOrCs with non-point sources. In addition, integration of spatial and temporal trends into LC-HRMS prioritization workflows could allow for more robust analysis of toxicological relevance by allowing for assessment of feature persistence fundamental to the PBT model of chemical risk assessment.

From a toxicological perspective, the scores plot clustering patterns of the individual ACT sites could suggest that unique TOrC mixtures can lead to the same biological endpoints (separation of ACT sites along PC2) or that certain TOrC components are critical to the endocrine disrupting potential of exposure mixtures (separation of ACT from REF sites along PC1). In the former case, there are "excess" prioritized compounds which may be interpreted as co-contaminants and/or non-estrogenic indicator compounds for EACs. This interpretation relies on the assumption that the elevated presence of certain TOrCs at active sites, rather than the absence of certain TOrCs at ACT sites that are present at REF sites, are responsible for biological observations. Our assumption is likely valid for application to an assay indicative of estrogen agonism based on previous observations of additive estrogenic effects of steroid estrogens in TOrC mixtures (Hernández et al., 2013). LC-HRMS features not prioritized by this analysis may still be associated with EAC properties (REF sites did not necessarily exhibit a complete absence of biological endpoints indicating endocrine disruption), but according to this workflow, they are not crucial to additive or synergistic effects that lead to pronounced differences in estrogen agonism between this study's sampling locations.

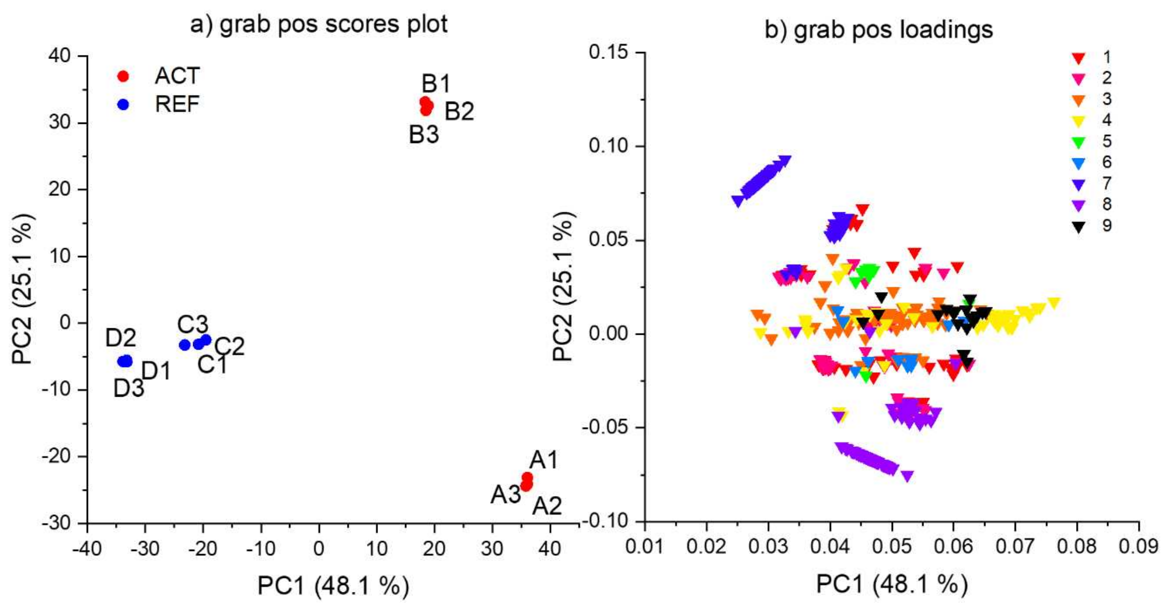

Figure 3.2 a) Scores and b) loadings plots from principal component analysis of t-test pre-treated data for the grab pos dataset. Sample site groupings are labeled in the scores plot as either endocrine active (ACT) or reference (REF). Individual features and their associated PCVG group (color coded 1-9) are displayed in the loadings plot. Refer to statistical prioritization of LC-HRMS section for further explanation.

The PCVG loadings plot indicates which features trend similarly with one another (Figure 3.2b). Each point within the plot represents a feature color coded according to its assigned group by the PCVG algorithm, PCVG groups 1 through 9. The benefit of PCA is the identification of features responsible for the separations observed in the scores plot without yet knowing these features' identities. The success of the PCVG algorithm is demonstrated in Figure 3.3, which provides visualization of the way in which PCVG groups corresponded to variable identification as occurring in both (AB) or in individual ACT sites (A or B). MarkerView provides a "view plot profile" feature which generates plots of response trends for individual features across all samples in a dataset. Figure 3.3 demonstrates how the plot profiles of features within the assigned PCVG groups allowed for the preliminary designation of features as A, B, or AB. The plot shows the average sum of all MLR normalized and log-transformed features within a given PCVG group across site replicates in the grab pos dataset (as an example). As expected from the PCA loadings plot, most groups are associated with AB features, while features unique to site A or site B exclusively were designated as PCVG group 8 or 7, respectively. The number of features classified within each PCVG group is also displayed in Figure 3.3, with AB groups containing the most total features, then A, then B for the grab pos data set. It is important to note that some groups, such as PCVG 3, indicate a more borderline classification of features with elevated presence at REF sites. Strategies for handling these features are discussed further in the text, particularly how questionable features are identified and final reported site distinctions are made.

3.4.3 Feature Reduction and Elimination of Redundancies

The overall workflow reduced the number of features in each dataset by an average of 97.9%: an average of 89.9% reduction between the initial feature definition and statistical prioritization that generated Interest List 1, and an average of a further 78.4% reduction during the suspect search that narrowed Interest List 1 to Interest List 2. Finally, the workflow resulted in an average reduction of 7.3% by removing redundancies between Interest List 2 and Interest List 3. Considering the amount of time necessary to run computer algorithms for nontarget environmental matrix screening using an entire spectral library, this strategic workflow considerably reduces the amount of data processing time of LC-HRMS data. Prioritizing features by environmental relevance prior to feature identification makes this technology more accessible for integration into research efforts and facilitates its application within regulatory frameworks.

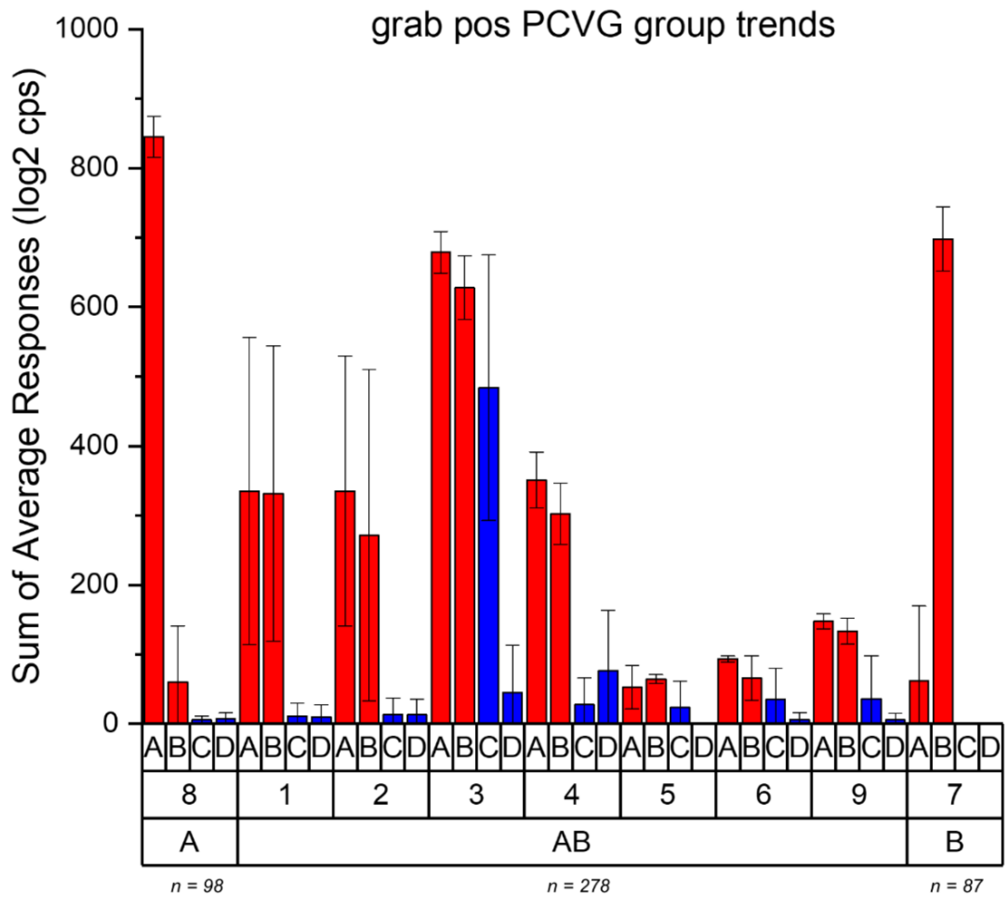

Figure 3.3 Sum of average feature responses in each principal component variable group for the A, AB, and B group classifications. Standard deviation represents the sum of standard deviation in feature response across site replicates. Note non-detects of features at site D in PCVG group 8 and sites C and D for group 7.

Interest List 1 features were further reduced in quantity after examination as a suspect list. The MasterView peak picking algorithm was able to identify certain Interest List 1 features with poor peak qualities, specifically features that were elevated noise baselines with too wide a RT width and features lacking consistent presence in site field replicates. Features that met MasterView algorithm parameters were compiled into Interest List 2. This list was then subjected to an adduct and isotope pattern search in R. Adducts were more commonly found in samples acquired in ESI + mode (M+H, M+NH4, M+Na, and M+K), with adducts from ESI – samples identified only in the *temp int site B* list of features (M-H, M+Na-2H, and M-H20-H). The isotope pattern search only identified Cl and Br isotopes in the ESI – features lists, with no isotopes identified for the ESI + feature lists. The *nontarget* package includes other adducts and isotopes to search for; however, the ones chosen were most appropriate for the sample matrix and instrument analysis procedures used in this study. Examination of the final (Interest List 3) feature lists indicates several cases of potential remaining redundant entries with the same retention time and similar masses. These listed features do not meet known isotope or adduct mass defects and could instead be in source fragments and other such artifacts introduced by instrument conditions. This was confirmed by reviewing the identity of two features, 2,4-D and dichlorophenol with the same retention time. Unfortunately, there is currently no appropriate method of establishing these instrument artifacts by feature mass, intensity, and RT alone.

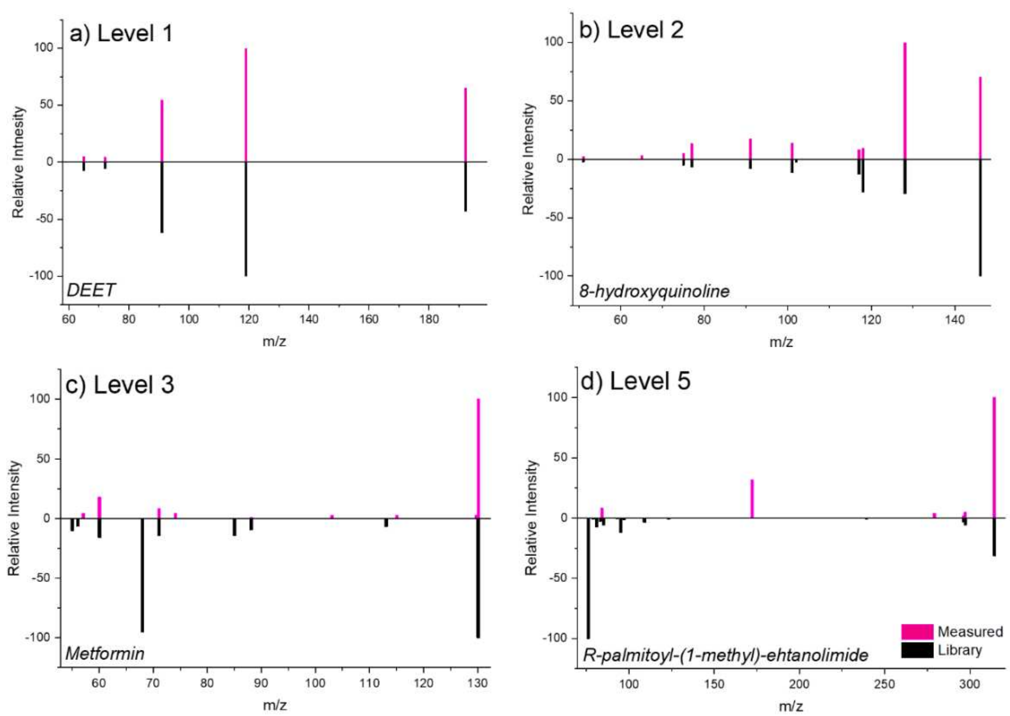

Figure 3.4 Representative comparisons between acquired and library reference MSMS spectra corresponding to assigned confidence levels of identification a) Level 1, b) Level 2, c) Level 3, and d) Level 5. Library reference spectra and acquired MS/MS spectra are displayed as negative and positive relative intensities, respectively.

Identification of features was made possible through library and formula finder searches. Representative examples of MS/MS spectra associated with Level 1, 2, 3, and 5 confidence levels for reported features are displayed in Figure 3.4. DEET is a Level 1 because it not only had a strong isotope and spectra match, but was also a chromatograph match to a standard. 8-

hydroxyquinoline's isotope and MS/MS spectra also matched very well for what is likely two isomers of a hydroxyquinoline compound. The feature with a retention time of 6.9 minutes is more likely 8-hydroxyquinoline based on comparison to the available library spectra. The peak at 6.5 minutes is likely a structural isomer with the –OH group placed on another carbon in the structure. Retention time effects were also important when examining structures more likely formed as in-source fragments, for instance the Level 1 identification 2,4-dichlorophenoxyacetic acid and its transformation product dichlorophenol as previously noted by Song in 2011 of dichlorophenol agricultural pesticides (Song, 2011). DCP was removed from the reported list of features in Interest List 2 for this reason.

Level 3 IDs were more subjective, and therefore, tentative identifications of LC-HRMS features (full list in Tables B.1 thru B.4). Even with multiple examinations of data by the same or different analysts, strong conclusions could not be made as to whether or not a match was made by comparison to library spectra. Scenarios that invoked a Level 3 ID ranged from MS/MS spectra with too few diagnostic fragments (<2) which would get a strong library match according to data processing algorithms, features that missed several fragments or contained extra low intensity fragments relative to the library match, or features that had multiple strong library matches.

The full list of Level 4 IDs, feature identifications resolved by the formula search feature in MasterView, is reported in Tables B.5 thru B.8. Several of these features did not have acquired MSMS, but the isotope spectra patterns provided conclusive evidence of likely elemental compositions. This search algorithm is separate from the library search algorithm. Other formula search algorithms and workflows are available to process LC-HRMS data, such as the Fiehn Laboratories' Seven Golden rules software (Kind and Fiehn, 2007). The Sciex software's ForumlaFinder function operates from the same principles and is already integrated into the MasterView software; therefore, only this method was pursued. Further consideration of elemental composition was thought to not contribute to a better understanding of the potential endocrine disrupting effects of prioritized chemicals and was deemed to be outside the scope of this study.

Level 5 IDs consisted of three classes of features unresolved by the library search: either because no MS/MS was acquired, the library search function could not find a suitable spectra match, or the library provided a very poor spectral match. A list of all Level 5 masses of interest can be referenced in Table B.9 thru B.12. Figure 3.4d provides an example of a poor library spectral match for mass of interest, 314.2/7.5, which the algorithm identified as a 34.8% match to

R-palmitoyl (1-methyl) ethanolimide. The identity of these chemicals remains unknown but could be improved by comparison to *in silico* fragmentation prediction software such as MetFrag (Wolf et al., 2010) or the Formula Finder and ChemSpider plug-ins available in Sciex's MasterView. This workflow's retention of unidentified features differentiates it from similar LC-HRMS studies focused in nontarget characterization of environmental matrices (Blum et al., 2017; Du et al., 2017b). Considering that soft ionization libraries are limited in their content because of inconsistencies in collision energy, limited standard availability, and struggles for replicability across instruments, approaches that eliminate features with poor spectral matches from their report lists unnecessarily censor data that could be important for research objectives and eventual regulatory measures/applications. Evaluation of feature raw intensities in MasterView showed slight differences in feature intensity trends than those generated in MarkerView by MLR normalization, particularly when considering significant differences between ACT and REF features. This called into question the prioritization of several features, particularly the log-fold change criterion, suggesting that future work may be made more efficient by setting a higher log-fold change cutoff for the t-test.

The reliability of library search algorithm cut-off parameters was brought into question by several instances of attempting feature identification. Specifically, we noted more favorable scores assigned to fragments with only a single diagnostic fragment, and favorability for matches for fragments with high relative intensity but poor matches for peaks with low relative intensity in the acquired MS/MS spectra. At this time, the library matching algorithm does not consider similarity of collision energies between library and acquired MS/MS spectra. We advise incorporating this, as well as the number of matching diagnostic fragments, into future iterations of library matching algorithms. It is important to note that examination of the isotope spectra of several features indicated interference during acquisition. We suspect the presence of these interferences could affect the quality of the MS/MS spectra matches to library spectra by adding extra diagnostic fragments or suppressing the signal of diagnostic fragments. Data independent acquisition is recommended by Roemmelt and colleagues to prevent important precursor ions from being missed, particularly during toxicological investigations (Roemmelt et al., 2014); however, deconvolution of MS/MS spectra necessary for identification remains time intensive (Arnhard et al., 2015). With advancements in data deconvolution strategies for data independent acquisition methods, we recommend toxicological studies shift to usage of this data acquisition method for feature prioritization in future work.

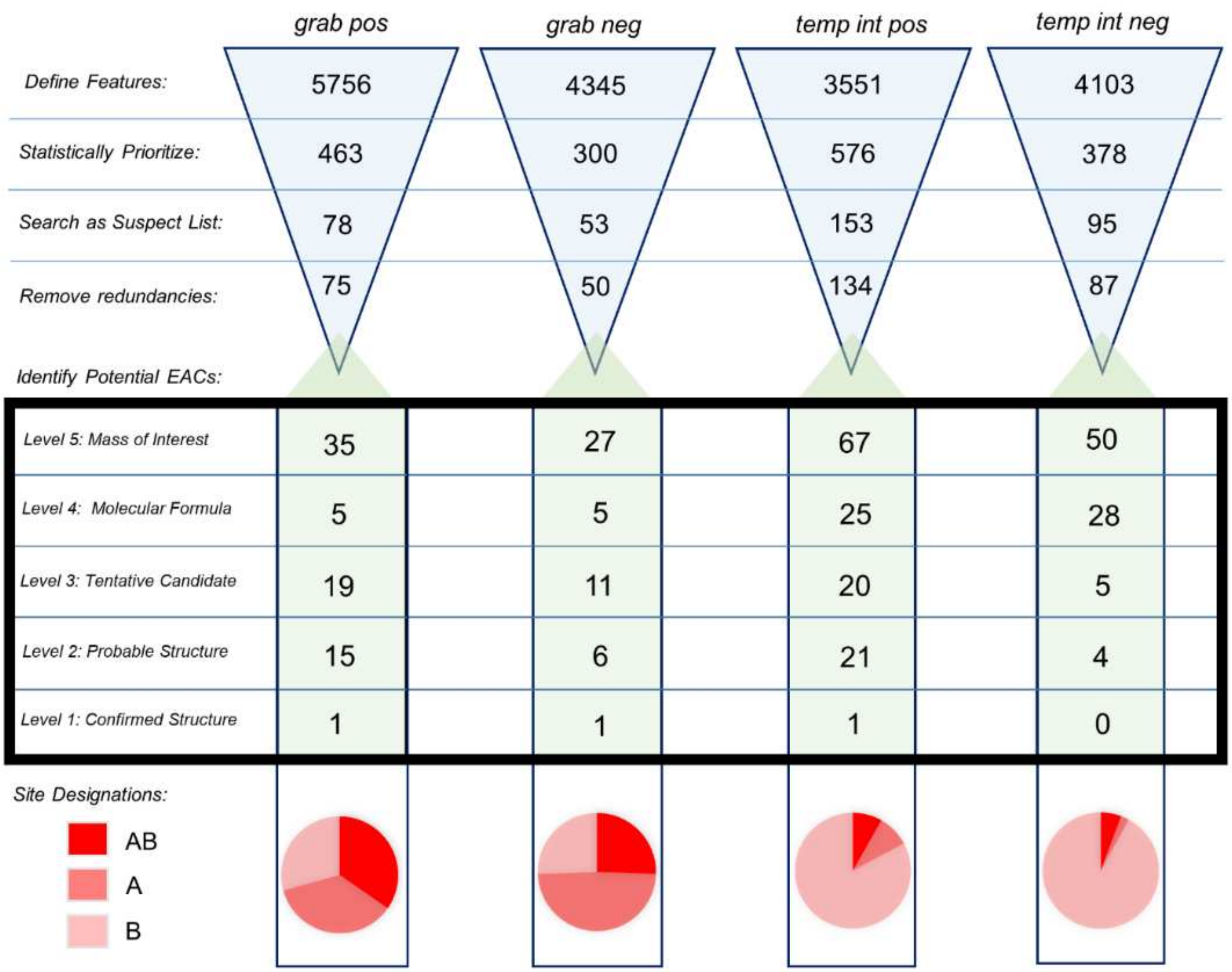

Figure 3.5 Feature reduction throughout each stage of the workflow and Schymanksi confidence level of each feature identified. See Data Processing Workflow subsection of the Materials and Methods section for explanation of data reduction process. Pie charts represent percentage of features within each dataset in A, B, or AB classifications.

Grab samples are traditionally used in field studies to provide an empirical understanding of transient environmental concentrations. The use of grab samples to study environmental TOrC concentrations is fairly limited because of the spatial and temporal heterogeneity of these contaminants' occurrence (Lapworth et al., 2012; Teerlink et al., 2012). In addition, considering endocrine disrupting effects occur with chronic exposure to EACs, it is understandable that previous efforts to align chemical occurrence in aqueous grab samples with observations of endocrine disruption in resident fish could not identify particular causative agents (Baker et al., 2014; Guyader et al., 2018a (Chpater 2)). Temporal integration of samples via 24-hour flow composites, as well as passive samplers, are alternative field sampling approaches thought to provide understanding of persistent and/or accumulative TOrCs within an environmental matrix. While the temporally integrated sample design used in this study may not have captured time scales necessary for chronic biological effects, such as endocrine disruption, they are useful for providing a look at continually occurring contaminants within each matrix. Differences between prioritized and identified features from grab and temporally integrated samples could arise from at least two sources in the workflow 1) more similar 12 hour compositions of TOrCs at each site or 2) more accumulation of matrix interferences in temporally integrated samples which inhibit LC-HRMS sensitivity for features prioritized in grab samples. The latter is particularly compelling when considering that humic substances are capable of detection in both ESI + and ESI – acquisition modes (Persson et al., 2000; Piccolo and Spiteller, 2003). Further investigation is required to examine how to interpret these different sample matrices in field studies of endocrine disruption and LC-HRMS analysis of field samples.

3.4.4 Evaluating EAC Potential

A complete list of features identified at the Level 1 or 2 confidence levels are reported in Tables 3.1-3.4. Library matches provided both confident and tentative matches for insect repellents, quinolones, pesticides and their transformation products, steroid hormones, fatty acids, food additives, industrial use chemicals, plant based derivatives, personal care products, flame retardants, and pharmaceuticals. Notable detections from targeted analysis of grab samples, 2,4-D for the *grab neg* data set and DEET for the *grab pos* data set, were retained throughout the workflow and are reported in Interest List 2 (Guyader et al., 2018a; refer to Chapter 2). It is useful to note that this alignment between analysis methods provides some confidence in the executed LC-HRMS feature prioritization workflow. Features at significantly higher intensities at both ACT sites relative to

REF sites that were previously omitted by targeted analysis include 8-hydroxyquinoline, simazine-2-hydroxy, hydroxycoumarins, mecoprop-p, malonic acid, and sumaresinolic acid. Prioritized features unique to ACT site A include plant essential oils, pesticides, and steroids. ACT site B uniquely identified chemicals include fatty acids and personal care products. The fatty acids identified are suspected to originate from tomatoes or from mechanically deboned meats (Kim et al., 2011; Püssa et al., 2009), considering that natural derived fatty acids would likely not be prioritized as different between ACT and REF sites. Although, naturally derived fatty acids cannot be completely ruled out as the source of these chemical components. Plant-derived A, B, and AB chemicals detected by this method could either come from shoreline foliage or from the anthropogenic uses of essential oils as naturopathic substances. The associated usage of many of these chemicals also suggest anthropogenic input, with some detections (i.e. human metabolites of volatile organic compounds, such phenylglyloxic), more likely to originate from wastewater streams while others (i.e. pesticide transformation products) more likely originating from outdoor residential usage. Considering the recreational fishing and boating at this lake, it is still possible for human metabolites of VOCs to originate from boating related activities.

Approximately half of the Level 1 and 2 IDs were not listed in the EDSP21 dashboard. Of the chemicals with EDSP21 dashboard listings, only two were noted to have some endocrine disrupting traits. Specifically, 8-hydroxyquinoline had a weak AR antagonist ranking and tris-(2-butoxyethyl) phosphate was listed as having active binding potency at the estrogen agonist receptor. Discrepancies between literature evaluations and EDSP21 dashboard listings of EAC potency were noticed, particularly for atrazine detected at Site A. Hayes and colleagues have noted that atrazine meets all nine Hill criteria for cause-effect endocrine disruption (Hayes et al., 2011). The EDSP21 program affirms that *in vitro* high throughput and computational model alternatives provide an accurate quantitative measure of specific endocrine receptor binding bioactivity and mechanisms that can serve as alternatives to the current Tier 1 assays (U.S. EPA, 2011). However, Filer and colleagues noted that the overall use of the ToxCast program for the assessment of EACs can fall subject to the Matthew principle (Filer et al., 2014). Explicitly, because ToxCast algorithms consider number of citations for potential EACs, they may include bias for potential EACs that have previously received substantial funding, rather than prioritizing chemicals that are most relevant from the perspective of toxicological assessment. To remedy this bias in the current EDSP21 work plan, we recommend the inclusion of nontarget screening strategies similar to this study.

Table 3.1 Schymanksi Level 1 (DEET only) and Level 2 (others listed) Chemical Structures Prioritized in *grab pos* dataset. Database outputs for chemicals listed in the EDSP21 Dashboard also provided. ER= estrogen receptor. AUC=area under the receiver operating characteristic curve. +/- indicates estrogen agonism and antagonism, respectively

Site	Feature (m/z_RT)	Identity	Tox Cast Model Predicted AUC (ER agonism, antagonism)	Consensus CERAPP QSAR ER Model Predictions (lit/QSAR, active assay(s), potency)
AB	146.0596_6.47	8-Hydroxyquinoline	0,0	lit, binding, weak
AB	146.0598_6.9	2-Hydroxyquinoline		
AB	160.0752_5.97	5-amino-1-naphthol		
AB	163.0753_8.75	6-methylchromone		
AB	184.1191_6.44	Simazine-2-hydroxy		
AB	192.1381_9.54	DEET	0,0	all inactive
B	233.1537_9.81	Isoalantolactone		
A	176.0703_7.28	8-methylquniolin-4-ol		
A	182.0489_6.13	6-methylthioguanine		
A	273.221_10.32	5α-androst-2-en-17-one		
A	305.2114_10.53	6β-hydroxytestosterone		
A	433.1122_7.42	Isovitexin		
A	579.1689_7.48	Vitexin-2"-rhamnoside		
B	267.1587_7.22	Sambucinol		
B	289.2161_10.2	5α-androstane-3,17-dione		
B	293.1203_10.09	Picrotoxinin		

Table 3.2 Schymanksi Level 1 (2,4-D only) and Level 2 (others listed) Chemical Structures Prioritized in *grab neg* dataset. Database outputs for chemicals listed in the EDSP21 Dashboard also provided. ER= estrogen receptor. AUC=area under the receiver operating characteristic curve. +/- indicates estrogen agonism and antagonism, respectively

Site	Feature (m/z_RT)	Identity	Tox Cast Model Predicted AUC (ER agonism, antagonism)	Consensus CERAPP QSAR ER Model Predictions (lit/QSAR, active assay(s), potency)
AB	149.0246_6.47	Phenylglyoxylic acid		
AB	213.0322_8.89	Mecoprop-P	0,0	all inactive
AB	218.9627_8.58	2,4-Dichlorophenoxyacetic acid	0,0	all inactive
A	161.0249_7.32	6-hydroxycoumarin		
A	207.0672_8.38	Ferulic Acid	0,0	QSAR, +/-/binding, very weak
A	267.0666_9.05	Formononetin		
A	577.1596_7.63	Vitexin-2"-rhamnoside		

Table 3.3 Schymanksi Level 1 (2,4-D only) and Level 2 (others listed) Chemical Structures Prioritized in *tempint pos* dataset. Database outputs for chemicals listed in the EDSP21 Dashboard also provided. ER= estrogen receptor. AUC=area under the receiver operating characteristic curve. +/- indicates estrogen agonism and antagonism, respectively

Site	Feature (m/z_RT)	Identity	Tox Cast Model Predicted AUC (ER agonism, antagonism)	Consensus CERAPP QSAR ER Model Predictions (lit/QSAR, active assay(s), potency)
AB	146.0597_6.36	8-hydroxyquinoline	0,0	lit, binding, weak
AB	163.075_8.72	4-hydroxycoumarin		
AB	184.1187_6.38	Simazine-2-hyrdoxy		
A	135.0803_6.96	Isobenzofuranone	0,0	all inactive
A	216.1009_8.99	Atrazine	0,0	all inactive
A	229.0857_10.27	2-hydroxy-4-methoxybenzophenone		
A	399.2502_10.32	tris(2-butoxyethyl) phosphate	0.0000118, 0	all inactive
B	115.076_7.36	2,5-hexanedione		
B	137.0963_8.88	Sabinene		
B	157.122_9.31	4-hydroxynonenal		
B	163.1114_9.15	6-hydrocycoumarin		
B	235.205_9.88	Valerenic Acid		
B	251.2002_10.16	Sclareolide	0,0	all inactive
B	253.2158_9.88	15-hydroxyculmorone		
B	277.2156_9.14	9,12-octadecadiyonic acid		
B	281.2463_9.75	Brefeldin A		
B	295.2262_10.35	9-oxo-10E,12E-octadecadienoic acid		
B	297.2413_10.35	9(10)-epoxy-12Z-octadecenoic acid		
B	295.2263_9.14	9-oxo-10E,12E-octadecadienoic acid_isomer		
B	297.207_10.02	12(13)-epoxy-9Z-octadecanoic acid		
B	311.221_10.14	9-oxo-11-(3-pentyl-2-oxiranyl)-10E-undecenoic acid		
B	315.2511_10	9,10-dihydroxy-12Z-octadecenoic acid		

Further literature reviews of analyte toxicity indicated mixed understandings of identified feature's anticipated endocrine activity and general toxicity, both individually and in mixtures. Many chemicals were noted to have more effects to androgen receptor binding. While androgens are precursors for estrogens after interactions with aromatase, the current understanding indicates that these chemicals should not affect male blood VTG concentrations (Sumpter and Johnson, 2008b). They are, however, understood to affect male reproductive development (Vajda and

Norris, 2011). 8-hydroxyquinoline is a particularly notable detection used as a metal chelating agent, fungicide in agricultural and textile operations, and antiseptic in medicinal products. Prachayasittikul and colleagues (2014) noted no endocrine activity of 8-hydroxyquinoline as a free ligand; however, they did note that copper and lithium complexes of the chemical to inhibit aromatase activity (Prachayasittikul et al., 2014). Literature classifications of EACs remains controversial and their regulatory criteria are still in development (Solecki et al., 2017; Sumpter and Johnson, 2008b). Recently, mode-of-action focused research has been considered important, particularly when attempting to predict additive or synergistic effects of mixtures, such as steroid estrogens (Brian et al., 2005). Kim and colleagues (2005) noted unexpected contributions of 2,4-D and DCP in disrupting androgenic activity of 17α-methyltestosterone (Kim et al., 2005). Diversifying the endocrine assays and number of samples used to prioritize chemicals could make these results more meaningful. Based on a literature review of the Level 1 and Level 2 contaminants, it appears that endocrine disrupting capabilities are not unique to AB prioritized features. This suggests that the presence of unique chemical substances at each site contribute to observations of endocrine activity.

Table 3.4 Schymanksi Level 1 (2,4-D only) and Level 2 (others listed) Chemical Structures Prioritized in *tempint neg* dataset. Database outputs for chemicals listed in the EDSP21 Dashboard also provided. ER= estrogen receptor. AUC=area under the receiver operating characteristic curve. +/- indicates estrogen agonism and antagonism, respectively

Site	Feature (m/z_RT)	Identity	Tox Cast Model Predicted AUC (ER agonism, antagonism)	Consensus CERAPP QSAR ER Model Predictions (lit/QSAR, active assay(s), potency)
AB	103.0045_2.62	Malonic Acid	0,0	all inactive
A	149.0244_6.86	Phenylglyoxylic acid		
B	159.1028_8.02	3-hydroxyoctanoic acid		
B	309.2073_9.77	9-octo-11-(3-pently-2-oxiranyl)-10E-undecenoic acid		

3.4.5 Suspect Search of a Reference Lake

After conducting a suspect search of the Interest List 3, 41 features across the *grab neg, grab pos, and temp int pos* datasets were also detected at the reference lake (list in Table B.13). None of these features were consistently detected across all sampling locations from the Cedar Lake sites. Features were more frequently detected at sites more proximal to shoreline residences (sites B and D), suggesting that they are more likely associated with residential proximity and are sourced from anthropogenic activity at residences. The only high confidence (Level 1 and 2) features identified were DEET and the glucoside derivative, indoxyl. Based on the method in which site designations were assigned in this investigation, these features also present at the reference lake may not be critical in the endocrine activity of the TOrC mixtures present at the endocrine active sites in Lake Mary.

3.5 Implications and Conclusions

This study aimed to develop a means of prioritizing and identifying potential EACs within advanced LC-HRMS analysis of environmental matrices. In isolation, these data do not provide a scientific basis to conclude that the identified chemicals have the potential for endocrine disruption. Instead, we view the workflow strategy used to address the research questions as a means of prioritizing and identifying potentially endocrine active features in locations where chemical sources, including potentially diffuse sources, are poorly understood. This prioritization technique could be advantageous for regulatory agencies wishing to integrate nontarget analyses into their monitoring efforts. Using biological endpoints to narrow the number of sites analytically assessed could also be attractive to local and regional resource managers with limited financial resources. There is also value of assessing these chemicals as mixtures or co-occurring contaminants when executing prioritized bioassays for unexpected chemicals and chemicals not included in a reference database such as the EDS21 Dashboard. This work also encourages the retention of unidentified masses of interest (Level 5 IDs) in prioritization frameworks to prevent bias in toxicological investigations towards the current knowledge base on chemicals of interest.

3.6 Acknowledgements

Funding for the present study was provided by the Minnesota Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (Grant M.L. 2015, Chp. 76, Sec. 2, Subd. 04c). Special thanks to Dr. Karl Oetjen for offering guidance with LC-HRMS analysis.

CHAPTER 4

SUSPECT ANALYSIS OF ON-SITE WASTEWATER TREATMENT TRACE ORGANIC CONTMAINANTS IN PASSIVE SAMPLERS DEPLOYED IN AN ADJACENT MINNESOTA LAKE

A paper to be submitted to *Science of the Total Environment*

Guyader, Meaghan E.[1,2]; Green, Emily [3]; Kiesling, Richard L.[4]; Schoenfuss, Heiko, L.[5]; Higgins, Christopher P. [1,6]

4.1 Abstract

On-site wastewater treatment systems (OWTSs) are a suspected source of widespread trace organic contaminant (TOrC) occurrence in Minnesota lakes. In this study, we compared TOrC composition within a shoreline residence's OWTS to that of porewater from an adjacent Minnesota lake. Both novel passive samplers and grab samples were concurrently collected from the septic tank (ST) and drain field (DF) of the OWTS over an 8-day period. Passive samplers were also deployed for 8 days in the sediment porewater of a lake location (HOME) proximal to the OWTS and a reference location (REF) from the same lake distal from any residences. Sample extracts were analyzed using liquid chromatography quadrupole time of flight mass spectrometry (LC-QToF/MS) in both positive (ESI +) and negative (ESI-) electrospray ionization modes. Features present across all field replicates and time points of the ST and/or DF were identified using a library spectra search algorithm. Chemicals identified with high confidence, including human waste chemicals, pharmaceuticals and their metabolites, hormonal steroids, personal care products and cleaning agents, flame retardants, surfactants, food and plant extracts, and amino acids were compiled into an interest list of OWTS-associated features. A suspect search of these chemical components was conducted on HOME and REF samplers, and spatiotemporal trends in their intensity and detection frequency were evaluated using principal component analysis. Strong indicators of wastewater presence, such as sucralose, were not detected at either lake location. The insect repellent DEET and high blood pressure medication, losartan, were the only notable

[1]Civil and Environmental Engineering Department, Colorado School of Mines
[2]Primary researcher and author
[3]Chemical Engineering Department, Colorado School of Mines
[4]Minnesota U.S. Geological Survey
[5]Biology, Department, St. Cloud State University, St. Cloud, MN, U.S.
[6]Corresponding author

detections unique to HOME versus REF locations. Lower sampling rates of passive samplers at lake locations, low velocity and unexpected groundwater flow paths, and low spatial resolution in samples between the OWTS and lake likely inhibited down gradient detection of suspect search features. Assessment of temporal trends in ST and DF chemical intensity in passive samplers indicate field monitoring application for a broad range of wastewater TOrCs, with compromised ability for the analysis of exceptionally polar or nonpolar chemicals.

4.2 Introduction

The numerous sources of uncertainty in understanding trace organic contaminant (TOrC) discharges from on-site wastewater treatment systems (OWTSs) make field monitoring efforts challenging. TOrCs detected in the environment typically have a number of potential anthropogenic and natural sources (Lapworth et al., 2012), as this group encompasses a broad range of chemical structures, including pesticides, pharmaceuticals, and food waste, so commonly consumed or used by humans that they are now detected at ng/L to µg/L concentrations throughout the environment. For this reason, it is often difficult to distinguish a particular source of detected TOrCs by their identity alone. Temporal and spatial resolution in field sampling and analysis can address this uncertainty for certain sources; however, TOrC composition in raw wastewater treated by an OWTS is highly variable (Teerlink et al., 2012). OWTSs are representative of the lower boundary for sewershed size, typically treating the wastewater of a single-family residence. The composition of TOrCs present at any given moment in these systems are subject to the habits of the residents, which are difficult to predict without an extended study of the individual waste generators' behaviors (Gerrity et al., 2011; Teerlink et al., 2012). Additionally, OWTSs can act as a non-point source of TOrCs, dispersing pretreated wastewater into native soil and subsequently allowing any of the unattenuated TOrCs to travel within shallow groundwater paths (Carrara et al., 2008; Conn et al., 2010a; Godfrey et al., 2007). These shallow groundwater paths are also variable, as depth to water table typically changes with the season and flow paths can change after individual precipitation events (Fovet et al., 2018; Woodward et al., 2016). While strategies have been proposed to accommodate the uncertainty associated with studying OWTS TOrC discharges, grab samples and targeted aqueous analysis are typically limited in their ability to represent these complex systems (Ort et al., 2010; Schaider et al., 2017).

OWTSs are a suspected source of widespread TOrC occurrence in lakes throughout Minnesota (Baker et al., 2014; Guyader et al., 2018a (Chapter 2)). In the past ten years, pharmaceuticals, steroids, insect repellents, and pesticides were detected in lakes from both populated and remote regions, indicating influence from diffuse sources (Ferrey et al., 2015; Writer et al., 2010). Concurrent evidence of endocrine disruption in the region, also associated with fish feminization in wastewater effluent dominated streams (Liu et al., 2004; Vajda et al., 2008), drives hypotheses that OWTSs commonly used in the area could contribute to TOrC loadings. Targeted analysis for a list of 33 TOrCs, including known endocrine disruptors and wastewater associated chemicals, suggested higher concentrations in sediment porewater closer to residential OWTSs, but the particular TOrCs detected were not strong indicators of wastewater occurrence (Guyader et al., 2018a; refer to Chapter 2). Nitrogen, also used as an indicator of wastewater, was at a higher concentration in sites distant from residential OWTSs, but could have likely also been linked to agricultural and manure applications in the region. Further investigations of OWTS impacts on these lakes should investigate other potential wastewater indicators, such as site-specific wastewater chemicals and their biotransformation products, which have the potential to have similar or more pronounced effects to exposed biota (Fatta-Kassinos et al., 2011; Lapworth et al., 2012).

Novel analytical and sampling strategies are now available to address the challenges of TOrC source attribution. Liquid chromatography with high resolution mass spectrometry (LC-HRMS) is an advanced analytical strategy capable of distinguishing unique chemicals that co-elute and have the same nominal mass (Krauss et al., 2010; Peter et al., 2018). Chemicals can then be identified with different levels of confidence by comparing their monoisotopic mass, retention time, isotope spectra, and, if available, MS/MS fragmentation pattens to standards, spectral libraries, or *in silico* fragmentation patterns(Schymanski et al., 2014a). This technology is particularly advantageous for the analysis of suspected or unknown chemical components in an environmental matrix (Hernández et al., 2015). Suspect searches are typically constructed from prior knowledge of an environmental matrix, such as understanding of pharmaceuticals prescribed within a region, recalcitrant chemicals likely to persist environmental degradation processes, or sources known to affect a sampled area (Gago-Ferrero et al., 2018; Hug et al., 2014b; McEachran et al., 2017). For environmental matrices with less pre-existing knowledge, nontarget analysis is more advantageous for examining all chemicals within a matrix (Rotander et al., 2015; Schymanski et al., 2015b). However, feature prioritization is necessary to restrict investigation of spurious features outside the scope of research objectives, such as natural organic matter molecules or features introduced by contamination (Du et al., 2017b; Schollée et al., 2015).

Passive samplers could improve understanding of TOrCs in systems associated with great temporal variability. Passive samplers are sorbent materials maintained in an environmentally safe container that uptake environmental contaminants and allow for *in situ* elimination of matrix interferences (Arditsoglou and Voutsa, 2008; Fernández et al., 2014; Kot-wasik, 2010). The material is usually selective for certain chemicals. Uptake typically occur in stages: a kinetic uptake regime which allows the sampler to act as an integrative sampling device whose extracts show all contaminants that have come into contact with the sampler during deployment; an intermediate regime; and an equilibrium regime which can allow for determination of a time weighted aqueous concentration for contaminants (Mazzella et al., 2007). Performance testing of different passive sampler configurations has occurred in numerous laboratory experiments (Fauvelle et al., 2014; Morrison and Belden, 2016; Togola and Budzinski, 2007), and field application of these tools is gaining popularity (Harman et al., 2011; Li et al., 2010; McDonough et al., 2014; Moschet et al., 2015; Vetter et al., 2018). Challenges that must still be overcome include: accounting for loss of sorbents during desorption, accounting for variable flow conditions in deployment locations (particularly for groundwater), the effect of biofouling common in field applications of contaminant uptake, and appropriate methods of handling variable contaminant loadings (Cornelissen et al., 2008).

The objective of this study was to assess the presence of OWTS-derived TOrC discharges in Minnesota lakewater. The following research questions were addressed: 1) How can sampling strategies affect TOrC characterization of systems associated with high temporal variability?, and 2) What are the spatial trends of OWTS-associated LC-HRMS feature occurrence in a Minnesota lake adjacent to an OWTS? We hypothesized that lake locations more proximal to an OWTS would have more frequent detections of OWTS-associated TOrCs. To test this hypothesis, we conducted a field sampling campaign of the septic tank (ST) and drain field (DF) of a Minnesota OWTS, as well as two lake locations proximal to (HOME) and distal from (REF) the same OWTS. Passive samplers were deployed in sampling wells at each location for 2, 4 and 8 days, though the samplers deployed at the REF sute were pulled at 2,6,and 8 days. Grab samples were concurrently collected at the ST and DF locations to compare passive sampler performance for OWTS characterization. Samples were extracted and analyze using liquid chromatography quadrupole time of flight mass spectrometry (LC-QTOF-MS) in both ESI + and ESI - modes. A suspect list of LC-HRMS OWTS-associated features was developed by prioritizing and identifying features present throughout all field and time replicates in both ESI + and ESI – modes with a library spectra search algorithm. Monoisotopic masses of likely biotransformation products were also calculated using their established mass difference and added to the suspect search list. Intensity and detection frequency of OWTS-associated features at HOME and REF sites were then compared using principal component analysis (PCA).

4.3 Materials and Methods

4.3.1 Field Sampling

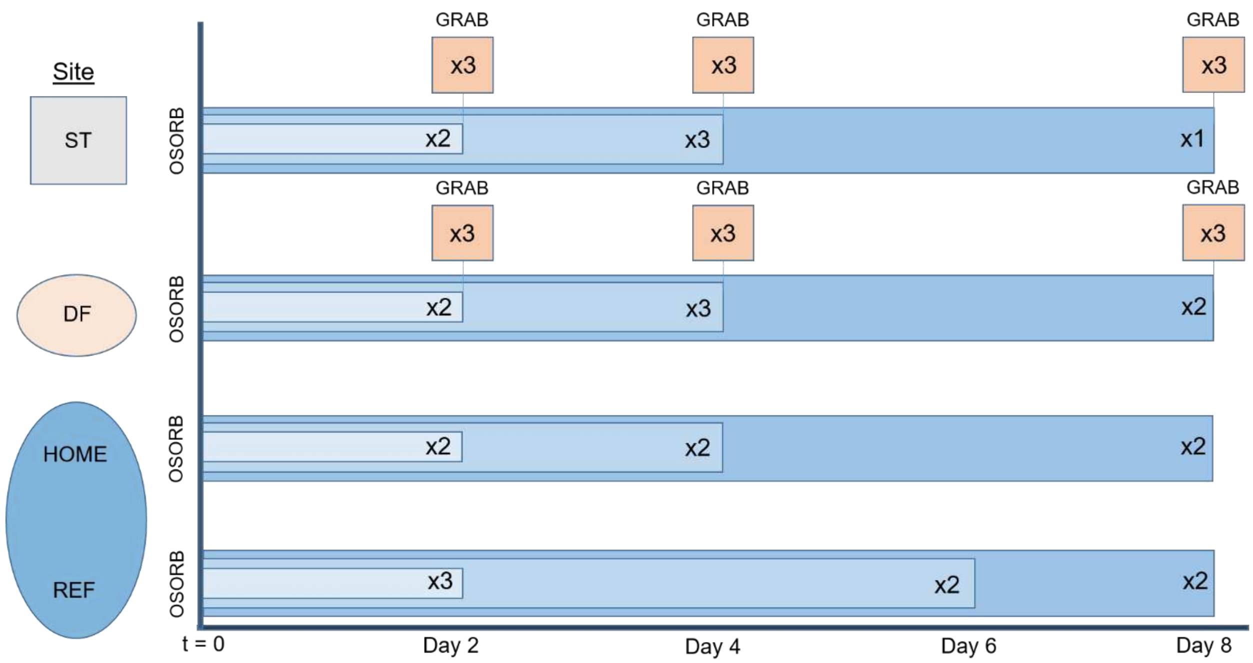

Figure 4.1 Sample inventory for each sampling location: ST=septic tank, DF=drain field, HOME=lake location close to OWTS, REF=reference lake location distant from OWTS. GRAB= grab samples extracted using mixed mode solid phase extraction, OSORB= sentinel passive samplers, time points for both GRAB and OSORB correspond to number of days after the OSORB samplers were deployed. Number of field replicates indicated by “xn”.

Sampling efforts were conducted at a single lake (Franklin Lake, Otter Tail County, MN). This lake was part of a previous sampling effort investigating TOrC concentrations in lake porewater proximal to and distal from shoreline residences whose wastewater is treated by single family OWTSs (Guyader et al., 2018a; refer to Chapter 2). Samples were collected from four locations: the septic tank (ST) and drain field (DF) of a shoreline residence's OWTS, and in the interstitial water of the lake's sediment at a location proximal to the sampled OWTS (HOME) and a lake location distant from any residential activity (REF). REF location selection was done using the same criteria in a previous field sampling effort (Guyader et al., 2018a; refer to Chapter 2). The septic system sampled was selected for its ease of access resulting from a pre-existing relationship with the property owners. Notably, at the time of sampling the system showed signs of compromised treatment indicated by water pooling at the DF. Resultant impacts on treatment efficiency are expected to affect the degree of biologic transformation of features identified in the ST. Additional groundwater sampling locations between the drain field and HOME sampling location were probed, but the groundwater table was deemed too low for establishing a passive sampling well. The available sampling equipment and the flow was at loo low a velocity to consistently replenish a supply sufficient for grab sampling within a reasonable timeframe.

Water samples were collected using grab samples and passive samplers during a field investigation conducted in September 2017. According to the manufacturer, the passive samplers (OSORB) used in this study are designed to monitor both pulse and continuous contaminant exposure events for a broad range of TOrCs (ABS Materials, 2016). A laboratory calibration study indicated faster sampling rates relative to traditionally used POCIS under flow conditions (Morrison and Belden, 2016). The configuration of the OSORB samplers was also preferred for the intended deployment in temporary groundwater sampling wells: a deployment holder for three POCIS is 15 cm x 16 cm, while a single OSORB is 6.5 cm x 1.5 cm and does not require a deployment holder (ABS Materials, 2016; Environmental Sampling Technologies, 2018). To the best of our knowledge, this is the first study to execute a field deployment of these samplers under anticipated highly variable exposure conditions (Teerlink et al., 2012).

Groundwater and lake porewater samples (DF, HOME, and REF sites) were acquired by establishing a well (~60 cm below ground surface) using 7.62 cm diameter polyvinyl chloride (PVC) piping with a slotted PVC well screen (0.5 mm slot opening). PVC piping was triple rinsed with DI water before placement. OSORB samplers were deployed in staggered duplicates and triplicates after presoaking in DI water for 20 minutes (Figure 4.1). Methanol-rinsed fishing line was tied through the ring and samplers were placed in replicate clusters to be immersed in the water that collected within the well. Samplers were then pulled at 2, 4, and 8 days after initial deployment to match the laboratory calibration study (Morrison and Belden, 2016), with the exception of the REF sampling location, for which samples were retrieved on day 6 in place of day 4. A PVC pipe was also placed in the ST to completely submerge the well screen and OSORB samplers were deployed in a similar manner. Grab samples were only collected from the ST and DF locations concurrently with the OSORB sample retrievals (i.e., at 2, 4 and 8 days). At the DF, a mini piezometer was placed to pump and collect drain field water in 1L amber glass bottles baked and pre-rinsed with methanol. At the ST, a stainless steel drive point piezometer (187 x 33.3 mm) was attached to the end of plastic tubing and pumped (35 mL/min), with water collected in 1 L amber glass bottles baked and pre-rinsed with methanol. Explicitly, the grab samples were collected to represent transient aqueous TOrC composition and are in no way temporally integrative. Grab samples were sent overnight on ice to the Colorado School of Mines (CSM; Golden, CO, U.S.). Passive samplers were kept frozen and shipped as a single batch after the last set of samples were pulled. Samples travelled overnight on ice in 50 mL polyethylene tubes (Falcon) to CSM.

4.3.2 Sample Preparation

Grab samples were extracted using a mixed mode solid phase extraction procedure developed by Eawag (Schollée et al., 2015). Briefly, triplicate 300 mL aliquots from the 1L amber glass bottles were vacuum filtered (GC/F filters, Whatman) and spiked with 15 ng of each surrogate standard (list in Table A.1). Mixed mode sample cartridges were prepared in according to the method described by the multilayered Eawag-cartridge type Silovo standard operating procedure with Envi Carb placed on bottom (Schymanski et al., 2014b). Sample loading was conducted with a Dionex AutoTrace280 (Thermo Scientific, Waltham, MA) instrument by preconditioning cartridges with methanol and MTBE, then loading the sample onto the cartridge. Samples were eluted with the cartridges in reverse flow direction on a manifold SPE to prevent analyte capture on the activated carbon material of the EnviCarb sorbent. Elution solvents were 6 mL of ethyl acetate and methanol buffered with formic acid, followed by 3 mL ethyl acetate and methanol (50:50 v/v) buffered with sodium hydroxide. Extracts were verified to be neutral with pH paper and then blown down and using an NEvap system to 100 µL. Extracts were reconstituted to 300 µL and transferred to 2mL autosampler vials where they were stored in at 4°C until analysis.

Passive samplers were extracted according to procedures recommended by the manufacturer (ABS Materials, Wooster, OH). The samplers were allowed to thaw at room temperature and then rinsed with DI water to remove sediment before being transferred to methanol-rinsed 40mL VOA vials. Samplers were extracted by adding 20mL acetonitrile to the vial and placing them on a shaker table for 1 hour. Samplers were then removed and 100 ng surrogate solution was added to all samples. Samples were then evaporated to dryness and reconstituted in 2 mL methanol and stored at 4°C until analysis.

4.3.3 LC-HRMS Data Acquisition

LC-HRMS features for grab sample and passive sampler extracts were acquired using a Sciex TripleTOF 5600 in both positive and negative electrospray ionization modes (ESI +/ ESI -, respectively). OSORB and grab sample extracts were diluted ten-fold in ultrafiltered water (Fisher) prior to 500 µL sample injections onto the instrument. Separation for ESI + mode analytes was achieved using a Luna C18 column (Luna, 3um C18(2) 100A 50x 4.6mm, fully porous silica, Phenomenex, Torrance, CA, U.S) eluted with water (Optima HPLC, Fisher) buffered with 0.1% formic acid and methanol. Separation for ESI – analytes was achieved using a Gemini C18 column (Gemini, 3um C18 11A 50x4.6mm fully porous organo-silica, Phenomenex, Torrance, CA, U.S.) eluted with water and methanol both buffered with 10mM ammonium fluoride (Fisher Scientific, Waltham, MA, U.S.). For both methods, the LC gradient used a 0.5 mL/min flowrate starting with 10% organic phase, ramping to 95% organic phase at 7 minutes, holding at 95% organic until 12 minutes, then returning to 10% organic and equilibrating until 18 minutes. The TripleTOF was operated using information dependent acquisition (IDA) with a minimum intensity of 1000 cps, dynamic exclusion, and no more than 12 candidates per cycle.

4.3.4 Developing Interest List of ST and DF Features

A suspect chemical list was developed from the passive sampler extracts deployed at the ST and DF. This characterization workflow started by conducting peak picking and feature alignment in Sciex MarkerView v 1.3.1 (input parameters informed by inspection of surrogate standard retention time and mass tolerance in Sciex PeakView v 2.2). Sample data was processed in batches according to sampling location (ST or DF) and ionization mode (ESI + or ESI-). Datasets are referred to herein as: *OSORBposST* for OSORB samplers deployed at the ST acquired in ESI + mode, *OSORBnegST* for OSORB samplers deployed at the ST acquired in ESI - mode, *OSORBposDF* for OSORB samplers deployed at the DF acquired in ESI + mode, *OSORBnegDF* for OSORB samplers deployed at the ST acquired in ESI – mode. Feature alignment for all datasest was conduicted using a mass tolerance of 5ppm and a retention time tolerance of 0.1 mintes.

Chemical components with with non-zero responses in all field replicates and all time point measurements from each dataset were prioritized. These prioritized features were then inspected in MasterView v 1.1, deprioritizing features with intensities less than five times the field blank corresponding to each dataset and executing a library search algorithm to identify features as chemical constituents (peak picking parameters in Figure B.1). Schymanksi Level 1 and 2 features that met MasterView peak picking standards in all field replicates and time measurements for each dataset were compiled into “Interest List 1”.

4.3.5 Performance Assessment of Passive Samplers

Intensity trends for Interest List 1 features were evaluated using normalized peaks tables from each of the OSORB datasets in Markerview. A correction factor was applied to feature intensities from each sample calculated using most likely ratio (MLR) normalization (Lambert et al., 2013). Correction factors are determined using select features shared across sample replicates, typically shared noise, chemical constituents, or spiked isotope standards. Linear regressions for the change in each feature’s intensity were then calculated in Microsoft Excel and plotted using Origin b9.5.195.

The feature prioritization, identification, and linear regression process was also executed on grab samples from the DF and ST. The datasets are referred to as *GRABposST, GRABnegST, GRABposDF, GRABnegDF*. Features used for this analysis were also required to meet MasterView peak picking parameters for all field replicates and time measurements within each respective dataset.

4.3.6 Multivariate Analysis of Suspect Search at Lake Locations

A comparison of Interest List 1 feature detection frequency and intensity at HOME versus REF sites was conducted using MarkerView’s principal component analysis (PCA) function. Interest List 1 was used as the XIC list and HOME and REF OSORB sampler intensity of these features was investigated in MasterView (peak picking parameters same as those used to develop Interest List 1). The library search function was also run to confirm the feature identities matched those proposed in Interest List 1 (input parameters in Figure B.1). Features needed to have an MS/MS spectra generated for at least one sample replicate in at least one time measurement; otherwise, the feature was not reported as being present in either lake location. Peak files were then generated and examined in MarkerView. Feature responses were exported as generic text files, log_2 transformed, and reimported before MLR normalizing the responses. These response values were log transformed to make the distribution of feature responses within and across all samples more Gaussian. Pareto-scaled principal component analysis (PCA) scores and loadings plots were then generated for these transformed data. Clustering of HOME and REF samples, and corresponding vectors on the loadings plot were interpreted to assess the chemical similarity of the two lake locations with respect to Interest List 1 features.

4.3.7 Quality Control and Quality Assurance

LC-HRMS data interpretation relies heavily on mass accuracy to confirm the identity of known, suspect, and unknown features. Calibration solution (APCI Pos and APCI Neg solutions from Sciex, Framingham, MA) was run at the beginning of each instrument run to calibrate the mass accuracy of the TOF and MS/MS scans. The same calibration solution was run throughout the run after every four samples and confirmed to pass mass accuracy requirements (<5 ppm mass error for all mass components). The mass accuracy of surrogate standards was also examined in regularly run calibration verification standard and spiked environmental samples. Blank contamination was also examined in laboratory and field blanks and the source was regularly rinsed by including double blank samples regularly throughout instrument runs.

4.4 Results and Discussion

4.4.1 Defining Interest List 1

Table 4.1 List of Level 2 Schymanski Scale identified human waste chemcials present in all field replicates and time points for passive samplers deployed in the ST and/or DF. The associated sample matrix, including grab samples from the ST and DF, is indicated by either a + or – indicating the ionization mode in which they were acquired.

Human Waste Chemcial	Feature (m/z_RT) (ESI+, ESI-)	Use	OSORB ST	OSORB DF	GRAB ST	GRAB DF
Sepantronium cation	363.0691_7.04	survivivin suppressant	+	+		
3-Furancarboxylic acid	255.1584_8.85	urine component, also in honey	+	+		
5-methylmyosmine	161.1323_10.4	metabolite of nictoine		+		+
Deoxycholic acid	391.2853_10.62	metabolic byproduct of intestinal bacteria	-			
Kynurenic acid	190.0491_8.07	human metabolite of trytophan, anticonvulsant	+			
2-Indolinone	134.0603_7.7	human metabolite of 1H-indole	+			

Table 4.1 continued

Human Waste Chemcial	**Feature (m/z_RT) (ESI+, ESI-)**	**Use**	**OSORB ST**	**OSORB DF**	**GRAB ST**	**GRAB DF**
2-Octenoyl-carnitine	286.2011_7.27	supplement or natural weigth loss chemical	+			
Ursodeoxycholic acid, MNH4	410.3264_10.77	dissolve gallstones and liver disorders	+			
2,3-dimethoxy-5methylbenzoquinone	183.081_9.1	e. coli and human metabolite	+			
Vitamin A acid	301.2153_10.11	metabolite of Vitamin A, eye supplement		+		
4-Pyridoxic acid	181.9916_6.43	urinary metabolite of vitamin b6	-			
Cholic acid	407.2795_10.26	bile acid	-			

Table 4.2 List of Level 2 Schymanski Scale identified hormonal steroids present in all field replicates and time points for passive samplers deployed in the ST and/or DF. The associated sample matrix, including grab samples from the ST and DF, is indicated by either a + or – indicating the ionization mode in which they were acquired.

Hormonal Steroid	**Feature (m/z_RT) (ESI+, ESI-)**	**Use**	**OSORB ST**	**OSORB DF**	**GRAB ST**	**GRAB DF**
Digitoxigenin	375.2883_10.35	steroid	+			
Digoxigenin	391.2834_10.09	steroid	+			
Beclomethasone	409.1758_10.07	steroid medication	+			

Table 4.3 List of Level 1 and 2 Schymanski Scale identified pharmaceuticals present in all field replicates and time points for passive samplers deployed in the ST and/or DF. The associated sample matrix, including grab samples from the ST and DF, is indicated by either a + or – indicating the ionization mode in which they were acquired.

Pharmaceutical	**Feature (m/z_RT) (ESI+, ESI-)**	**Use**	**OSORB ST**	**OSORB DF**	**GRAB ST**	**GRAB DF**
Losartan	423.1695_9.51	angiotensin reeptor agonist (cozaar)	+	+		
Amprolium cation	243.232_11.68	antiprotozoal in chicken ag	+			
Diphenydramine	256.2995_7.31	benadryl, antihistamine	+			
O-Desmethylvenlafaxine	264.1954_6.42	antidepressant	+			
Venlafaxine	278.211_7.13	antidepressant (effexor)	+			
Losartancarboxaldehyde	421.155_10.23	metabolite of losartan		+		
Loxoprofen	493.3225_8.9	NSAID	+	+		
Lansoprazole	368.0677_9.28	heartburn medication (prevacid)	-		-	-
Phenylglyoxylic acid	149.0614_8.15	cosmetic, food additive,	-			
N-(4-Hydroxyphenyl)propanamide	166.086_6.75	metabolite of fenatyl	+			

Table 4.4 List of Level 1 (DEET and oxybenzone) and 2 (all others) Schymanski Scale personal care product or cleaning agents present in all field replicates and time points for passive samplers deployed in the ST and/or DF. The associated sample matrix, including grab samples from the ST and DF, is indicated by either a + or – indicating the ionization mode in which they were acquired.

Personal Care or Cleaning Agent	**Feature (m/z_RT) (ESI+, ESI-)**	**Use**	**OSORB ST**	**OSORB DF**	**GRAB ST**	**GRAB DF**
1-Cyclohexylethanol	111.1173_10.05	polish, wax, cleaning products	+			
Quinolin-8-ol	146.06_7.43	antiseptic, chelating agent, liquid bandages	+			
Quinolin-2-ol	146.06_7.85	antiseptic, chelating agent, liquid bandages	+			
DEET	192.1382_9.35	insect repellent	+			
Oxybenzone	229.086_10.21	sunscreen	+			

Table 4.5 List of Level 2 Schymanski Scale flame retardants present in all field replicates and time points for passive samplers deployed in the ST and/or DF. The associated sample matrix, including grab samples from the ST and DF, is indicated by either a + or – indicating the ionization mode in which they were acquired.

Flame Retardant	**Feature (m/z_RT) (ESI+, ESI-)**	**Use**	**OSORB ST**	**OSORB DF**	**GRAB ST**	**GRAB DF**
Tri(3-chloropropyl) phosphate	327.3367_9.6	flame retardant	+			
Tris(2-butoxyethyl) phosphate	399.2496_10.61	flame retardant	+			

Table 4.6 List of Level 2 Schymanski Scale surfactants present in all field replicates and time points for passive samplers deployed in the ST and/or DF. The associated sample matrix, including grab samples from the ST and DF, is indicated by either a + or – indicating the ionization mode in which they were acquired.

Surfactant	**Feature (m/z_RT) (ESI+, ESI-)**	**Use**	**OSORB ST**	**OSORB DF**	**GRAB ST**	**GRAB DF**
Pentaethylene glycol	239.1483_6.25	surfactant	+	+		
Hexaethylene glycol	283.1746_6.61	surfactant	+	+		
PEG-7mer Ammonium adduct	344.2281_6.87	surfactant	+	+		
PEG-8mer Ammonium adduct	388.2534_8.64	surfactant	+	+		
Nonaethylene glycol	415.2534_7.24	surfactant	+	+		
Undecaethylene glycol	503.305_7.52	surfactant	+	+		
PEG-9mer Ammonium adduct	432.2794_7.36	surfactant		+		

Table 4.6 continued

PEG-10mer Ammonium adduct	476.3052_7.51	surfactant		+		
PEG-11mer Ammonium adduct	520.331_7.64	surfactant		+		
PEG-12mer Ammonium adduct	564.357_7.76	surfactant		+		
Nonapropylene glycol	541.3924_10.17	surfactant		+		
Dodecyl sulfate	265.1477_10.64	surfactant in personal care products	-			

Table 4.7 List of Level 2 Schymanski Scale food and plant extracts present in all field replicates and time points for passive samplers deployed in the ST and/or DF. The associated sample matrix, including grab samples from the ST and DF, is indicated by either a + or – indicating the ionization mode in which they were acquired.

Food and Plant Extract	**Feature (m/z_RT) (ESI+, ESI-)**	**Use**	**OSORB ST**	**OSORB DF**	**GRAB ST**	**GRAB DF**
4-Hydroxy-6-methoxycoumarin	193.1433_9.74	flavoring	+	+		
2,7,8-Trimethyl-2-(.beta.-carboxyethyl)-6-hydroxychroman	265.143_9.2, 263.1287_6.6	derivative of culmorin, mycotoxin	+		-	-
Sucralose	395.0063_7.39	artifical sweetener	-	-	-	
Benzaldehyde	107.0498_9.44	food extract with industrial uses	+			
2,2-Dimethyl-4-pentenoic acid	129.1275_10.02	flavouring ingredient	+			
Lavandulol	155.1433_9.56	monoterpene in essential oils	+			
4-Hydroxynonenal	157.1586_10.35	oxidation of lipids	+			
1H-Indole-3-carboxylic acid	162.0549_7.61	plant extract and pharmaceutival ingredient	+			
2'-Hydroxy-6'-methoxyacetophenone	167.0703_10.04	plant extract and pharmaceutical ingredient	+			
3-Indoleacetic acid	176.0703_8.12	auxin, plant hormone	+			
Indoleacrylic acid	188.0701_9.01	plant growth hormone	+			
Caffeine	195.0872_7.04	stimulant	+			
Piperine	286.143_9.88	black pepper extract	+			
(-)-Riboflavin	377.1466_7.27	supplement in food	+			
Acetic Acid	223.154_7.57	used in vinegar		+		
Isoalantolactone	233.1164_9.1	sesquiterpene with antifungal properties		+		
Atractylenolie III	249.1844_9.81	bioactive isolate of root system		+		
10-Methylundecanoic acid	199.1708_10.68	fatty acid most common in coconut oil	-			
6,4'-Dihydroxyflavone	253.0502_8.84	isoflavone from soy protein	-			

High confidence features (level 1 and 2) from Interest List 1 are reported in Tables 4.1 thru 4.7. Features are listed according to the likelihood of their associated with wastewater: human waste chemicals including metabolites of drug products and biosynthesized chemicals, hormonal steroids, pharmaceuticals, personal care and/or cleaning agents, flame retardants, surfactants, and food and plant extracts. The food and plant extracts could also be naturally derived chemicals from flora and fauna in the area, with the exception of sucralose, an artificial sweetener, which is a strong indicator of wastewater presence (Oppenheimer et al., 2011). Far more features were identified with high confidence in ESI + mode (85% of high confidence features) relative to ESI – mode. DEET was one notable detection from the ST feature list. This insect repellent was identified as an environmental TOrC at higher concentrations in lake porewater sampled closer to residences with OWTSs relative to locations distant from residential activity; however, its associated usage with outdoor activity made it a weak indicator of OWTS effluent (Guyader et al., 2018a; refer to Chapter 2). Irregular occurrence on this contaminant in the sampled OWTS, explicitly its detection in the ST but not the DF (Table 4.1), validate concerns that this contaminant is a poor indicator for OWTS discharges in the environment (see further discussion in Section 3.2.2). Also notable was the detection in the ST of quinolin-8-ol (i.e., 8-hydroxyquinoline), which was also identified in a previous study of another Minnesota lake at locations associated with endocrine activity (Guyader et al., 2018b; refer to Chapter 3). The detected chemicals DEET, caffeine, tris (2-butoxyehtyl) phosphate and the flame retardants detected in the ST were also prioritized as contaminants of concern in previous studies conducting LC-HRMS feature prioritization within wastewater treatment systems of varying scales (Blum et al., 2017; Erickson et al., 2014). The ST and DF sampled also contained less-common chemicals unique to the residents of this household, such as the high blood pressure medication, losartan.

Many features were reported across multiple ST and DF feature lists. Features reported across multiple lists were deemed to have more importance as suspect search chemicals considering the increased likelihood of the chemical being detected. Most commonly, features were detected at both ST and DF locations by passive samplers acquired in ESI + mode. These features include a component of urine, 3-furancarboxylic acid; the biogenic suppressant of a protein inhibitor, sepatronium; the high blood pressure medication, losartan; the non-steroidal anti-inflammatory drug (NSAID), loxoprofen; and several surfactants. Sucralose (log K_{ow} = - 1.5) was the only feature detected in *both* ST and DF passive samples acquired in ESI – mode. Sucralose is a strong indicator of wastewater occurrence in environmental compartments (Oppenheimer et al., 2011). Several of the Interest List 1 features, derived from the OSORB samples, were also detected in all grab sample replicates and time measurements taken from the same location and acquired in the same ionization mode. These features include the nicotine metabolite, 5-methylmyosmine, detected in both DF GRAB and OSORB samplers acquired in ESI - mode, as well as sucralose found in both ST OSORB and GRAB samples. Although sucralose was identified in the DF ESI – OSORB samplers, it was not identified in the GRAB DF samples. OSORB samplers apparently did not always capture some chemicals observed in the grab samples, including lansoprazole, which was detected in ESI – mode in the grab samples from both DF and ST locations, but was only detected at the ST of the OSORB samplers. It is important to note that DF features are less conclusively OWTS-associated because this sampling location is both a component of an engineered wastewater treatment system and an environmental compartment capable of receiving TOrC loadings from other non-point sources.

Several Interest List 1 features were identified to link to each other through biotransformation reactions, specifically human metabolism of chemicals. Examination of these pairs is valuable for predicating or contextualizing the relative abundance of OWTS associated parent compounds and transformation products (TPs) in down gradient environments. Deoxycholic acid and ursodeoxycholic acid, for instance, are secondary metabolites of cholic acid formed in the human intestine (Devlin and Fischbach, 2015). Deoxycholic acid is known to act as a strong indicator of human fecal matter versus other mammalian species, lasting hundreds of years and used in archaeological studies of septic systems (Zocatelli et al., 2017). Pharmaceuticals, their metabolites, and other TPs were also identified: venlafaxine and its metabolite desmethylvenlafaxine (both in ST only) (Dubey et al., 2013), and losartan's (found in both ST and DF) hydrogenation TP losartan carboxyaldehyde (found only in DF) (Jiang et al., 2015). Both venlafaxine compounds are, therefore, not expected to occur down-gradient, while losartan carboxaldehyde, which likely required aerobic conditions for transformation, is the expected down-gradient form of this OWTS-associated chemical.

4.4.2 Evaluating OSORB Performance

Polarity of Chemicals in OSORB. The polarity of chemicals unique to OSORB versus GRAB samples is presented in Figure 4.2. The median log K_{ow} values of compounds detected in OSORB samplers are slightly less than that of the compounds detected in GRAB samples from the ST (Figure 4.2a), and are very similar in the DF (Figure 4.2b). The widest range of log K_{ow} values was observed in the features unique to the GRABnegST dataset, with 1-oleoyl-sn-glycero-3-phosphoethanolamine (log K_{ow} = 10.4), a human waste chemical, being the most nonpolar chemical identified, and the sugar product, D-mannitol, representing the most polar (log K_{ow} = -4.7). The range of log K_{ow} values in the OSORBposDF, OSORBposST, and OSORB neg ST was still wider than that from the features unique to the GRABposDF and GRABnegDF. This suggests that while OSORB samplers are appropriate for the analysis of chemicals with a range of polarities (both positive and negative log K_{ow}'s), they are not selective for extremely polar or nonpolar chemicals with consistent presence at a deployment location.

In both the ST (Figure 4.2a) and DF (Figure 4.2b), there were more ESI + features identified in the OSORB versus the GRAB samples, and more ESI – features identified in the GRAB samples versus OSORB samplers. The ESI negative features missed by the OSORB sampler from the same sampling location and ionization mode include pharmaceuticals, human waste chemicals, plasticizers, fatty acids, steroids, and food products. Chemicals in both ST and DF GRAB samples missed by OSORB include the plasticizer 4,4'-sulfonylbisphenol (log K_{ow} = 1.9) and the human waste chemical, monobutyl phthalate (log K_{ow} = 3.1). Both of these chemicals fall within the range of log K_{ow} values for features unique to OSORB samplers, suggesting polarity is not the only chemical property which controls OSORB sampling rates. This is consistent with findings that other factors, such as flow rate, pulse versus continuous exposure conditions, and acid/base properties of contaminants can affect their uptake onto sorbent materials, particularly passive samplers (Morrison and Belden, 2016). Consideration of slower uptake in low flow conditions could also influence the fewer unique features identified in the DF versus ST OSORB samplers. Explicitly, analogous datasets for the ST and DF are from total fewer features detected at the DF relative to the ST, rather than the DF containing more shared features with grab samples or features without published log K_{ow} values. In addition, the DF sampling location could have altered uptake ability as a result of different matrix conditions, such as temperature or natural organic matter composition (Ghosh et al., 2014). While several chemicals were missed, the same compound classes were identified uniquely by OSORB and GRAB samplers, the ease with which the OSORB samplers were deployed make them a viable alternative for groundwater and temporally variable OWTSs.

Temporal trends in Interest List 1 Features. The linearity of temporal changes in individual Interest List 1 feature intensity were assessed using Pearson's r correlation. The r value for each feature, plotted on the x-y plane as their retention time and mass, is represented using a colorscale in Figure 4.3. For the n = 6 OSORB samples collected in the ST and DF, the r-critical value is 0.729 using a one-way t-test. The closer Pearson r values without significant increasing or decreasing trends are displayed in light gray. Pearson r-values with significantly increasing ($r > 0.729$; green) and significantly decreasing ($r < -0.729$) represent features with increasing and decreasing intensity trends, respectively.

Temporal trends in feature intensity varied according to feature mass, a proxy for the size of each chemical; feature retention time, correlated with the hydrophilicity or hydrophobicity of each chemical; as well as sampling location and sample ionization mode. In the OSORBposST dataset (Figure 4.3a), a number of features had either constant or slightly increasing ($r < 0.5$) changes in feature intensity, such as diphenhydramine (r = 0.0082). It appears that more hydrophobic features, particularly those eluting after 9 minutes, had more irregular feature intensity throughout the sampling period, with both strong increasing and decreasing feature trends. Notably increasing features from this dataset include 2-hydroxy-4-methoxybenzophenone (r = 0.91947), losartan (r = 0.74359), and DEET (r = 0.66101). Decreasing features include tris(3-chloropropyl) phosphate (r = -0.85784) and beclomethasone (r = -0.74789).

Other features, including diphenhydramine (r = 0.0082), demonstrated constant feature intensity over the deployment period. In the OSORBposDF it appeared mass, rather than polarity, had an affect on the dataset (Figure 4.3a), a number of features had either constant or slightly increasing ($r < 0.5$) changes in feature intensity, such as diphenhydramine (r = 0.0082). It appears that more hydrophobic features, particularly those eluting after 9 minutes, had more irregular feature intensity throughout the sampling period, with both significantly increasing and decreasing feature trends. Notably increasing features from this dataset include 2-hydroxy-4-methoxybenzophenone (r = 0.91947), losartan (r = 0.74359), and DEET (r = 0.66101). Decreasing features include tris(3-chloropropyl) phosphate (r = -0.85784) and beclomethasone (r = -0.74789). The decreasing features are expected to desorb over the length of the deployment, while the increasing features are still sorbing to the sampler material.

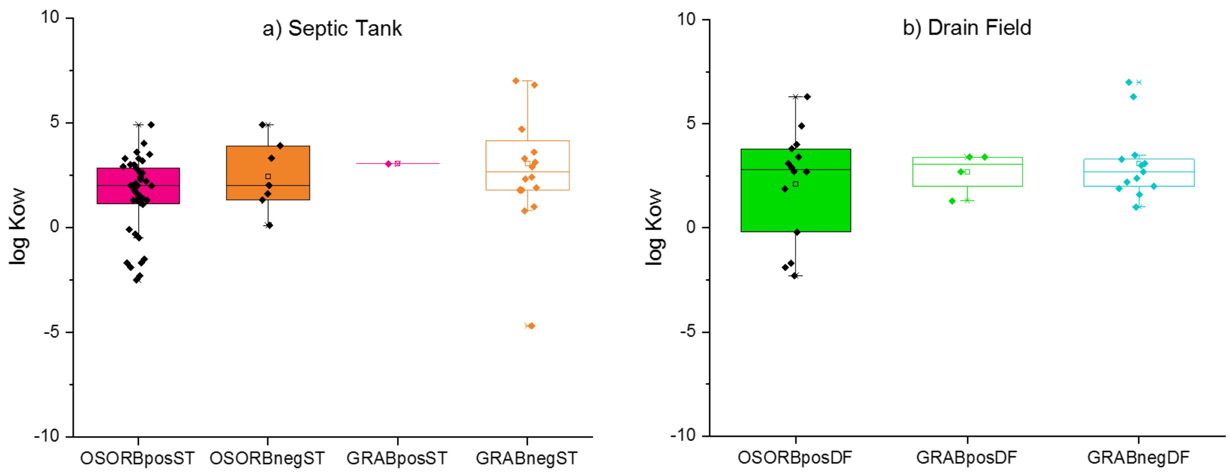

Figure 4.2 Comparison of feature polarity unique to OSORB and GRAB samples taken from the a) septic tank and b) drain field. Features shared by OSORB and GRAB samples omitted. There were no features unique to the OSORBnegDF dataset. Each point represents the log K_{ow} of a feature unique to its respective dataset.

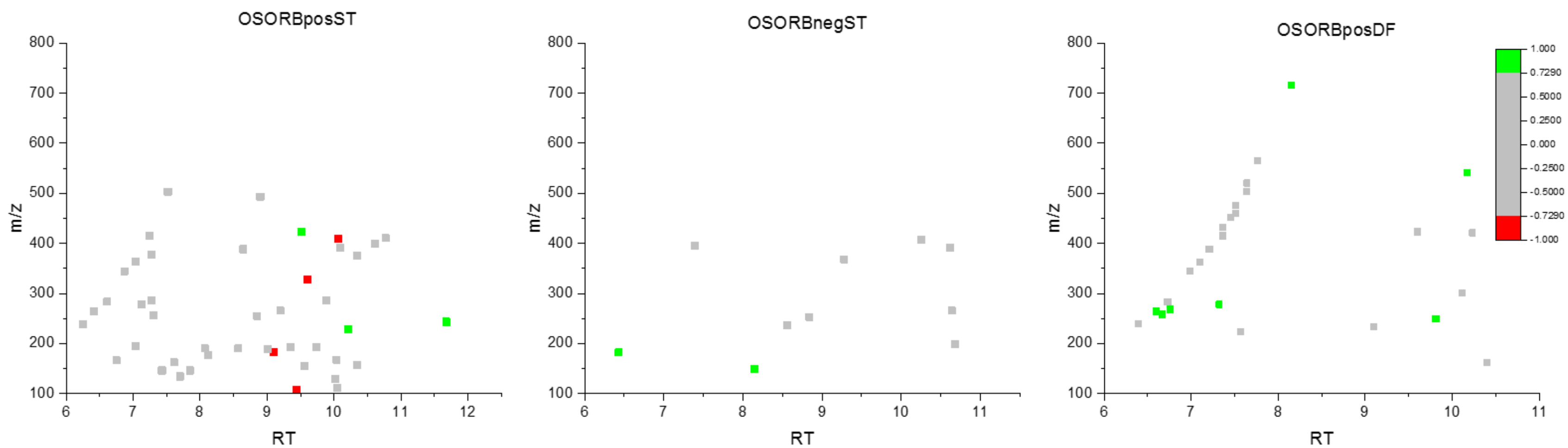

Figure 4.3 Pearson r correlation from OSORB samplers deployed in septic tank acquired in ESI + mode, OSORB samplers deployed in septic tank acquired in ESI – mode, and OSORB samplers deployed in drain field acquired in ESI + mode. Features identified by mass (m/z) in Daltons and retention time (RT) in minutes.

In addition, features, including diphenhydramine (r = 0.0082), demonstrated constant feature intensity over the deployment period. In the OSORBposDF it appeared mass, rather than polarity, had an affect on feature intensity trends. Chompounds with molecular weight ions between 200 Da – 300 Da and above 600 Da had increasing intensity trends. The linear cluster of features in the OSORBposDF dataset are surfactant structures, particularly different chain lengths of polyethylene glycol. All the detected entities of this homologous series had no significant change in feature intensity with time. The OSORBnegST dataset had only constant or significantly increasing feature intensity trends (Figure 4.2b). While there are fewer features in this list, it appears that the more polar (RT < 8.5 min) and hydrophobic features had increasing intensity trends. Feature mass did not appear to have an effect on temporal intensity trends. The OSORBnegDF dataset only had a single feature, sucralose which did not significantly change in intensity. Sucralose had a similar temporal pattern in the OSORBnegST dataset.

Overall, it appears that the OSORB samplers were capable of consistently measuring the presence of a wide range of OWTS-associated chemicals; however, their uptake was inconsistent. Longer deployment times for these samplers should be assessed; however, it appears that some chemicals are likely to desorb during the deployment period.

4.4.3 Suspect Search and Multivariate Analysis of OWTS-Associated Compounds at Lake Locations

A matrix of feature responses for ESI + Interest List 1 features in all HOME and REF replicates was analyzed in multivariate space using PCA (Figure 4.4). Again, this interest list of features was developed from characterization by OSORB samplers only, as only passive samplers were used to monitor the lake locations. The scores plot, Figure 4.4a, indicates a clustering in the linear combination of feature responses for REF samples along principal component 1 (PC1; attributed 50.2% of variance in log 2 transformed response values). There is also a noticeable separation of field replicates for the 2,4 and 8 day time measurements along PC1. The shorter field deployments (Day 2 and Day 4) cluster closer to the REF samples, while the 8 day deployment replicates show the furthest separation from all other samples. No ESI – OWTS-associated features were identified at either HOME or REF sites, hindering a PCA plot for these data. Furthemore, the strong indicator of wastewater presence, sucralose, was not detected at either lake location.

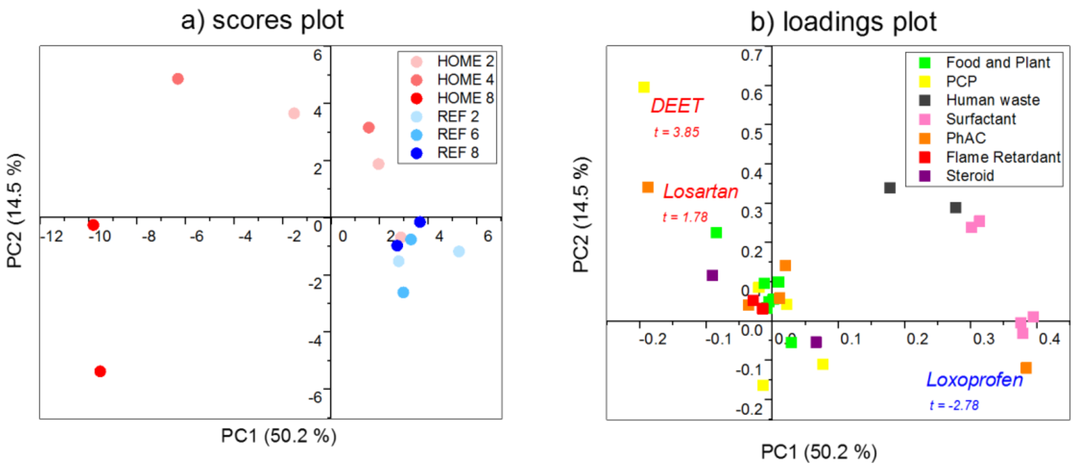

Figure 4.4 Principal component analysis a) scores and b) loadings of ESI + Interest List feature responses in HOME and REF OSORB samplers. The notable detections DEET, losartan, and loxoprofen are annotated on the loadings plot with their corresponding t-values from a Welch t-test of feature intensity in HOME vs REF samples. Loading plot features are color coded to indicate feature presence on the ST, DF or both ST and DF feature lists.

Investigation of the loadings plot (Figure 4.4b) identified features with the largest impact on sample separation on the scores plot. DEET was a major driver of separation for the HOME versus REF samples (t-value from Welch t-test = 3.85 > t-critical). This observation of higher DEET concentrations closer to OWTSs is consistent with previous studies conducted at this and other Minnesota lakes (Guyader et al., 2018a, 2018b (refer to Chapters 2 and 3), Writer et al., 2010). However, only the high blood pressure medication, losartan, correlated with this chemical's presence. This study verifies DEET can originate from a residential OWTS in Minnesota and plausibly contributes to DEET occurrence in Minnesota lakes. However, we agree with previous studies that it is still a weak indicator of wastewater occurrence, particularly if it is detected on its own in environmental matrices (Del Rosario et al., 2014). Interestingly, the non-steroidal anti-inflammatory drug, loxoprofen was at a higher concentration at the REF site (t = -2.78). Other studies of on-site wastewater treatment systems identify this as a notable detection in OWTSs (Gago-Ferrero et al., 2017; Gros et al., 2017), but suggest that this compound has a high limit of detection which typically prevents its detection in environmental compartments (Gago-Ferrero et al., 2017). The presence of this chemical at the HOME location is, perhaps, a line of evidence for the presence of OWTS impact at the HOME sampling location. In addition, this drug product is only approved for manufacturing and distribution in countries outside of the United States, which brings into question the confidence of this chemical's identity (Drugs.com, 2019). The t-values associated with both DEET and loxoprofen are still relatively close to t-critical for this dataset (df = 12, t = 2.2), suggesting that the difference in concentrations between HOME and REF sites is still relatively small. Several surfactants were also at higher concentrations in the REF and HOME 2 and 4-day time measurements, indicating these features are also poor indicators exclusively of OWTS impact, as they could come from other sources. Integration of LC-HRMS into suspect screening for this case study proved a useful tool for site-specific investigation; however, in the future it should be paired with higher spatial resolution in down gradient samples or sampling of additional OWTSs to encourage feature prioritization. Toxicological relevance could also be combined to improve feature prioritization, such as integrating endocrine disruption bioassays as done in a previous study of Minnesota lakes (Guyader et al., 2018b; refer to Chapter 3). A regional hydrogeological assessment of the lake rated it's geologic sensitivity as groundwater vertical travel time (to the target zone/water table) as years to decades based on tritium tracer studies (Ekman, 2002). More rapid non-point transport pathways, such as recreation and stormwater runoff, likely still contribute to TOrC presence in Minnesota lakes.

4.5 Implications and Conclusions

This study prioritized an expanded list of OWTS-associated features to further investigate the presence of wastewater in Minnesota lakes. DEET, commonly detected in environmental studies of TOrC occurrence, was identified within the sampled ST, suggesting it can occur in regional wastewater. However, its absence from the DF suggests it can still be attenuated and likely has other sources. The absence of notable wastewater indicator sucralose further challenges conclusions of OWTS occurrence at the sampled lake location proximal to the OWTS. More through characterization of groundwater flow paths is recommended to address potential uncertainty introduced by low spatial resolution in sampling locations. OSORB passive samplers provided adequate understanding of OWTS associated compounds, but could have limited understanding of OWTS feature presence at lake locations, particularly for polar or nonpolar substances. Overall, this LC-HRMS analysis provided a preliminary assessment for field monitoring efforts for complex TOrC loadings to Minnesota lakes.

4.6 Acknowledgements

Funding for the present study was provided by the Minnesota Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (Grant M.L. 2015, Chp. 76, Sec. 2, Subd. 04c). We extend special thanks to the family who offered their septic tank up to science!

CHAPTER 5
CONCLUSIONS

The objective of this dissertation was to investigate OWTSs as sources of TOrCs in Minnesota lakes as well as an analysis of TOrCs as potential causative agents for endocrine disruption in resident fish. Studies prior to this work identified OWTSs as one of a number of potential diffuse sources to Minnesota lakes. Causative agents of endocrine disruption at these lakes were largely speculated from studies of fish exposed to the effluent of centralized wastewater treatment plants. Overall, the goals of Chapter 2 were to investigate OWTS contribution to TOrC occurrence and endocrine disruption using traditional methods of aqueous analysis. In Chapters 3 and 4, LC-HRMS aqueous analysis was used to provide a more thorough understanding of chemical agents contributing to observations of pronounced endocrine disruption and site-specific indicators of wastewater presence in lake locations, respectively. Discussion of the findings and broader significance from these research efforts is provided in the following section. This dissertation concludes by discussing recommendations for future work.

5.1 Summary of Findings

Below is a summary of the findings for each research objective in this dissertation, presented as a discussion of how the findings relate to each objective and hypothesis.

5.1.1. Objective 1: Examining chemical and biological signatures of OWTS proximity in five Minnesota lakes

The aim of the study described in Chapter 2 was to evaluate spatial changes in Minnesota lake TOrC concentrations with respect to OWTS proximity. Resident fish from each of the sampling locations were also assessed for biomarkers of endocrine disruption. Previous efforts to align chemical and biological datasets assessing non-point sources of TOrCs to Minnesota lakes were hindered by only collecting a single, whole-lake sample. This approach fails to assess the localized TOrC composition more representative of fish exposure conditions. Below is a discussion of how the findings of this study related to Hypothesis 1:

Hypothesis 1: Elevated TOrC concentrations and positive endocrine disruption assays will occur at lake locations proximal relative to those distal from shoreline residences

The most significant findings of this work include:

- Elevated TOrC concentrations and fish blood vitellogenin concentrations are associated with residential proximity, which supports hypotheses of influence from OWTS discharges
- The use of targeted aqueous analysis alone in field monitoring studies may omit important components of TOrC mixtures that act as causative agents of endocrine disruption
- Influence from additional diffuse was also suggested by elevated nitrogen concentrations and detection of 17α- ethinylestradiol at reference sites

5.1.2. Objective 2: Investigating potential causative agents of endocrine disruption using LC-HRMS

The aim of the study described in Chapter 3 was to use biological endpoints to inform an LC-HRMS prioritization workflow in an effort to prioritize potential endocrine active chemicals within a Minnesota lake. Hormonal steroids are usually assumed to be the dominant causative agents of endocrine disruption in fish exposed to wastewater effluent. This assumption does not align with findings from Chapter 2, in which the targeted hormonal steroids and other known endocrine disruptors were not detected at lake locations (limits of detection for hormonal steroids ranged from 1ng/L to 9ng/L) associated with pronounced evidence of estrogen agonism. An LC-HRMS analysis and a feature prioritization workflow was executed to identify unexpected chemicals which may have contributed individually or additively with co-contaminants to the endocrine activity of the environmental mixture. Below is a discussion of how the findings of this study related to Hypothesis 2:

Hypothesis 2: Sites with positive VTG assays will contain unique LC-HRMS features associated with estrogen agonist potency

The most significant findings of this work include:

- A suite of chemical components were at significantly higher abundance in grab and temporally integrated samples from the active location
- A greater total number of features were prioritized in temporally integrated samples, but the identity of many remain masses of interest

- More than half of confidently identified features were not currently included in the EDSP21 dashboard, and those which were present were not listed as endocrine active

5.1.3 Objective 3: Delineating OWTS-associated LC-HRMS features at an adjacent Minnesota lake

The aim of the study described in Chapter 4 was to conduct a suspect search of OWTS-associated TOrCs specific to the OWTS of a single Minnesota shoreline residence at nearby lake locations. A suspect list of wastewater chemicals typically pertains to prioritizing chemicals based on their widespread usage, persistence, or toxic properties. This prediction method is preferred for centralized treatment plants, but is less applicable to OWTSs which treat single family wastewater streams and are subject to microscale anthropogenic habits. Detection or a lack of detection of these chemicals, such as the targeted list used to address the research objectives of Chapter 2, is then difficult to interpret. Below is a discussion of how the findings of this study related to Hypothesis 3:

Hypothesis 3: A lake location closer to an OWTS will contain higher intensities of LC-HRMS TOrCs present in a shoreline residence's septic tank and drain field

The most significant findings of this work include:

- A suite of 63 human waste chemicals, hormonal steroids, flame retardants, pharmaceuticals, surfactants, and food and plant extracts were prioritized as OWTS-associated chemicals
- Higher intensities of DEET and losartan occurred at the lake location closer to the HOME
- OWTS-associated compounds still occurred at reference locations, suggesting additional prioritization of OWTS features should be integrated into this suspect search strategy
- Performance assessment of passive samplers indicated some biases introduced by this sampling technique, but we find them overall suitable for monitoring TOrCs in an OWTS

5.2 Overall Significance and Broader Implications

Known as the land of 10,000 lakes, lake water quality and consequent ecological health are a major concern of Minnesotans. These water bodies provide habitat for a variety of wildlife and vegetation, and act as centers for natural enjoyment highly valued by the state's residents and visitors. Fishing and boating are two popular recreational activities supported by these systems, as evidenced by public and private docks and boating entrances along the shoreline of many of these lakes and public programs to stock and monitor fish populations. The Minnesota Department of Natural Resources (MNDNR) maintains an active log of lake water quality throughout the state, which is supported by monitoring efforts conducted by employees and volunteer citizen scientists. Reports issued by the MNDNR indicate the agency is actively working to integrate non-point sources and contaminants of emerging concern into their lake preservation strategies. Developing strategies to understand and mange threats to water quality from septic systems and unknown components of endocrine active mixtures are two priorities listed by the MNDNR and Minnesota Pollution Control Agency (Engelking and Kovacevic, 2016; Robinson and Schultz, 2015). The first phase of this dissertation work (Chapter 2) demonstrated that residential proximity increased TOrC concentrations in lake water. Classical contaminants, such as nitrogen were still at higher concentrations in unpopulated reference locations, suggesting that other nonpoint sources also influence the lakes sampled. The MNDNR will need to establish methods for prioritizing the numerous potential sources and organic and inorganic components within lake contaminant mixtures, and aggregate multiple lines of evidence to prioritize mitigation or remediation efforts the most at-risk or degraded lakes.

Lake locations proximal to residences also corresponded to more pronounced estrogen agonism in resident fish. However, the targeted detected TOrCs at these locations are considered endocrine inactive when fish are exposed to them as individual contaminants. This suggests that either grab samples were not representative of exposure conditions or unexpected causative agents or unexpected co-contaminant interactions were present at these lake locations. The next stage of this work (Chapter 3) executed LC-HRMS on both grab and 12-hour temporally integrated samples to identify potentially causative agents omitted by targeted analysis. A basic univariate statistical method prioritized features with higher intensity at locations associated with endocrine activity. Multivariate analysis identified features shared by both active sampling locations as well as features unique to each location, potentially reflective of unique TOrC discharges associated with proximity to different households. While additional features were identified at higher intensities in these lakes, current metrics for assessing endocrine activity suggested these features were also inactive. Unidentified features remain within the prioritized list which could have potential endocrine active effects. In addition, the workflow did not prioritize contaminants in equal abundance at both endocrine active and reference sites. Co-contaminant interactions with these unextracted features, or amongst the prioritized features should also be considered in toxicology studies. Aggregate databases, such as the Endocrine Disruptor Screening Program in the 21st century (EDSP21) dashboard, should anticipate integrating co-contaminant information into their algorithms.

The source of several contaminants detected at higher concentrations more proximal to residences in Chapter 2 were associated with more outdoor TOrC applications. To evaluate the influence of OWTSs on TOrC composition proximal to residences at these lakes, an OWTS in the region was characterized using LC-HRMS (Chapter 4). A passive sampler previously only tested in laboratory calibration studies was implemented. The dimensions of this sampler made it compatible with temporary groundwater well monitoring for a wide range of TOrCs, though uptake was inconsistent over the sampling period. Only losartan and DEET were detected at higher intensities in the proximal lake location versus the reference location. Interestingly, 2,4-D which was associated with residential proximity at the lake locations sampled in Chapter 2 was absent form the sampled OWTS. While sampling occurred at an entirely different lake, this observation could also support that DEET is more likely to enter wastewater streams as its intended function is to be applied to humans, while 2,4-D is intended for direct application to invasive plant species on both residential plots and along roadsides. OWTS-associated chemicals, particularly surfactants were still identified at the reference site indicating that these TOrCs either have different sources of unexpected transport mechanisms. These polyethylene glycol chemicals were present at intensities in environmental samples at least five times greater than field and equipment blank sampler extracts, suggesting their presence was not an artifact of experimental methods. Overall, this study did not provide strong evidence of impact specifically from the sampled OWTS at the adjacent lake location. However, low spatial resolution in lake samples and uncertainty of groundwater flow paths from the drain field to the sampled lake location make this study insufficient to rule out OWTS impact on TOrC occurrence in Minnesota lakes.

5.3 Recommendations for Future Directions

Increasing spatial resolution of sampling between OWTSs and lakes. Spatial resolution in samples analyzed during this research effort were insufficient to address the many sources of uncertainty associated with these systems. Subsurface transport of TOrCs in variable shallow groundwater paths contributes to poor understanding of which lake locations are truly down-gradient of OWTSs. Furthermore, analysis of a single OWTS limits understanding of the many wastewater streams which may combine in lakewater systems. This research effort should be considered a preliminary analysis motivating field monitoring and aqueous analysis strategies before stronger conclusions can be made with regards to the impacts of OWTS-associated TOrCs on Minnesota lakes. The discrete size and ease with which passive samplers are deployed could encourage greater spatial resolution in sampling locations, but careful consideration should be taken to understand the bias introduced by the selectivity of these field monitoring tools. Field monitoring efforts should start with deployment of these samplers in multiple OWTSs from lakes across the state. Lake selection should, again focus on regions where shallow groundwater feeds into the lake, and bromide tracer experiments can be executed to prioritize OWTSs with wastewater plumes that flow from the drain field to shoreline locations. Older septic systems should also be prioritized to account for potential long horizontal flow paths within the subsurface. LC-HRMS analysis of these samples can encourage a broad understanding of the wastewater indicators unique to each household. Once identified, samplers deployed at nearby lake locations should be analyzed for these wastewater compounds.

Investigate uptake, metabolism, and toxicity for exposed fish with endocrine disruption. The chronic time scales associated with biological endpoints of endocrine disruption make analysis of transient aqueous samples insufficient for truly understanding causative agents. Stronger conclusions on the particular causative agents and potential synergistic or additive chemical interactions occurring at these lakes would require prolonged exposure and multigenerational studies representative of chronic toxicity effects. In particular, interactions between DEET, 2,4-D, 8-hydroxyquinoline, and several other notable detections from Chapter 3 should be investigated with laboratory exposure studies. Further prioritization could also occur by examining feature uptake prior to these exposure studies to understand the biologically available components present in the analyzed aqueous samples. Similar to effect directed analysis and wastewater exposure studies paired with LC-HRMS, the brain, blood, and bile of Minnesota lake fish demonstrating endocrine disrupting effects should be analyzed for TOrCs. Non-target feature prioritization could rely on unique chemical signatures in an endocrine disrupted set of fish versus a reference set of fish from Minnesota lakes which did not contain evidence of endocrine disruption. Alternatively, laboratory fish dosed with varying concentrations of Minnesota lakewater could be analyzed in a similar manner. Features could then be prioritized using multivariate analysis to identify TOrC parent compounds or metabolites, which increased in intensity with increased dosage or increased blood vitellogenin concentrations, provided that this assay is also included in the laboratory study. Results from this study could also prioritize masses of interest from Chapter 3 for more thorough identification. Feature identification without reference spectra is rather difficult and time intensive, typically requiring assistance from *in silico* fragmentation tools. Further prioritization will allow for efforts to be focused on features more likely to influence endocrine activity in Minnesota lake TOrC mixtures.

Evaluate TOrC transport from other diffuse sources. Sampled lake locations showed weak evidence regarding the presence of wastewater indicators. However, numerous TOrCs were still at higher concentrations or relative abundances at locations proximal to residences. Alternative diffuse transport pathways are suspected to contribute to TOrC occurrence, such as atmospheric deposition, surface water runoff, and recreational activity, particularly boating activity occurring at docks common amongst many residences in the region. Air, stormwater runoff, and shallow groundwater throughout the region should be sampled from multiple locations proximal and distal from shoreline residences. Stormwater is typically sampled with flow weighted composite auto-samplers, although passive samplers are now being developed and implemented for this particular monitoring application. Deploying various passive samplers to sample each environmental compartment would allow for higher spatial resolution and encourage understanding of portioning behaviors and transport mechanisms for these contaminants. Septic density was also expected to influence TOrC loadings to lakes in Chapter 2. Study areas should focus on regions with multiple houses proximal to one another, as well as reference sites, to encourage prioritization of at-risk and likely endocrine active lakeshore locations for further analysis regarding endocrine active mixtures.

REFERENCES

ABS Materials, 2016. SENTINEL: Passive Sampler & Monitoring Tool.

Al-Salhi, R., Abdul-Sada, A., Lange, A., Tyler, C.R., Hill, E.M., 2012. The xenometabolome and novel contaminant markers in fish exposed to a wastewater treatment works effluent. Environ. Sci. Technol. 46, 9080–9088. doi:10.1021/es3014453

Arditsoglou, A., Voutsa, D., 2008. Passive sampling of selected endocrine disrupting compounds using polar organic chemical integrative samplers. Environ. Pollut. 156, 316–324. doi:10.1016/j.envpol.2008.02.007

Arnhard, K., Gottschall, A., Pitterl, F., Oberacher, H., 2015. Applying 'Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra' (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry. Anal. Bioanal. Chem. 407, 405–414. doi:10.1007/s00216-014-8262-1

Baker, B.H., Martinovic-Weigelt, D., Ferrey, M., Barber, L.B., Writer, J.H., Rosenberry, D.O., Kiesling, R.L., Lundy, J.R., Schoenfuss, H.L., 2014. Identifying Non-point Sources of Endocrine Active Compounds and Their Biological Impacts in Freshwater Lakes. Arch. Environ. Contam. Toxicol. 67, 374–388. doi:10.1007/s00244-014-0052-4

Barber, L.B., Vajda, A.M., Douville, C., Norris, D.O., Writer, J.H., 2012. Fish Endocrine Disruption Responses to a Major Wastewater Treatment Facility Upgrade.

Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. Sci. Total Environ. 575, 265–275. doi:10.1016/J.SCITOTENV.2016.09.135

Borgert, C.J., Stuchal, L.D., Mihaich, E.M., Becker, R.A., Bentley, K.S., Brausch, J.M., Coady, K., Geter, D.R., Gordon, E., Guiney, P.D., Hess, F., Holmes, C.M., LeBaron, M.J., Levine, S., Marty, S., Mukhi, S., Neal, B.H., Ortego, L.S., Saltmiras, D.A., Snajdr, S., Staveley, J., Tobia, A., 2014. Relevance Weighting of Tier 1 Endocrine Screening Endpoints by Rank Order. Birth Defects Res. Part B Dev. Reprod. Toxicol. 101, 90–113. doi:10.1002/bdrb.21096

Brack, W., Ait-Aissa, S., Burgess, R.M., Busch, W., Creusot, N., Di Paolo, C., Escher, B.I., Mark Hewitt, L., Hilscherova, K., Hollender, J., Hollert, H., Jonker, W., Kool, J., Lamoree, M., Muschket, M., Neumann, S., Rostkowski, P., Ruttkies, C., Schollee, J., Schymanski, E.L., Schulze, T., Seiler, T.-B., Tindall, A.J., De Aragão Umbuzeiro, G., Vrana, B., Krauss, M., 2016. Effect-directed analysis supporting monitoring of aquatic environments — An in-depth overview. Sci. Total Environ. 544, 1073–1118. doi:10.1016/j.scitotenv.2015.11.102

Brian, J. V., Harris, C.A., Scholze, M., Kortenkamp, A., Booy, P., Lamoree, M., Pojana, G., Jonkers, N., Marcomini, A., Sumpter, J.P., 2007. Evidence of estrogenic mixture effects on the reproductive performance of fish. Environ. Sci. Technol. 41, 337–344. doi:10.1021/es0617439

Brian, J. V, Harris, C.A., Scholze, M., Backhaus, T., Booy, P., Lamoree, M., Pojana, G., Jonkers, N., Runnalls, T., Bonfà, A., Marcomini, A., Sumpter, J.P., 2005. Accurate prediction of the response of freshwater fish to a mixture of estrogenic chemicals. Environ. Health Perspect.

113, 721–8. doi:10.1289/EHP.7598

Carrara, C., Robertson, W.D., Blowes, D.W., 2008. Fate of Pharmaceutical and Trace Organic Compounds in Three Septic System Plumes , Ontario , Canada 42, 2805–2811.

Colby, J.M., Rivera, J., Burton, L., Cox, D., Lynch, K.L., 2017. Improvement of drug identification in urine by LC-QqTOF using a probability-based library search algorithm. Clin. Mass Spectrom. 3, 7–12. doi:10.1016/j.clinms.2017.04.001

Conn, K.E., Lowe, K.S., Drewes, J.E., Hoppe-Jones, C., Tucholke, M.B., 2010a. Occurrence of Pharmaceuticals and Consumer Product Chemicals in Raw Wastewater and Septic Tank Effluent from Single-Family Homes. Environ. Eng. Sci. 27, 347–356. doi:10.1089/ees.2009.0364

Conn, K.E., Siegrist, R.L., Barber, L.B., Meyer, M.T., 2010b. Fate of trace organic compounds during vadose zone soil treatment in an onsite wastewater system. Environ. Toxicol. Chem. 29, 285–293. doi:10.1002/etc.40

Cornelissen, G., Pettersen, A., Broman, D., Mayer, P., Breedveld, G.D., 2008. Field testing of equilibrium passive samplers to determine freely dissolved native polycyclic aromatic hydrocarbon concentrations. Environ. Toxicol. Chem. 27, 499–508. doi:10.1897/07-253.1

Del Rosario, K.L., Mitra, S., Humphrey, C.P., O'Driscoll, M.A., 2014. Detection of pharmaceuticals and other personal care products in groundwater beneath and adjacent to onsite wastewater treatment systems in a coastal plain shallow aquifer. Sci. Total Environ. 487, 216–223. doi:10.1016/j.scitotenv.2014.03.135

Desbrow, C., Routledge, E.J., Brighty, G.C., Sumpter, J.P., Waldock, M., 1998. Identification of estrogenic chemicals in STW effluent. 1. Chemical fractionation and in vitro biological screening. Environ. Sci. Technol. 32, 1549–1558. doi:10.1021/es9707973

Devlin, A.S., Fischbach, M.A., 2015. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. Nat. Chem. Biol. 11, 685–690. doi:10.1038/nchembio.1864

Drugs.com, 2019. Loxoprofen [WWW Document]. URL https://www.drugs.com/international/loxoprofen.html (accessed 1.14.19).

Du, B., Lofton, J.M., Peter, K.T., Gipe, A.D., James, C.A., McIntyre, J.K., Scholz, N.L., Baker, J.E., Kolodziej, E.P., 2017a. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. Environ. Sci. Process. Impacts 19, 1185–1196. doi:10.1039/C7EM00243B

Du, B., Lofton, J.M., Peter, K.T., Gipe, A.D., James, C.A., McIntyre, J.K., Scholz, N.L., Baker, J.E., Kolodziej, E.P., 2017b. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. Environ. Sci. Process. Impacts 1185–1196. doi:10.1039/C7EM00243B

Dubey, S.K., Saha, R.N., Jangala, H., Pasha, S., 2013. Rapid sensitive validated UPLC–MS method for determination of venlafaxine and its metabolite in rat plasma: Application to pharmacokinetic study. J. Pharm. Anal. 3, 466–471. doi:10.1016/J.JPHA.2013.05.002

Ekman, J.C., 2002. Geologic Sensitivity to Pollution of Near-Surface Groundwater: Regional Hydrogeologic Assessment for Ottertail Area, West Central Minnesota.

Engelking, P., Kovacevic, A., 2016. 2016 Pollution Report to the Legislature.

Environmental Sampling Technologies, 2018. Environmental Sampling Technologies - The Leader in Polar Organic Chemical Integrative Samplers (POCIS) [WWW Document]. URL http://www.est-lab.com/pocis.php (accessed 12.21.18).

EPA, 2014. TSCA Work Plan for Chemical Assessments: 2014 Update.

Erickson, M.L., Langer, S.K., Roth, J.L., Kroening, S.E., 2014. Scientific Investigations Report 2014–5096 Contaminants of Emerging Concern in Ambient Groundwater in Urbanized Areas of Minnesota.

Fatta-Kassinos, D., Meric, S., Nikolaou, A., 2011. Pharmaceutical residues in environmental waters and wastewater: current state of knowledge and future research. Anal. Bioanal. Chem. 399, 251–275. doi:10.1007/s00216-010-4300-9

Fauvelle, V., Mazzella, N., Belles, A., Moreira, A., Allan, I.J., Budzinski, H., 2014. Optimization of the polar organic chemical integrative sampler for the sampling of acidic and polar herbicides. Anal. Bioanal. Chem. 406, 3191–3199. doi:10.1007/s00216-014-7757-0

Fernández, D., Vermeirssen, E.L.M., Bandow, N., Muñoz, K., Schäfer, R.B., 2014. Calibration and field application of passive sampling for episodic exposure to polar organic pesticides in streams. Environ. Pollut. 194, 196–202. doi:10.1016/J.ENVPOL.2014.08.001

Ferrey, M.L., Heiskary, S., Grace, R., Hamilton, M.C., Lueck, A., 2015. Pharmaceuticals and other anthropogenic tracers in surface water: A randomized survey of 50 Minnesota lakes. Environ. Toxicol. Chem. 34, 2475–2488. doi:10.1002/etc.3125

Filer, D., Patisaul, H.B., Schug, T., Reif, D., Thayer, K., 2014. Test driving ToxCast: endocrine profiling for 1858 chemicals included in phase II. Curr. Opin. Pharmacol. 19, 145–152. doi:10.1016/J.COPH.2014.09.021

Fovet, O., Dupas, R., Gruau, G., Jeanneau, L., Jaffrézic, A., Ruiz, L., Humbert, G., Denis, M., Gu, S., Strohmenger, L., Odoux, C.G., 2018. Groundwater storage as a major control of seasonal stream water quality dynamics during both base and storm flows 20, EGU2018-6532.

Gago-Ferrero, P., Gros, M., Ahrens, L., Wiberg, K., 2017. Impact of on-site, small and large scale wastewater treatment facilities on levels and fate of pharmaceuticals, personal care products, artificial sweeteners, pesticides, and perfluoroalkyl substances in recipient waters. Sci. Total Environ. 601–602, 1289–1297. doi:10.1016/j.scitotenv.2017.05.258

Gago-Ferrero, P., Krettek, A., Fischer, S., Wiberg, K., Ahrens, L., 2018. Suspect screening and regulatory databases: A powerful combination to identify emerging micropollutants. Environ. Sci. Technol. acs.est.7b06598. doi:10.1021/acs.est.7b06598

Gerrity, D., Trenholm, R.A., Snyder, S.A., 2011. Temporal variability of pharmaceuticals and illicit drugs in wastewater and the effects of a major sporting event. Water Res. 45, 5399–5411. doi:10.1016/J.WATRES.2011.07.020

Ghosh, U., Kane Driscoll, S., Burgess, R.M., Jonker, M.T.O., Reible, D., Gobas, F., Choi, Y., Apitz, S.E., Maruya, K.A., Gala, W.R., Mortimer, M., Beegan, C., 2014. Passive sampling methods for contaminated sediments: practical guidance for selection, calibration, and implementation. Integr. Environ. Assess. Manag. 10, 210–23. doi:10.1002/ieam.1507

Godfrey, E., Woessner, W.W., Benotti, M.J., 2007. Pharmaceuticals in On-Site Sewage Effluent and Ground Water , Western Montana 45, 263–271. doi:10.1111/j.1745-6584.2006.00288.x

Gros, M., Blum, K.M., Jernstedt, H., Renman, G., Rodríguez-Mozaz, S., Haglund, P., Andersson, P.L., Wiberg, K., Ahrens, L., 2017. Screening and prioritization of micropollutants in wastewaters from on-site sewage treatment facilities. J. Hazard. Mater. 328, 37–45. doi:10.1016/J.JHAZMAT.2016.12.055

Guyader, M.E., Warren, L.D., Green, E., Bertram, R., Proudian, A.P., Kiesling, R.L., Schoenfuss, H.L., Higgins, C.P., 2018a. Trace organic contaminant (TOrC) mixtures in Minnesota littoral zones: Effects of on-site wastewater treatment system (OWTS) proximity and biological impact. Sci. Total Environ. 626, 1157–1166. doi:10.1016/J.SCITOTENV.2018.01.123

Guyader, M.E., Warren, L.D., Green, E., Butt, C., Ivosev, G., Kiesling, R.L., Schoenfuss, H.L., Higgins, C.P., 2018b. Exploitation of Toxicological Endpoints for the Prioritization of Liquid Chromatography Tandem High Resolution Mass Spectrometry (LC-HRMS) Features: Identifying Potential Endocrine Active Chemicals in Lakewater. Prep.

Hansen, P.-D., Dizer, H., Hock, B., Marx, A., Sherry, J., McMaster, M., Blaise, C., 1998. Vitellogenin – a biomarker for endocrine disruptors. TrAC Trends Anal. Chem. 17, 448–451. doi:10.1016/S0165-9936(98)00020-X

Harman, C., Allan, I.J., B, P.S., 2011. The PRC and the POCIS.

Hayes, T.B., Anderson, L.L., Beasley, V.R., de Solla, S.R., Iguchi, T., Ingraham, H., Kestemont, P., Kniewald, J., Kniewald, Z., Langlois, V.S., Luque, E.H., McCoy, K.A., Muñoz-de-Toro, M., Oka, T., Oliveira, C.A., Orton, F., Ruby, S., Suzawa, M., Tavera-Mendoza, L.E., Trudeau, V.L., Victor-Costa, A.B., Willingham, E., 2011. Demasculinization and feminization of male gonads by atrazine: consistent effects across vertebrate classes. J. Steroid Biochem. Mol. Biol. 127, 64–73. doi:10.1016/j.jsbmb.2011.03.015

Helsel, D.R., 2012. Statistics for Censored Environmental Data Using Minitab and R. John Wiley & Sons, Inc., Hoboken, NJ.

Hernández, A.F., Parrón, T., Requena, M., Alarcón, R., López-Guarnido, O., 2013. Toxic effects of pesticide mixtures at a molecular level: Their relevance to human health. Toxicology 307, 136–145. doi:10.1016/J.TOX.2012.06.009

Hernández, F., Ibáñez, M., Portolés, T., Cervera, M.I., Sancho, J. V., López, F.J., 2015. Advancing towards universal screening for organic pollutants in waters. J. Hazard. Mater. 282, 86–95. doi:10.1016/J.JHAZMAT.2014.08.006

Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014a. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. Environ. Pollut. 184, 25–32. doi:10.1016/J.ENVPOL.2013.07.048

Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014b. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. Environ. Pollut. 184, 25–32. doi:10.1016/J.ENVPOL.2013.07.048

Ivosev, G., Burton, L., Bonner, R., 2008. Dimensionality Reduction and Visualization in Principal Component Analysis. Anal. Chem. 80, 4933–4944. doi:10.1021/ac800110w

Jiang, P., Loyau, S., Tchitchinadze, M., Ropers, J., Jondeau, G., Jandrot-Perrus, M., 2015. Inhibition of Glycoprotein VI Clustering by Collagen as a Mechanism of Inhibiting Collagen-Induced Platelet Responses: The Example of Losartan. PLoS One 10, e0128744. doi:10.1371/journal.pone.0128744

Jobling, S., Nolan, M., Tyler, C.R., Brightly, G., Sumpter, J.P., 1998. Widespread Sexual Disruption in Wild Fish. Environ. Sci. Technol.

Kawanishi, M., Takamura-Enya, T., Ermawati, R., Shimohara, C., Sakamoto, M., Matsukawa, K., Matsuda, T., Murahashi, T., Matsui, S., Wakabayashi, K., Watanabe, T., Tashiro, Y., Yagi, T., 2004. Detection of genistein as an estrogenic contaminant of river water in Osaka. Environ. Sci. Technol. 38, 6424–6429. doi:10.1021/es049764v

Kim, H.-J., Park, Y.I., Dong, M.-S., 2005. Effects of 2,4-D and DCP on the DHT-Induced Androgenic Action in Human Prostate Cancer Cells. Toxicol. Sci. 88, 52–59. doi:10.1093/toxsci/kfi287

Kim, Y.-I., Hirai, S., Takahashi, H., Goto, T., Ohyane, C., Tsugane, T., Konishi, C., Fujii, T., Inai, S., Iijima, Y., Aoki, K., Shibata, D., Takahashi, N., Kawada, T., 2011. 9-oxo-10(E),12(E)-octadecadienoic acid derived from tomato is a potent PPAR α agonist to decrease triglyceride accumulation in mouse primary hepatocytes. Mol. Nutr. Food Res. 55, 585–593. doi:10.1002/mnfr.201000264

Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. doi:10.1186/1471-2105-8-105

Kot-wasik, A., 2010. Passive sampling as a tool for obtaining reliable analytical information in environmental quality monitoring 273–296. doi:10.1007/s00216-009-3244-4

Krauss, M., Singer, H., Hollender, J., 2010. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. Anal. Bioanal. Chem. 397, 943–951. doi:10.1007/s00216-010-3608-9

Lambert, J.P., Ivosev, G., Couzens, A.L., Larsen, B., Taipale, M., Lin, Z.Y., Zhong, Q., Lindquist, S., Vidal, M., Aebersold, R., Pawson, T., Bonner, R., Tate, S., Gingras, A.C., 2013. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. Nat. Methods. doi:10.1038/nmeth.2702

Lapworth, D.J., Baran, N., Stuart, M.E., Ward, R.S., 2012. Emerging organic contaminants in groundwater: A review of sources, fate and occurrence. Environ. Pollut. 163, 287–303. doi:10.1016/j.envpol.2011.12.034

Li, H., Helm, P. a., Metcalfe, C.D., 2010. Sampling in the great lakes for pharmaceuticals, personal care products, and endocrine-disrupting substances using the passive polar organic chemical integrative sampler. Environ. Toxicol. Chem. 29, 751–762. doi:10.1002/etc.104

Liu, R., Zhou, J.., Wilding, A., 2004. Simultaneous determination of endocrine disrupting phenolic compounds and steroids in water by solid-phase extraction–gas chromatography–mass spectrometry. J. Chromatogr. A 1022, 179–189. doi:10.1016/j.chroma.2003.09.035

Loos, M., 2016. Package "nontarget" Type Package Title Detecting Isotope, Adduct and Homologue Relations in LC-MS Data.

Mazzella, N., Dubernet, J.-F., Delmas, F., 2007. Determination of kinetic and equilibrium regimes in the operation of polar organic chemical integrative samplers. Application to the passive sampling of the polar herbicides in aquatic environments. J. Chromatogr. A 1154, 42–51. doi:10.1016/j.chroma.2007.03.087

McDonough, C.A., Khairy, M.A., Muir, D.C.G., Lohmann, R., 2014. Significance of Population Centers As Sources of Gaseous and Dissolved PAHs in the Lower Great Lakes. Environ. Sci. Technol. 48, 7789–7797. doi:10.1021/es501074r

McEachran, A.D., Sobus, J.R., Williams, A.J., 2017. Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard. Anal. Bioanal. Chem. 409, 1729–1735. doi:10.1007/s00216-016-0139-z

Morrison, S.A., Belden, J.B., 2016. Calibration of nylon organic chemical integrative samplers and sentinel samplers for quantitative measurement of pulsed aquatic exposures. J. Chromatogr. A 1449, 109–117. doi:10.1016/j.chroma.2016.04.072

Moschet, C., Vermeirssen, E.L.M., Singer, H., Stamm, C., Hollender, J., 2015. Evaluation of in-situ calibration of Chemcatcher passive samplers for 322 micropollutants in agricultural and urban affected rivers. Water Res. 71, 306–17. doi:10.1016/j.watres.2014.12.043

Nürenberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on liquid chromatography – high resolution mass spectrometry analysis for the evaluation of different wastewater treatment options. J. Chromatogr. A 1426, 77–90. doi:10.1016/J.CHROMA.2015.11.014

Oetjen, K., Chan, K.E., Gulmark, K., Christensen, J.H., Blotevogel, J., Borch, T., Spear, J.R., Cath, T.Y., Higgins, C.P., 2018. Temporal characterization and statistical analysis of flowback and produced waters and their potential for reuse. Sci. Total Environ. 619–620, 654–664. doi:10.1016/J.SCITOTENV.2017.11.078

Oppenheimer, J., Eaton, A., Badruzzaman, M., Haghani, A.W., Jacangelo, J.G., 2011. Occurrence and suitability of sucralose as an indicator compound of wastewater loading to surface waters in urbanized regions. doi:10.1016/j.watres.2011.05.014

Ort, C., Lawrence, M.G., Rieckermann, J., Joss, A., 2010. Sampling for Pharmaceuticals and Personal Care Products (PPCPs) and Illicit Drugs in Wastewater Systems: Are Your Conclusions Valid? A Critical Review. Environ. Sci. Technol. 44, 6024–6035. doi:10.1021/es100779n

Persson, L., Alsberg, T., Kiss, G., Odham, G., 2000. On-line size-exclusion chromatography/electrospray ionisation mass spectrometry of aquatic humic and fulvic acids. Rapid Commun. Mass Spectrom. 14, 286–292. doi:10.1002/(SICI)1097-0231(20000229)14:4<286::AID-RCM879>3.0.CO;2-C

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol. acs.est.8b03287. doi:10.1021/acs.est.8b03287

Piccolo, A., Spiteller, M., 2003. Electrospray ionization mass spectrometry of terrestrial humic substances and their size fractions. Anal. Bioanal. Chem. 377, 1047–1059. doi:10.1007/s00216-003-2186-5

Prachayasittikul, V., Pingaew, R., Nantasenamat, C., Prachayasittikul, S., Ruchirawat, S., Prachayasittikul, V., 2014. Investigation of aromatase inhibitory activity of metal complexes of 8-hydroxyquinoline and uracil derivatives. Drug Des. Devel. Ther. 8, 1089–96. doi:10.2147/DDDT.S67300

Püssa, T., Raudsepp, P., Toomik, P., Pällin, R., Mäeorg, U., Kuusik, S., Soidla, R., Rei, M., 2009. A study of oxidation products of free polyunsaturated fatty acids in mechanically deboned meat. J. Food Compos. Anal. 22, 307–314. doi:10.1016/J.JFCA.2009.01.014

Robinson, C., Schultz, P., 2015. 2015 SSTS Annual Report Subsurface Sewage Treatment Systems in Minnesota Minnesota Pollution Control Agency Photo credit.

Roemmelt, A.T., Steuer, A.E., Poetzsch, M., Kraemer, T., 2014. Liquid Chromatography, in Combination with a Quadrupole Time-of-Flight Instrument (LC QTOF), with Sequential Window Acquisition of All Theoretical Fragment-Ion Spectra (SWATH) Acquisition: Systematic Studies on Its Use for Screenings in Clinical and Forensic Toxicology and Comparison with Information-Dependent Acquisition (IDA). Anal. Chem. 86, 11742–11749. doi:10.1021/ac503144p

Rotander, A., Kärrman, A., Toms, L.M.L., Kay, M., Mueller, J.F., Gómez Ramos, M.J., 2015. Novel fluorinated surfactants tentatively identified in firefighters using liquid chromatography quadrupole time-of-flight tandem mass spectrometry and a case-control approach. Environ. Sci. Technol. 49, 2434–2442. doi:10.1021/es503653n

Routledge, E.J., Sumpter, J.P., 1996. Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. Environ. Toxicol. Chem. 15, 241–248. doi:10.1002/etc.5620150303

Rstudio Team, 2015. RStudio: Integrated Development for R.

Schaider, L.A., Rodgers, K.M., Rudel, R.A., 2017. Review of Organic Wastewater Compound Concentrations and Removal in Onsite Wastewater Treatment Systems. Environ. Sci. Technol. 51, 7304–7317. doi:10.1021/acs.est.6b04778

Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J., 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. Anal. Chem. 87, 12121–12129. doi:10.1021/acs.analchem.5b02905

Schreiber, A., Cox, D., Pace, N., Borton, C., 2010. Metabolomic Profiling of Accurate Mass LC-MS / MS Data to Identify Unexpected Environmental Pollutants 1–5.

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014a.

Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ. Sci. Technol. 48, 2097–2098. doi:10.1021/es5002105

Schymanski, E.L., Singer, H.P., Longreé, P., Loos, M., Ruff, M., Stravs, M.A., Ripolles Vidalles, C., Hollender, J., 2014b. SOP for Solid Phase Extraction with multilayered Eawag- Cartridge type Silovo 6–7.

Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I.M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., Thomaidis, N.S., Bletsou, A., Zwiener, C., Ibáñez, M., Portolés, T., de Boer, R., Reid, M.J., Onghena, M., Kunkel, U., Schulz, W., Guillon, A., Noyon, N., Leroy, G., Bados, P., Bogialli, S., Stipaničev, D., Rostkowski, P., Hollender, J., 2015a. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal. Bioanal. Chem. 407, 6237–6255. doi:10.1007/s00216-015-8681-7

Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I.M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., Thomaidis, N.S., Bletsou, A., Zwiener, C., Ibáñez, M., Portolés, T., de Boer, R., Reid, M.J., Onghena, M., Kunkel, U., Schulz, W., Guillon, A., Noyon, N., Leroy, G., Bados, P., Bogialli, S., Stipaničev, D., Rostkowski, P., Hollender, J., 2015b. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal. Bioanal. Chem. 407, 6237–6255. doi:10.1007/s00216-015-8681-7

Schymanski, E.L., Williams, A.J., 2017. Open Science for Identifying “Known Unknown” Chemicals. Environ. Sci. Technol. 5357–5359. doi:10.1021/acs.est.7b01908

Solecki, R., Kortenkamp, A., Bergman, Å., Chahoud, I., Degen, G.H., Dietrich, D., Greim, H., Håkansson, H., Hass, U., Husoy, T., Jacobs, M., Jobling, S., Mantovani, A., Marx-Stoelting, P., Piersma, A., Ritz, V., Slama, R., Stahlmann, R., van den Berg, M., Zoeller, R.T., Boobis, A.R., 2017. Scientific principles for the identification of endocrine-disrupting chemicals: a consensus statement. Arch. Toxicol. 91, 1001–1006. doi:10.1007/s00204-016-1866-9

Song, F., 2011. "Cross-Talk" in Scheduled Multiple Reaction Monitoring Caused by In-Source Fragmentation in Herbicide Screening with Liquid Chromatography Electrospray Tandem Mass Spectrometry. J. Agric. Food Chem 59, 4361–4364. doi:10.1021/jf200592n

Sumpter, J.P., Jobling, S., 1995. Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ. Health Perspect. 103 Suppl 7, 173–8.

Sumpter, J.P., Johnson, A.C., 2008a. 10th Anniversary Perspective: Reflections on endocrine disruption in the aquatic environment: From known knowns to unknown unknowns (and many things in between). J. Environ. Monit. 10, 1476–1485. doi:10.1039/b815741n

Sumpter, J.P., Johnson, A.C., 2008b. 10th Anniversary Perspective: Reflections on endocrine disruption in the aquatic environment: from known knowns to unknown unknowns (and many things in between). J. Environ. Monit. 10, 1476. doi:10.1039/b815741n

Sun, Q., Deng, S., Huang, J., Shen, G., Yu, G., 2008. Contributors to estrogenic activity in wastewater from a large wastewater treatment plant in Beijing, China. Environ. Toxicol. Pharmacol. 25, 20–26. doi:10.1016/J.ETAP.2007.08.003

Teerlink, J., Hering, A.S., Higgins, C.P., Drewes, J.E., 2012. Variability of trace organic chemical concentrations in raw wastewater at three distinct sewershed scales. Water Res. 46, 3261–3271. doi:10.1016/j.watres.2012.03.018

Thorpe, K.L., Cummings, R.I., Hutchinson, T.H., Scholze, M., Brighty, G., Sumpter, J.P., Tyler, C.R., 2003. Relative potencies and combination effects of steroidal estrogens in fish. Environ. Sci. Technol. 37, 1142–1149. doi:10.1021/es0201348

Togola, A., Budzinski, H., 2007. Development of polar organic integrative samplers for analysis of pharmaceuticals in aquatic systems. Anal. Chem. 79, 6734–6741. doi:10.1021/ac070559i

U.S. Environmental Protection Agency, 2017. Endocrine Disruptor Screening Program for the 21st Century [WWW Document]. URL https://actor.epa.gov/edsp21/ (accessed 11.1.18).

U.S. EPA, 2011. Endocrine Disruptor Screening Program for the 21 st Century: (EDSP21 Work Plan) The Incorporation of In Silico Models and In Vitro High Throughput Assays in the Endocrine Disruptor Screening Program (EDSP) for Prioritization and Screening Summary Overview A Part of the EDSP Comprehensive Management Plan.

Vajda, A.M., Barber, L.B., Gray, J.L., Lopez, E.M., Woodling, J.D., Norris, D.O., 2008. Reproductive Disruption in Fish Downstream from an Estrogenic Wastewater Effluent. Environ. Sci. Technol. 42, 3407–3414. doi:10.1021/es0720661

Vajda, A.M., Norris, D.O., 2011. Endocrine-active Chemicals (EACs) in Fishes 1, 245–264.

Vetter, W., Kaserzon, S., Gallen, C., Knoll, S., Gallen, M., Hauler, C., Mueller, J.F., 2018. Occurrence and concentrations of halogenated natural products derived from seven years of passive water sampling (2007–2013) at Normanby Island, Great Barrier Reef, Australia. Mar. Pollut. Bull. 137, 81–90. doi:10.1016/J.MARPOLBUL.2018.09.032

Wode, F., Van Baar, P., Dü Nnbier, U., Hecht, F., Taute, T., Jekel, M., Reemtsma, T., 2015. Search for over 2000 current and legacy micropollutants on a wastewater infiltration site with a UPLC-high resolution MS target screening method. Water Res. 69, 274–283. doi:10.1016/j.watres.2014.11.034

Wolf, S., Schmidt, S., Müller-Hannemann, M., Neumann, S., 2010. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 11, 148. doi:10.1186/1471-2105-11-148

Woodward, S.J.R., Wöhling, T., Stenger, R., 2016. Uncertainty in the modelling of spatial and temporal patterns of shallow groundwater flow paths: The role of geological and hydrological site information. J. Hydrol. 534, 680–694. doi:10.1016/J.JHYDROL.2016.01.045

Writer, J.H., Barber, L.B., Brown, G.K., Taylor, H.E., Kiesling, R.L., Ferrey, M.L., Jahns, N.D., Bartell, S.E., Schoenfuss, H.L., 2010. Science of the Total Environment Anthropogenic tracers , endocrine disrupting chemicals , and endocrine disruption in Minnesota lakes. Sci. Total Environ. 409, 100–111. doi:10.1016/j.scitotenv.2010.07.018

Zocatelli, R., Lavrieux, M., Guillemot, T., Chassiot, L., Le Milbeau, C., Jacob, J., 2017. Fecal biomarker imprints as indicators of past human land uses: Source distinction and preservation potential in archaeological and natural archives. J. Archaeol. Sci. 81, 79–89. doi:10.1016/J.JAS.2017.03.010

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

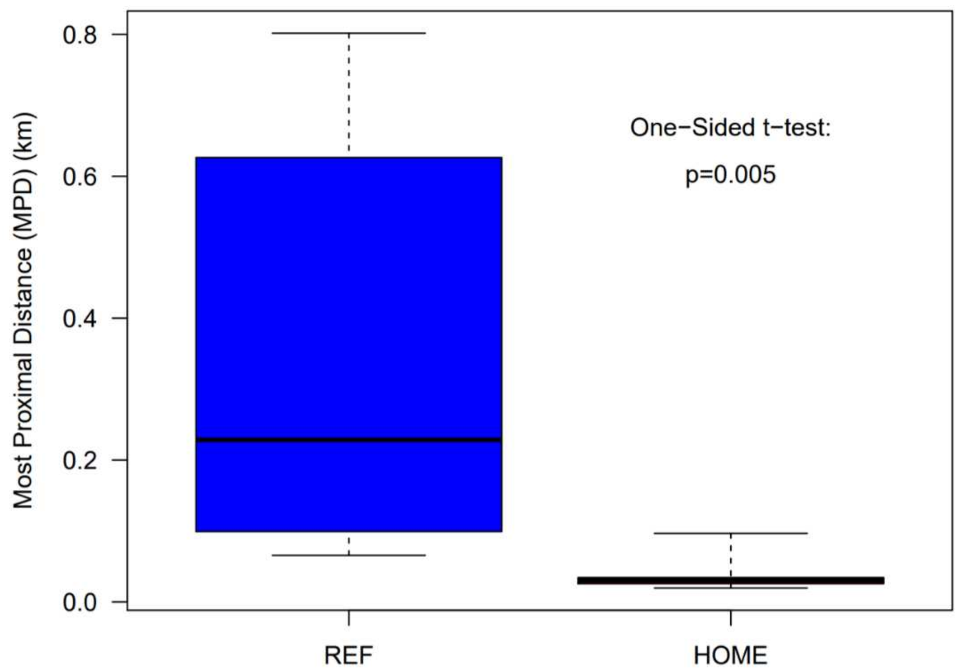

Figure A.1 Boxplot assessing distribution of most proximal distance to residences (MPDs) in HOME and REF site groupings.

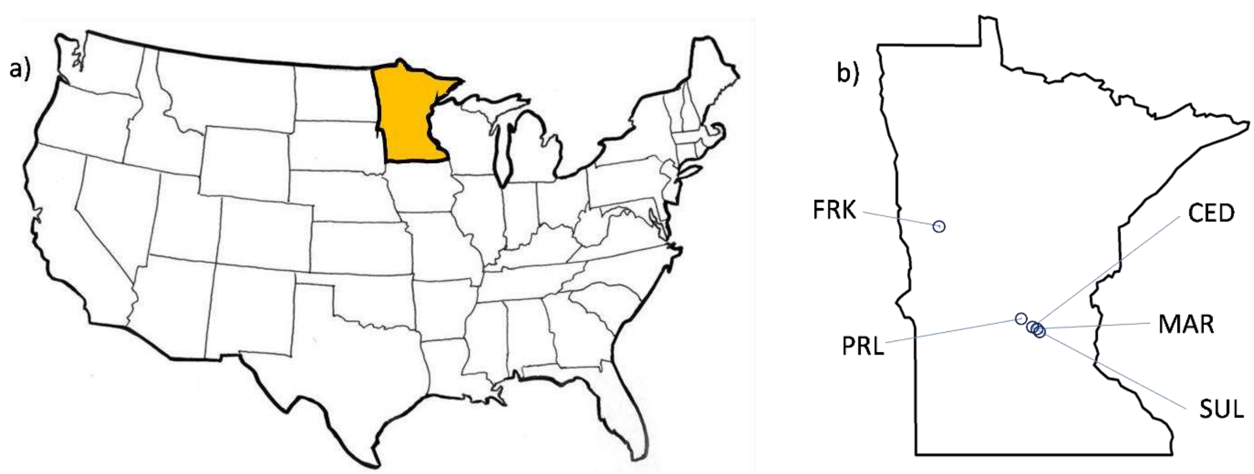

Figure A.2 Map illustrating the location of a) Minnesota (highlighted in yellow) in the United States and b) the lakes sampled within Minnesota. Lake labels are in alignment with abbreviations used throughout the manuscript: CED= Cedar Lake, MAR= Lake Mary, SUL= Sullivan Lake, FRK= Franklin Lake, PRL= Pearl Lake

Table A.1 Average matrix spike recoveries and surrogate recoveries for methods of TOrC sample preparation and analysis. Compound usages include the abbreviations are used for pharmaceuticals (PhACs) and personal care products (CPs). Known endocrine disruptors indicated with * following the analyte's name. Surrogate recoveries were averaged for all porewater grab samples across all lakes (n = 60). The number of samples used to calculate average matrix spike recovery were controlled by the number of valid matrix spike recovery samples. Reporting limits (RLs) set as either the average concentration in field and laboratory blanks (n=10) plus three times the standard deviation or just three times the measured concentration in a single blank. Analytes without detections in blanks only have a set LOQ.

Analyte	Compound Use	Isotope Standard	LC-MS/MS Method	Average Matrix Spike Recovery (%)	Average Surrogate Recovery (%)
4-n-nonylphenol*	Detergent	4-nonylphenol-d4	ESI -	85 +/- 3 (n=5)	41 +/- 10
4-tert-octylphenol*	Detergent	4-n-octylphenol-d17	ESI -	100 +/- 26 (n=5)	73 +/- 15
17α-ethinylestradiol*	Steroidal Hormone	17α-ethinylestradiol-d4	ESI -	86 +/- 12 (n=5)	88 +/- 26
2,4-D	Herbicide	2,4-D-d3	ESI -	128 +/- 21 (n=5)	71 +/- 24
Amitriptyline	PhAC - antidepressant	Amitrityline	ESI +	88 +/- 3 (n=5)	44 +/- 12
Androstenedione*	Steroidal Hormone	Testosterone-d2	ESI +	106 +/- 5 (n=5)	28 +/- 19
Atenolol	PhAC - beta blocker	Atenolol-d7	ESI +	78 +/- 5 (n=5)	74 +/- 21
Bisphenol A*	PCP - plasticizer	Bisphenol A-d16	ESI -	85 +/- 9 (n=3)	90 +/- 35
Caffeine	PCP - stimulant	Caffeine-d9	ESI +	128 +/- 76 (n=4)	35 +/- 22
Carbamazepine	PhAC - antiepileptic	Carbamazepine-d10	ESI +	104 +/- 1 (n=3)	36 +/- 27
Cimetidine	PhAC - histamine blocker	Cimetidine-d3	ESI +	89 +/- 5 (n=5)	25 +/- 17
DEET	PCP - insecticide	DEET-d4	ESI +	116 +/- 8 (n=2)	39 +/- 29
Diazepam	PhAC - muscle relaxant	Diazepam-d5	ESI +	90 +/- 5 (n=5)	37 +/- 21
Dilantin	PhAC - antiepileptic	Dilantin-d10	ESI +	101 +/- 9 (n=5)	38 +/- 27
Diuron	Herbicide	Diuron-d6	ESI +	84 +/- 6 (n=5)	33 +/- 24
Estriol*	Steroidal Hormone	Estriol-d2	ESI -	103 +/- 6 (n=5)	78 +/- 28
Estrone*	Steroidal Hormone	Estrone-d4	ESI -	80 +/- 3 (n=5)	97 +/- 29
Estrone-3-sulfate*	Steroidal Hormone	Estrone-d4	ESI -	111 +/- 43 (n=3)	97 +/- 29
Fluoxetine	PhAC - antidepressant	Fluoxetine-d5	ESI +	84 +/- 7 (n=5)	30 +/- 11
Ibuprofen	PhAC - analgesic	Ibuprofen-d3	ESI -	98 +/- 28 (n=1)	64 +/- 24
Iopromide	Contrast Agent	Iopromide-d3	ESI +	97 +/- 4 (n=2)	100 +/- 19
Meprobamate	PhAC - anxiolytic	Meprobamate-d3	ESI +	96 +/- 8 (n=5)	24 +/- 17
Methylparaben*	PCP - antimicrobial	Methylparaben-d4	ESI -	80 +/- 11 (n=5)	101 +/- 30
Naproxen	PhAC - analgesic	Naproxen-d3	ESI -	87 +/- 39 (n=3)	60 +/- 23
Norfluoxetine	PhAC - metabolite	Norfluoxetine-d5	ESI +	73 +/- 19 (n=5)	12 +/- 8
Oryzalin	Herbicide	FipronildesF3	ESI -	80 +/- 3 (n=5)	127 +/- 26
Oxybenzone*	PCP - UV Stabilizer	Oxybenzone-d5	ESI +	110 (n=1)	22 +/- 8
Pendimethalin	Herbicide	Pendimethalin-d5	ESI +	95 +/- 8 (n=5)	30 +/- 17
Primidone	PhAC - antidepressant	Primidone-d5	ESI +	79 +/- 3 (n=5)	43 +/- 30
Progesterone*	Steroidal Hormone	Progesterone-d9	ESI +	93 +/- 25 (n=5)	36 +/- 14
Prometon	Herbicide	Prometon-d3	ESI +	91 +/- 5 (n=5)	55 +/- 30
Propylparaben*	PCP - antimicrobial	Propylparaben-d4	ESI -	88 +/- 13 (n=5)	112 +/- 35
Sulfamethoxazole	PhAC - antibiotic	Sulfamethoxazole-d4	ESI +	88 +/- 2 (n=5)	27 +/- 20
TCEP	Flame Retardant	TCEP-d12	ESI +	119 +/- 14 (n=5)	29 +/- 25
Testosterone*	Steroidal Hormone	Testosterone-d2	ESI +	83 +/- 7 (n=5)	26 +/- 16
Triclocarban	PCP - antimicrobial	Triclocarban-d4	ESI -	84 +/- 6 (n=5)	40 +/- 11
Trimethoprim	PhAC - antibiotic	Trimthoprim-d9	ESI +	91 +/- 17 (n=5)	57 +/- 22

Table A.2 Concentrations measured in the porewater equipment blanks (EBs) and lab blanks included in the study. The number of blanks with detections above the limit of quantitation (LOQ) are reported along with the calculated reporting limit (RL), if necessary.

Analyte	SUL_ EB	SUL_ LB	CED_ EB	CED_ LB	PRL_ EB	PRL_ LB	FRK_ EB	FRK_ LB	MAR_ EB	MAR_ LB
4-n-nonylphenol	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2
4-tert-octylphenol	<4	<4	<4	<4	<4	<4	5	<4	<4	<4
17a ethinylestradiol	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2
2,4-D	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2
Amitriptyline	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Androstenedione	<1	<1	<1	<1	<1	<1	<1	2	<1	<1
Atenolol	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Bisphenol A	<4	<4	11	29	4	10	<4	71	4	16
Caffeine	3	<1	<92	1	43	<1	10	3	7	1
Carbamazepine	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9
Cimetidine	<1	<1	1	<1	<1	<1	<1	<1	57	71
DEET	<2	<2	2	<2	3	<2	<2	<2	<2	<2
Diazepam	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2
Dilantin	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Diuron	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2
Estriol	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Estrone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Estrone 3 sulfate	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9
Fluoxetine	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Ibuprofen	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9
Iopromide	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9
Meprobamate	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Methylparaben	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4
Norfluoxetine	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Naproxen	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9
Oryzalin	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Oxybenzone	6	5	142	9	23	7	25	20	39	9
Pendimethalin	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Primidone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Progesterone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Prometon	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2
Propylparaben	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4
Sulfamethoxazole	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
TCEP	5	<1	7	5	2	5	2	5	5	10
Testosterone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Triclocarban	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Trimethoprim	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1

Table A.3 Detected concentrations of TOrCs in HOME versus REF sampling locations. Limits of quantitation (LOQs) and reporting limits (RLs) also provided.

Analyte	LOQ (ng/L)	RL (ng/L)	HOME Detection Frequency (%)	Concentration Range (ng/L)
DEET	2	3.5497	100	86 - 360
2,4-D	2	-	40	4.7 - 2200
Oxybenzone	4	152.476	40	190 - 380
Meprobamate	1	-	30	1.5 - 10
Methylparaben	4	-	30	4.8 - 11
Estrone	1	-	20	1.3 - 1.9
Pendimethalin	1	-	20	1.0 - 6.0
Testosterone	1	-	20	4.0 - 35
4-tert-octylphenol	4	14.7906	10	18
Androstenedione	1	-	10	5.8
Carbamazepine	9	-	10	430
Dilantin	1	-	10	14
Ibuprofen	9	-	10	26
Prometon	2	-	10	2.2
Analyte	LOQ	RL	REF Detection Frequency (%)	Concentration Range (ng/L)
DEET	2	3.5497	70	3.6 - 130
Methylparaben	4	-	50	4.7 - 11
Oxybenzone	4	152.476	40	180 - 310
17α-ethinylestradiol	2	-	20	5.0 - 25
Estrone	1	-	20	1.2 - 1.4
2,4-D	2	-	10	430
Androstenedione	1	-	10	6

Table A.4 Average grab porewater TOrC concentrations measured at each sampling location (ng/L)). Site distinctions are displayed for each site as either HOME or REF. P-values assess if there is a significant difference in average concentrations in HOME versus REF sites across all lakes. Analytes without no detections (n.d.) could not have a two-group hypothesis test performed on them.

	SUL_A	SUL_B	SUL_C	SUL_D	CED_A	CED_B	CED_C	CED_D	PRL_A	PRL_B	PRL_C	PRL_D	FRK_A	FRK_B	FRK_C	FRK_D	MAR_A	MAR_B	MAR_C	MAR_D	
Analyte	HOME	HOME	REF	REF	REF	HOME	REF	HOME	HOME	HOME	REF	REF	HOME	HOME	REF	REF	HOME	HOME	REF	REF	p-value
4-n-nonylphenol	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	n.d.
4-tert-octylphenol	<4	<4	<15	<15	<15	<4	<4	<4	<15	18	<4	<4	<15	<15	<15	<15	<15	<4	<15	<15	0.317
17a ethinylestradiol	<2	<2	<2	25	5	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	0.147
2,4-D	<2	5	<2	<2	<2	<2	<2	<2	<2	26	<2	<2	<2	<2	<2	<2	2213	2220	433	<2	0.136
Amitriptyline	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Androstenedione	<1	<1	<1	<1	6	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	6	<5	<1	<1	0
Atenolol	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Bisphenol A	<4	<4	<92	<4	<92	<92	<92	<92	<4	<92	<4	<4	<92	<92	<92	<92	<92	<92	<92	<92	n.d.
Caffeine	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	<51	n.d.
Carbamazepine	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	426	<9	<9	<9	<9	<9	<9	<9	0.317
Cimetidine	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
DEET	358	125	127	72	4	282	19	86	129	274	<2	4	99	117	<2	<2	189	159	49	43	<0.001
Diazepam	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	n.d.
Dilantin	<1	14	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	0.317
Diuron	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	n.d.
Estriol	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Estrone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	1	<1	2	1	1	<1	0.914
Estrone-3-sulfate	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	n.d.
Fluoxetine	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Ibuprofen	<9	<9	<9	<9	<9	26	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	0.317
Iopromide	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	n.d.
Meprobamate	<1	<1	<1	<1	<1	3	<1	<1	<1	2	<1	<1	<1	10	<1	<1	<1	<1	<1	<1	0.068
Methylparaben	<4	<4	5	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	5	11	5	11	9	6	5	0.521
Naproxen	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	<9	n.d.
Norfluoxetine	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Oryzalin	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Oxybenzone	<152	<152	<152	<152	<152	<152	<152	<152	<152	<152	<152	<152	294	384	314	254	223	190	182	221	0.865
Pendimethalin	<1	<1	<1	<1	<1	<1	<1	<1	6	1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	0.147
Primidone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Progesterone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Prometon	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	<2	2	<2	<2	<2	0.317
Propylparaben	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	<4	n.d.
Sulfamethoxazole	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
TCEP	<1	<1	<12	<12	<1	<1	<1	<12	<12	<12	<12	<12	<12	<12	<1	<1	<12	<12	<12	<1	n.d.
Testosterone	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	4	35	<1	<1	0.147
Triclocarban	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.
Trimethoprim	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	n.d.

Table A.5 Basic water Quality measurements at each site. Porewater(PW) and surface water(SW) measurements both provided. The field probe was not working at Lake Mary so conductivity values are not available for this lake.

SiteName	Type	pH	Conductivity PW (µS/cm)	Temperature PW (°C)	Dissolved Oxygen PW (mg/L)	Conductivity SW (µS/cm)	Temperature SW (°C)	Dissolved Oxygen SW (mg/L)
SUL_C_PW	REF	6.8	710	13	3.3	422	15.79	11
SUL_D_PW	REF	6.7	1100	12	2.7	421	15.45	12
PRL_C_PW	REF	7.3	730	10	9	467	13.75	14
PRL_D_PW	REF	7.3	680	11	6	457	13.55	11
CED_A_PW	REF	7	590	14	4.4	413	14.79	14
CED_C_PW	REF	6.7	1200	12	5.4	418	15.18	11
MAR_C_PW	REF	6.6	550	18	1	-	18.9	9.4
MAR_D_PW	REF	6.7	640	21	1.1	-	22.7	8.6
FRK_C_PW	REF	6.4	1000	14	2.8	939	22.57	14
FRK_D_PW	REF	6.5	940	14	2.2	789	19.69	12
SUL_A_PW	HOME	6.7	650	14	2.9	423	15.63	11
SUL_B_PW	HOME	6.7	1500	14	4.3	430	14.43	11
PRL_A_PW	HOME	7.3	460	14	3.1	443	14.29	15
PRL_B_PW	HOME	7.4	480	13	3.1	418	13.11	16
CED_B_PW	HOME	6.8	980	14	3.8	419	16.78	13
CED_D_PW	HOME	7.2	1200	15	4.1	418	16.83	13
MAR_A_PW	HOME	6.6	380	21	1	-	21.6	8.5
MAR_B_PW	HOME	6.5	670	20	1	-	20.1	8.6
FRK_A_PW	HOME	6.7	990	17	4.1	840	19.48	9.5
FRK_B_PW	HOME	6.9	1300	17	2.6	839	19.9	9.3

Table A.6 Average concentration (mg/L) of major nutrients measured at each sampling location. TOC=totoal organic carbon and TN= total nitrogen. Reporting levels were determined by the methods described in the quality control and quality assurance section of the manuscript using field and laboratory blank concentration reported beneath the environmental concentrations. "n.a." indicates there were no detections in the blank samples and a reporting level was not applicable for this analyte. P-values assess if there is a significant difference in average concentrations in HOME versus REF sites across all lakes. Analytes without no detections (n.d.) could not have a t-test performed on them.

SiteName	SiteType	Chloride	Bromide	Fluoride	Nitrite as NO2	Nitrate as NO3	Phosphate	Sulfate	TOC	TN as N
CED_A_PW	REF	8.87	<0.353	<0.241	<0.331	<0.1	<0.5	31.67	5.96	<0.17
CED_C_PW	REF	<0.1	<0.353	<0.1	<0.1	<0.497	<0.5	<0.429	16.37	1.552
FRK_C_PW	REF	<0.1	<0.353	0.33	<0.1	<0.1	1.67	2.75	29.45	9.441
FRK_D_PW	REF	<0.1	<0.353	0.36	<0.1	<0.497	<0.5	2.75	22.65	6.523
MAR_C_PW	REF	29	<0.1	<0.1	<0.1	<0.1	<0.5	3.64	<0.17	<0.17
MAR_D_PW	REF	25.73	<0.353	<0.241	<0.331	<0.1	<0.5	<0.429	13.35	3.034
PRL_C_PW	REF	18.58	<0.1	<0.241	<0.1	135.61	<0.5	27.45	<3.25	22.633
PRL_D_PW	REF	25.48	<0.1	<0.241	<0.1	0.52	<0.5	33.92	<3.25	0.542
SUL_C_PW	REF	22.91	<0.353	<0.241	<0.1	<0.1	1.43	<0.429	8.91	4.397
SUL_D_PW	REF	17.74	<0.1	0.31	<0.1	<0.1	<0.5	<0.429	18.68	5.421
CED_B_PW	HOME	<0.1	<0.353	<0.1	<0.1	<0.1	<0.5	40.88	11.47	0.523
CED_D_PW	HOME	<0.1	<0.1	<0.1	<0.1	<0.1	<0.5	31.99	<3.25	<0.17
FRK_A_PW	HOME	<0.1	<0.353	0.4	<0.331	<0.497	<0.5	11.95	7.02	<0.17
FRK_B_PW	HOME	<0.1	<0.353	0.37	<0.1	<0.1	<0.5	39.18	5.67	0.693
MAR_A_PW	HOME	23.74	<0.353	<0.1	<0.331	<0.497	<0.5	7.22	<0.17	<0.17
MAR_B_PW	HOME	53.61	<0.353	0.27	<0.331	<0.1	<0.5	4.55	17.56	0.122
PRL_A_PW	HOME	19.62	<0.1	<0.241	<0.331	<0.1	<0.5	6.67	4.44	0.011
PRL_B_PW	HOME	23.97	<0.353	<0.241	<0.331	<0.1	<0.5	5.3	12.61	0.053
SUL_A_PW	HOME	28.15	<0.1	0.33	<0.1	<0.1	<0.5	24.64	8.79	0.542
SUL_B_PW	HOME	24.89	<0.1	<0.241	<0.1	<0.1	<0.5	<0.429	14.27	8.542
	p-value	0.88	n.d	0.57	n.d.	0.15	0.15	0.07	0.45	0.02
Blank 1		<0.1	<0.1	<0.1	0.234	<0.1	<0.5	0.146	1.107	<0.17
Blank 2		0.51106	0.17325	0.11448	<0.1	0.30861	<0.5	0.3089	1.255	<0.17
Blank 3		0.55844	0.16806	<0.1	<0.1	0.19265	<0.5	0.20667	1.997	<0.17
Blank 4		0.466	0.259	0.163	0.263	<0.1	<0.5	0.155	1.472	<0.17
Blank 5		0.38	<0.1	<0.1	0.282	<0.1	<0.5	<0.1	0.78	<0.17
Blk Avg		0.479	0.20012	0.13858	0.2595	0.25063	n.a.	0.20402	1.416	n.a.
Blk StDev		0.07576	0.05111	0.03407	0.02397	0.082	n.a.	0.07488	0.61041	n.a
RL		0.70628	0.35346	0.2408	0.3314	0.49662	n.a.	0.42867	3.24755	n.a

Table A.7 Average concentration of trace metals measured at each sampling site (mg/L). Reporting levels were determined by the methods described in the quality control and quality assurance section of the manuscript using field and laboratory blank concentration reported beneath the environmental concentrations. "n.a." indicates there were no detections in the blank samples and a reporting level was not applicable for this analyte. p-values assess if there is a significant difference in average concentrations in HOME versus REF sites across all lakes. Analytes without no detections (n.d.) could not have a t-test performed on them.

Site Name	Group	Al	As	B	Ba	Be	Ca	Cd	Co	Cr	Cu	Fe	K	Li	Mg	Mn	Na	Ni	P	Pb	S	Se	Si	Sr	Tl	V	Zn	Sn	Mo	Sb	Ti
CED_A_PW	REF	<0.038	<0.005	<0.05	0.04	0.0002	75	<0.0003	<0.0002	<0.0003	<0.003	0.015	0.59	0.02	21	0.07	6.54	0.001	0.03	<0.0020	10.98	0.05	15	0.16	<0.0027	<0.0008	0.013	<0.0017	0.005	<0.0051	<0.0002
CED_C_PW	REF	<0.038	<0.005	0.14	0.26	<0.0002	157	<0.0003	<0.0002	<0.0003	<0.001	4.015	3.22	0.01	47	0.92	2.72	<0.0005	0.03	<0.0020	3.8	<0.01	17	0.31	<0.0027	<0.0008	0.017	<0.0017	0.003	<0.0051	<0.0002
FRK_C_PW	REF	<0.038	<0.005	0.15	0.11	<0.0002	98	<0.0003	<0.0002	<0.0003	<0.001	0.033	3.66	0.03	60	0.33	9.11	0.001	0.69	<0.0020	4.9	<0.01	22	0.32	<0.0027	<0.0008	0.012	<0.0017	<0.001	<0.0051	<0.0002
FRK_D_PW	REF	<0.038	<0.005	0.17	0.1	<0.0002	111	<0.0003	<0.0002	<0.0003	<0.001	1.489	2.87	0.04	48	0.3	6.78	0.001	0.24	<0.0020	3.94	<0.01	20	0.34	<0.0027	<0.0008	0.017	<0.0017	<0.001	<0.0051	<0.0002
MAR_C_PW	REF	<0.038	<0.005	<0.09	0.11	0.0002	60	<0.0003	<0.0002	<0.0003	<0.001	1.726	5.54	0.01	17	0.74	4.31	0.001	0.02	<0.0020	2.46	<0.01	11	0.15	<0.0027	<0.0008	0.016	<0.0017	0.003	<0.0051	<0.0002
MAR_D_PW	REF	<0.02	<0.005	<0.05	0.08	<0.0002	66	<0.0003	<0.0002	<0.0003	<0.001	0.145	5.06	0.01	18	0.76	4.67	<0.0005	<0.01	<0.0020	1.82	<0.01	12	0.13	<0.0027	<0.0008	0.011	<0.0017	0.002	<0.0051	<0.0002
PRL_C_PW	REF	<0.02	<0.005	<0.05	0.04	<0.0002	85	<0.0003	<0.0002	<0.0003	<0.001	<0.003	1.43	<0.0007	25	<0.00004	4.43	0.001	0.03	<0.0020	9.42	<0.01	10	0.07	<0.0027	<0.0008	0.011	<0.0017	0.004	<0.0051	<0.0002
PRL_D_PW	REF	<0.02	<0.005	<0.05	0.07	<0.0002	84	<0.0003	<0.0002	<0.0003	<0.001	<0.013	1.51	<0.0007	24	0.09	6.94	<0.0005	0.07	<0.0020	11.35	<0.01	13	0.08	<0.0027	<0.0008	0.013	<0.0017	0.005	<0.0051	<0.0002
SUL_C_PW	REF	<0.02	<0.005	<0.05	0.09	<0.0002	82	<0.0003	<0.0002	<0.0003	<0.001	0.019	7.32	0.01	27	1.65	10.17	0.001	0.03	<0.0020	1.95	<0.01	12	0.15	<0.0027	<0.0008	<0.01	<0.0017	0.004	<0.0051	<0.0002
SUL_D_PW	REF	<0.038	<0.005	<0.05	0.24	<0.0002	156	<0.0003	0.001	<0.0003	<0.001	0.051	5.44	0.02	39	2.56	9.25	0.001	0.03	<0.0020	3.74	<0.01	16	0.38	<0.0027	<0.0008	0.014	<0.0017	0.009	<0.0051	<0.0002
CED_B_PW	HOME	<0.038	<0.005	0.16	0.07	0.0002	117	<0.0003	<0.0002	<0.0003	<0.001	0.751	3.12	0.02	32	0.59	21.95	0.001	0.03	<0.0020	14.78	0.03	15	0.24	<0.0027	<0.0008	0.015	<0.0017	0.006	<0.0051	<0.0002
CED_D_PW	HOME	<0.038	<0.005	0.14	0.21	<0.0002	98	<0.0003	<0.0002	<0.0003	<0.001	<0.013	4.94	0.01	53	0.48	32.6	0.001	0.03	<0.0020	11.39	0.01	5	0.16	<0.0027	<0.0008	0.013	<0.0017	0.002	<0.0051	<0.0002
FRK_A_PW	HOME	<0.02	0.01	0.24	0.05	<0.0002	105	<0.0003	<0.0002	<0.0003	<0.001	0.016	2.27	0.06	55	1.14	15.46	0.002	<0.01	<0.0020	6.07	<0.01	10	0.19	<0.0027	<0.0008	0.012	<0.0017	0.002	<0.0051	<0.0002
FRK_B_PW	HOME	<0.02	0.01	0.16	0.1	<0.0002	147	<0.0003	<0.0002	<0.0003	<0.001	0.019	11.5	0.07	57	1.13	18.03	0.002	0.01	<0.0020	14.86	<0.01	16	0.42	<0.0027	<0.0008	0.014	<0.0017	<0.001	<0.0051	<0.0002
MAR_A_PW	HOME	<0.038	<0.005	<0.05	0.07	<0.0002	41	<0.0003	<0.0002	<0.0003	<0.001	0.276	3.59	0.01	12	0.68	4.47	0.001	0.03	<0.0020	3.35	<0.01	3	0.08	<0.0027	<0.0008	0.02	<0.0017	0.002	<0.0051	<0.0002
MAR_B_PW	HOME	<0.02	<0.005	<0.05	0.09	<0.0002	74	<0.0003	<0.0002	<0.0003	<0.001	0.59	2.31	0.01	18	0.73	5.71	0.001	<0.01	<0.0020	3.44	<0.01	6	0.16	<0.0027	<0.0008	<0.01	<0.0017	<0.001	<0.0051	<0.0002
PRL_A_PW	HOME	<0.02	<0.005	<0.05	0.07	<0.0002	46	<0.0003	<0.0002	<0.0003	<0.001	<0.013	3.81	<0.0007	19	0.77	5.99	<0.0005	0.05	<0.0020	3.09	<0.01	14	0.07	<0.0027	<0.0008	<0.01	<0.0017	0.003	<0.0051	<0.0002
PRL_B_PW	HOME	<0.02	<0.005	<0.09	0.06	<0.0002	44	<0.0003	<0.0002	<0.0003	<0.001	0.681	1.33	<0.0007	19	0.08	14.37	<0.0005	0.09	<0.0020	2.7	<0.01	8	0.05	<0.0027	0.002	<0.01	<0.0017	0.003	<0.0051	<0.0002
SUL_A_PW	HOME	<0.02	<0.005	0.11	0.12	<0.0002	86	<0.0003	<0.0002	<0.0003	<0.001	0.135	4.52	0.01	18	2.11	9.93	0.001	0.03	<0.0020	2.03	<0.01	17	0.17	<0.0027	<0.0008	0.011	<0.0017	0.004	<0.0051	<0.0002
SUL_B_PW	HOME	<0.02	<0.005	0.09	0.12	<0.0002	169	<0.0003	<0.0002	<0.0003	<0.001	0.052	5.14	0.02	45	2.27	43.57	0.003	0.04	<0.0020	3.85	<0.01	16	0.33	<0.0027	<0.0008	0.016	<0.0017	0.008	<0.0051	<0.0002
		n.a.	0.15	0.27	0.74	0.54	0.88	n.a.	0.32	n.a.	n.a.	0.91	0.94	0.88	0.88	0.33	0.02	0.18	0.26	n.a.	0.65	0.63	0.23	0.97	n.a.	0.32	0.6	n.a.	0.52	n.a.	n.a.
		Al	As	B	Ba	Be	Ca	Cd	Co	Cr	Cu	Fe	K	Li	Mg	Mn	Na	Ni	P	Pb	S	Se	Si	Sr	Tl	V	Zn	Sn	Mo	Sb	Ti
Blank 1		<0.0205	<0.0047	<0.0459	0.0002	<0.0002	0.0842	<0.0001	<0.0003	<0.0003	<0.0011	<0.0003	0.0485	<0.0014	0.0181	0.0001	0.1729	<0.0005	<0.0106	<0.0020	0.0431	<0.0100	0.0156	0.0004	<0.0019	<0.0008	0.0069	<0.0017	<0.0011	<0.0051	<0.0002
Blank 2		0.0303	<0.0047	<0.0459	0.0003	<0.0002	0.0201	0.0003	<0.0003	<0.0003	0.0015	<0.0003	0.0383	<0.0014	0.0022	0.0002	<0.01345	<0.0005	<0.0106	<0.0020	<0.0276	<0.0100	0.0231	<0.0001	<0.0019	<0.0008	0.0004	<0.0017	<0.0011	<0.0051	<0.0002
Blank 3		<0.0205	<0.0047	0.0604	0.0004	<0.0002	0.0326	<0.0001	<0.0003	<0.0003	<0.0011	0.0044	<0.0323	0.0018	0.0018	<0.00004	<0.1345	<0.0005	<0.0106	<0.0020	<0.0276	<0.0100	0.0693	0.0002	<0.0019	<0.0008	0.0014	<0.0017	<0.0011	<0.0051	<0.0002
Blank 4		0.0253	<0.0042	0.0424	<0.0001	<0.0001	0.0406	<0.0003	<0.0003	<0.0003	0.0007	<0.0003	<0.0404	<0.0007	0.005	<0.0003	0.0106	<0.0004	<0.0083	<0.0018	0.0296	<0.0061	0.0709	<0.0001	<0.0027	<0.0027	0.0018	<0.0005	<0.0011	<0.0022	<0.0002
Blank 5		<0.0205	<0.0047	<0.0459	0.0002	<0.0002	0.0449	<0.0001	<0.0003	<0.0003	<0.0011	<0.0003	0.0541	<0.0014	0.0101	<0.0001	0.1514	<0.0005	<0.0106	<0.0020	<0.0276	<0.0100	0.0325	<0.0001	0.003	<0.0008	0.0015	<0.0017	<0.0011	<0.0021	<0.0002
Blank 6		<0.0205	<0.0047	<0.0459	<0.0001	<0.0002	<0.0064	<0.0001	0.0003	<0.0003	<0.0011	<0.0003	<0.0323	<0.0014	0.0005	<0.0001	<0.1345	<0.0005	<0.0106	<0.0020	<0.0276	<0.0100	<0.0114	<0.0001	<0.0019	<0.0008	<0.0003	<0.0017	<0.0011	<0.0021	<0.0002

Table A.8 Average blood concentrations of vitellogenin and 11-ketotestosterone measured at each site. Average values for body condition factor, gonadal somatic index (GSI), and hepatosomatic index (HSI) at each site are also displayed.

Site Name	Group	Vitellogenin (mg/mL)	11-Ketotestosterone (mg/mL)	Body Condition Factor	Hepatosomatic Index	Gonadal Somatic Index
CED_B	REF	593.02	10.6596	2.3626	1.3016	0.9604
CED_C	REF	567.92	1.5478	2.269	1.0749	0.217
FRK_C	REF	590.53	0.7999	2.3733	0.9252	0.0726
FRK_D	REF	1137.26	0.5538	2.2916	0.9574	0.0959
MAR_C	REF	398.84	10.9269	1.9652	0.9254	1.4426
MAR_D	REF	458.24	9.9662	1.9718	0.9846	1.4354
SUL_C	REF	1382.82	28.911	1.9183	0.8989	1.1411
SUL_D	REF	201.7	23.2721	1.7304	0.9966	1.0697
CED_A	HOME	745.6	6.5947	2.314	1.3111	0.9909
FRK_A	HOME	1933.42	12.106	2.3078	1.1331	0.3658
FRK_B	HOME	689.89	0.5555	2.4762	1.1591	0.145
MAR_A	HOME	2692.01	13.4626	2.0032	1.1519	1.661
MAR_B	HOME	1533.42	9.5319	1.9832	0.9016	1.0394
SUL_A	HOME	1270.85	30.4394	2.1095	0.9509	0.9704
SUL_B	HOME	1551.7	23.6049	2.0155	0.9505	1.2294

APPENDIX B

SUPPORINTG INFORMATION FOR CHAPTERS 3 AND 4

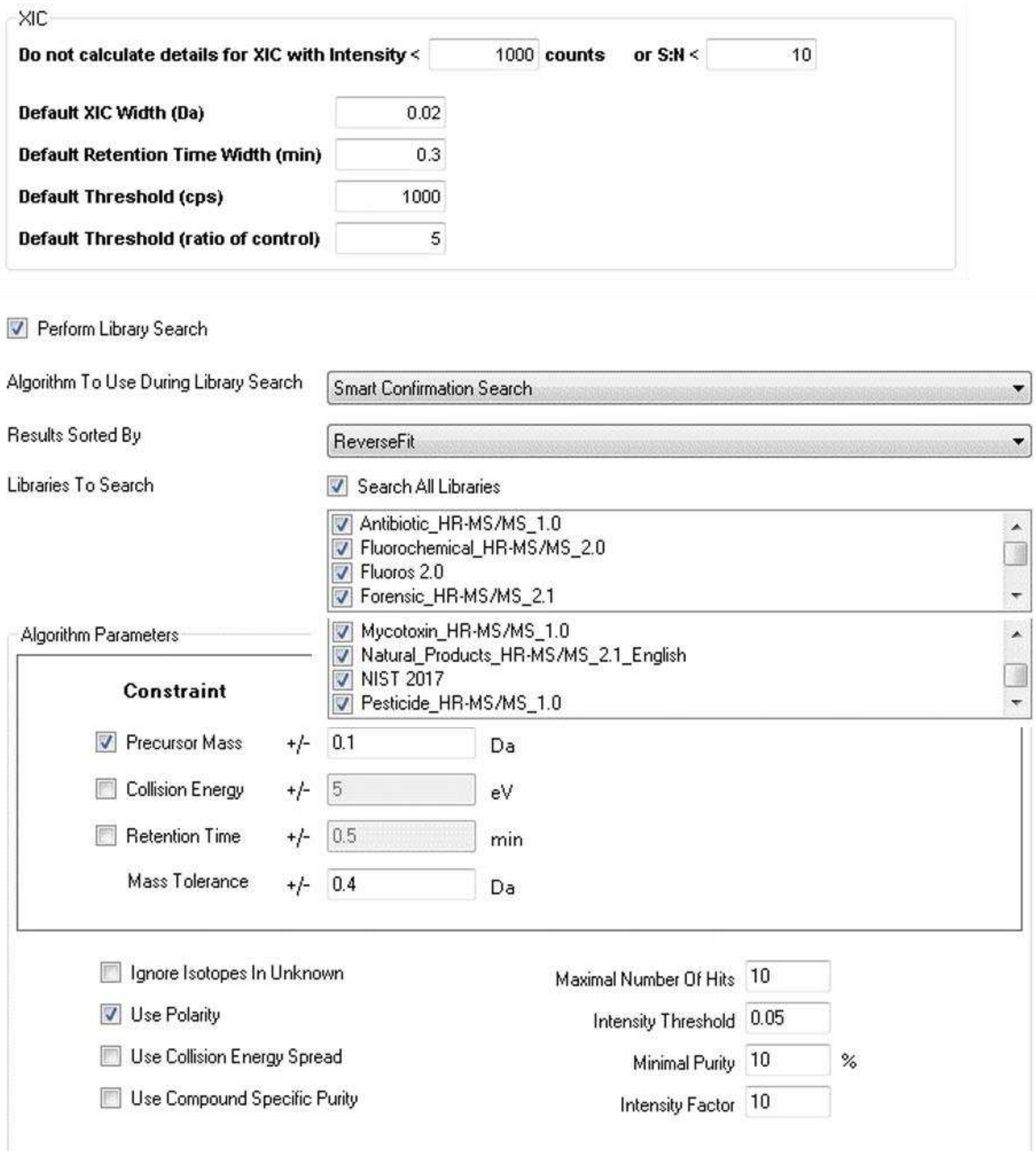

Figure B.1 Masterview peak pick and library search algorithm peak picking input parameters

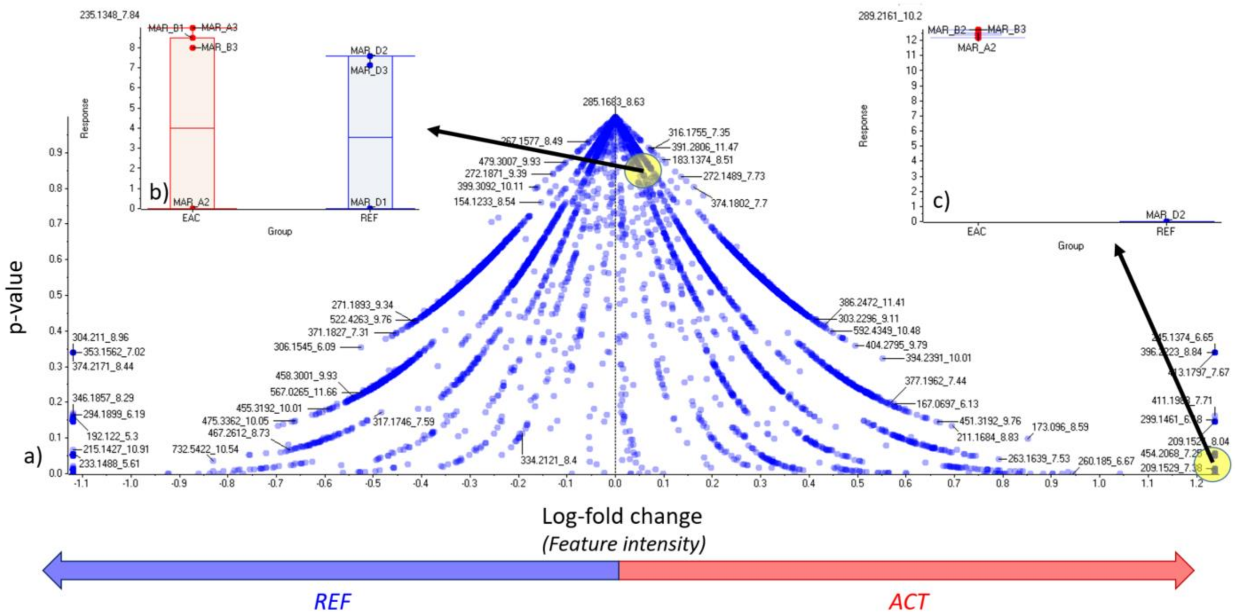

Figure B.2 a) “Volcano plot” visualization of t-test feature prioritization for the grab pos dataset. Representative box plots comparing ACT and REF feature responses for b) excluded and c) retained features after the t-test analysis are also displayed.

Increasing Confidence

Schymanksi Scale	Feature Characteristics	
Level 5: Mass of Interest	Poor match or No match to Library Spectra or formula finder, Acquires MS/MS not required	Unresolved by Library and Formula Search
Level 4: Unequivocal Molecular Formula	>70% match in formula finder across all replicates, Acquired MS/MS not required	Resolved by Formula Search
Level 3: Tentative Candidate	< 2 matched fragments, multiple library matches, skewed relative intensities of fragments	Resolved by Library Search
Level 2: Probable Structure	All diagnostic fragments present with suitable relative intensity matches	
Level 1: Confirmed Structure	Confirmed retention time, isotope spectra, and fragmentation with reference standard	

Figure B.3 Detailed description of confidence metrics associated with feature identifications. Adopted from Schymanski et al., 2014.

Table B.1 Level 3 IDs from the *grab neg* dataset

Feature (m/z_RT)	Site Designation	Confidence	ID
130.1228_7.53	AB	Level 3	Metformin
161.0958_8.3	AB	Level 3	Napthorescorcinol
174.054_6.72	AB	Level 3	5,8-dimethylquinolin-4-ol
270.0791_8.55	AB	Level 3	Acetochlor
397.3838_11.96	AB	Level 3	Vitamin D2
287.2006_10.89	AB	Level 3	Androstan-4,6-diene-17-beta-ol-3-one
287.2322_6.93	AB	Level 3	16 alpha hydroxyestrone
289.2114_7.3	AB	Level 3	16 beta hydroxyestradiol
174.0909_6.69	A	Level 3	5,8-dimethylquinolin-4-ol
287.1998_10.65	A	Level 3	9(11)-didehydroestriol
292.191_7.44	A	Level 3	Cyclopnetoloate
287.0907_8.96	A	Level 3	Caterarin
423.3603_9.64	A	Level 3	Tetradecanimide
595.1636_7.38	A	Level 3	Saponarian
595.1645_7.65	A	Level 3	Vitexin-4-O-glucoside
160.0758_8.78	B	Level 3	2-methylquinolin-8-ol
191.1062_8.59	B	Level 3	Lingustilide
148.0754_8.47	B	Level 3	1-phyenylpyrorolidine
234.0582_9.25	B	Level 3	Normeperedine

Table B.2 Level 3 IDs form the *grab neg* dataset

Feature (m/z_RT)	Site Designation	Confidence	ID
149.0612_7.67	AB	Level 3	4-ethylbenzoic acid
207.0297_6.39	A	Level 3	Chorismic acid
563.1448_7.24	A	Level 3	Isoschaftoside
199.1343_8.91	A	Level 3	Dodecanoic acid
223.1127_9.86	A	Level 3	4-hydroxychalcone
253.051_9.36	A	Level 3	3-furancarboxylic acid
255.0666_9.22	A	Level 3	Pinocembrin
249.1502_9.25	B	Level 3	Citrinin
162.0563_6.86	B	Level 3	8-fluoroquinolin-4-ol
253.087_9.2	B	Level 3	Daidzein
319.101_9.31	B	Level 3	alpha-Zearalenol

Table B.3 Level 3 IDs form the *temp int pos* dataset

Feature (m/z_RT)	Site Designation	Confidence	ID
473.361_11.43	AB	Level 3	Sumaresinolic Acid
179.1062_8.26	A	Level 3	Coniferyl aldehyde
134.0712_7.96	B	Level 3	Indoxyl
149.132_9.14	B	Level 3	4-chromanone
159.1376_8.72	B	Level 3	4,6-dioxoheptanoic acid
185.1168_9.13	B	Level 3	Methyl gallate
211.1324_9.31	B	Level 3	Jasmonic Acid
233.2256_10.38	B	Level 3	Costunolide
308.2579_10.57	B	Level 3	Androsterone, ammonium adduct
311.2216_9.34	B	Level 3	cis-13-eicosenoic acid
145.1221_7.12	B	Level 3	4-hydroxycyclohexylcarboxylic acid
161.1168_8.1	B	Level 3	2,3-dihydroxynapthalene
173.117_8.88	B	Level 3	Corey lactone diol
179.1425_9.14	B	Level 3	7-methoxychromanone
180.1588_7.09	B	Level 3	Adenochrome
199.1324_10.21	B	Level 3	delta-dodecalactone
231.158_10.21	B	Level 3	Dehydrocostus lactone
253.2517_10.43	B	Level 3	Palmitic acid alkyne
301.2369_10.05	B	Level 3	Retenoic Acid
313.273_10.43	B	Level 3	Norgesterol
473.361_11.43	AB	Level 3	Sumaresinolic Acid

Table B.4 Level 3 IDs from the *temp int neg* dataset

Feature (m/z_RT)	Site Designation	Confidence	ID
132.0572_7.75	B	Level 3	Indoxyl
159.1025_7.16	B	Level 3	2-hydroxyoctanoic acid
173.0819_6.54	B	Level 3	Suberic acid
185.0816_7.85	B	Level 3	Endothal
197.0453_7.22	B	Level 3	Ethyl gallate

Table B.5 Level 4 IDs from the *grab pos* dataset

Feature (m/z_RT)	Site Designation	Confidence	ID
252.1048_9.49	AB	Level 4	C13H17NO2S
272.2329_6.04	AB	Level 4	C14H29N3O2
273.217_6.83	AB	Level 4	C14H28N2O3
445.2433_5.99	AB	Level 4	C22H36O9
268.1915_9.81	B	Level 4	C15H25NO3

Table B.6 Level 4 IDs from the *grab neg* datset

Feature (m/z_RT)	Site Designation	Confidence	ID
305.2483_11.06	AB	Level 4	C20H34O2
195.0305_6.86	A	Level 4	C10H12O4
285.078_8.38	A	Level 4	C16H14O5
251.1655_9.48	B	Level 4	C15H24O3
285.113_8.31	B	Level 4	C17H18O4

Table B.7 Level 4 IDs from the *temp int pos* dataset

Feature (m/z_RT)	Site Designation	Confidence	ID
497.2867_11.37	AB	Level 4	C25H40N2O8
506.3472_11.31	AB	Level 4	C29H47NO6
546.3783_11.68	AB	Level 4	C32H51NO6
294.2419_8.16	A	Level 4	C18H31NO2
263.1636_11.18	B	Level 4	C16H22O3
293.2108_10.14	B	Level 4	C18H28O3
301.2634_10.43	B	Level 4	C20H32N2
311.2576_10.35	B	Level 4	C19H34O3
320.2426_8.67	B	Level 4	C16H33NO5
325.237_10.14	B	Level 4	C19H32O4
327.216_7.83	B	Level 4	C18H30O5
329.2318_9.31	B	Level 4	C18H32O5
330.2635_9.9	B	Level 4	C18H35NO4
334.258_8.95	B	Level 4	C17H35NO5
334.2948_9.79	B	Level 4	C18H39NO4
348.2741_9.17	B	Level 4	C18H37NO5
360.2737_9.7	B	Level 4	C19H37NO5
360.2743_10.18	B	Level 4	C19H37NO5_isomer
364.2691_8.58	B	Level 4	C18H37NO6
423.3249_11.18	B	Level 4	C29H42O2
451.3196_11.18	B	Level 4	C30H42O3

Table B.7 continued

Feature (m/z_RT)	Site Designation	Confidence	ID
469.3302_11.18	B	Level 4	C30H44O4
471.3457_11.88	B	Level 4	C30H46O4
487.3404_11.18	B	Level 4	C30H46O5
500.2791_9.71	B	Level 4	C32H37NO4

Table B.8 Level 4 IDs from the *temp int neg* dataset

Feature (m/z_RT)	Site Designation	Confidence	ID
233.1542_9.92	AB	Level 4	C15H22O2
239.129_9.34	AB	Level 4	C13H20O4
145.0873_7.31	B	Level 4	C7H14O3
157.0872_8.03	B	Level 4	C8H14O3
157.1235_8.5	B	Level 4	C9H18O2
183.1393_9.99	B	Level 4	C11H20O2
187.0978_7.81	B	Level 4	C9H16O4
187.1341_9.36	B	Level 4	C10H20O3
201.113_8.06	B	Level 4	C10H18O4
203.1287_7.77	B	Level 4	C10H20O4
213.1492_9.62	B	Level 4	C12H22O3
215.1285_7.75	B	Level 4	C11H20O4
247.1184_6.91	B	Level 4	C11H20O6
247.1547_7.8	B	Level 4	C12H24O5
257.212_10.42	B	Level 4	C15H30O3
259.1548_8.08	B	Level 4	C13H24O5
259.191_9.47	B	Level 4	C14H28O4
261.1703_8.27	B	Level 4	C13H26O5
273.2064_10.12	B	Level 4	C15H30O4
279.1962_9.99	B	Level 4	C17H28O3
285.1854_10.5	B	Level 4	C19H26O2
287.2229_10.34	B	Level 4	C16H32O4
297.2435_10.6	B	Level 4	C18H34O3
315.2533_9.99	B	Level 4	C18H36O4
317.2482_10.59	B	Level 4	C21H34O2
361.2232_8.05	B	Level 4	C18H34O7
373.2595_9.72	B	Level 4	C20H38O6
591.4636_10.36	B	Level 4	C36H64O6

Table B.9 Masses of interest from the *grab pos* dataset along with their reason for Level 5 IDs

Feature (m/z_RT)	Site Designation	Confidence	Reason for Level 5 ID
259.2007_6.4	AB	Level 5	No match/No formula found
259.2012_6.79	AB	Level 5	No match/No formula found
292.2262_8.51	AB	Level 5	Poor match
313.2483_7.21	AB	Level 5	Poor match
318.2423_9.16	AB	Level 5	Poor match
369.3506_11.64	AB	Level 5	Poor macth
499.2149_11.77	AB	Level 5	Poor match
565.1542_7.42	AB	Level 5	Poor match
188.069_7.62	A	Level 5	No match/No formula found
205.1326_6.69	A	Level 5	No match/No formula found
221.1275_6.14	A	Level 5	No match/No formula found
278.247_9.49	A	Level 5	Poor match-too few frags
298.2121_7.98	A	Level 5	Poor match
308.2584_10.32	A	Level 5	No acquired MSMS
314.2063_7.53	A	Level 5	Poor match
372.1427_6.42	A	Level 5	Poor macth
441.3718_9.65	A	Level 5	Poor match
469.3297_9.76	A	Level 5	Poor match
525.3421_10.14	A	Level 5	Poor match
535.1431_7.64	A	Level 5	Poor match
711.2119_7.47	A	Level 5	Poor match
773.1915_8.1	A	Level 5	Poor match
275.2221_6.86	B	Level 5	No match/No formula found
280.1359_10.37	B	Level 5	Poor match
318.2435_9.37	B	Level 5	Poor match
320.2589_9.2	B	Level 5	Poor match
342.2638_9.35	B	Level 5	No match/No formula found
352.1574_9.96	B	Level 5	No match/No formula found
358.1837_11.6	B	Level 5	Poor match-1 diagnostic fragment
360.2887_9.74	B	Level 5	Poor match
374.2686_9.33	B	Level 5	Poor match
374.2688_9.08	B	Level 5	Poor match
376.2856_9.1	B	Level 5	Poor match
392.2796_8.08	B	Level 5	Poor match
411.1838_9.04	B	Level 5	Poor match

Table B.10 Masses of interest from the *grab neg* dataset along with their reason for Level 5 IDs

Feature (m/z_RT)	Site Designation	Confidence	Reason for Level 5 ID
107.0513_7.83	AB	Level 5	No match/no formula match
219.1387_9.81	AB	Level 5	Poor match, too few frags
220.9591_8.56	AB	Level 5	No acquired MSMS
222.9565_8.58	AB	Level 5	Poor match
271.2025_6.04	AB	Level 5	Poor match
301.217_10.78	AB	Level 5	C20H30O2
370.0548_9.01	AB	Level 5	No match/no formula match
417.1262_9.66	AB	Level 5	Poor match
144.0457_6.38	A	Level 5	Poor match-too few frags
261.114_8.19	A	Level 5	Poor match
281.0671_7.54	A	Level 5	No acquired MSMS/Poor match
299.0764_4.75	A	Level 5	Poor match
339.0735_5.74	A	Level 5	Poor Match
396.1983_9.41	A	Level 5	Too few fragments
498.9311_9.81	A	Level 5	Poor match-too few frags
533.1327_7.6	A	Level 5	No match
563.1428_7.54	A	Level 5	Poor match
765.448_9.97	A	Level 5	Too few frags
779.4267_9.38	A	Level 5	Poor match-too few frags
793.4082_9.39	A	Level 5	Poor match-too few frags
189.0922_8.08	B	Level 5	Poor match
235.0763_8.98	B	Level 5	Poor match-too few frags
248.3374_9.25	B	Level 5	No acquired MSMS/No formula
279.1057_8.92	B	Level 5	Poor match
393.1746_8.95	B	Level 5	Poor match

Table B.11 Masses of interest from the *temp int neg* dataset along with their reason for Level 5 IDs

Feature (m/z_RT)	Site Designation	Confidence	Reason for Level 5 ID
178.1336_6.79	AB	Level 5	Poor match-too few frags
229.1804_6.72	AB	Level 5	Poor match, no formula
294.2057_10.3	AB	Level 5	Poor match, no formula
362.2531_8.03	AB	Level 5	poor match, no formula
182.0486_6.06	A	Level 5	No acquired MSMS, No formula
252.1049_9.45	A	Level 5	No match, no formula
288.216_8.24	A	Level 5	Poor match, poor formula match
362.2683_10.37	A	Level 5	poor match, noformula
390.2635_9.92	A	Level 5	Poo rmatch, no formula
642.2989_5.93	A	Level 5	No match, no formula
141.1274_8.72	B	Level 5	Poor match
142.1224_7.78	B	Level 5	No acquired MSMS, No formula
150.1487_7.72	B	Level 5	Poor match, no formula
159.1376_7.53	B	Level 5	Poor match, no formula
162.1484_8.61	B	Level 5	No match, no formula
172.1323_6.4	B	Level 5	No match, no formula
175.1328_8.68	B	Level 5	No match, no formula
178.1434_8.3	B	Level 5	No match, no formula
180.159_7.29	B	Level 5	Isomer of 180.1588_7.09, poor library match
187.0961_7.85	B	Level 5	Poor match, no formula
187.1323_9.41	B	Level 5	No acquried, no alignment of formula
189.0576_6.78	B	Level 5	Poor match, no formula
189.1118_8.71	B	Level 5	No acquired MSMS, No formula
192.1592_6.83	B	Level 5	Poor match, no formula
195.1013_9.94	B	Level 5	Poor match, no formula
195.1373_9.93	B	Level 5	Poor match
203.1273_8.73	B	Level 5	No match, no formula
204.1588_8.77	B	Level 5	No match, no formula
206.138_8.16	B	Level 5	Poor match, no formula
206.1743_7.8	B	Level 5	Poor match, no formula
207.2103_9.86	B	Level 5	No acquired MSMS, No formula
215.1861_5.82	B	Level 5	Poor match
217.143_9.15	B	Level 5	No acquired MSMS, No formula
226.2161_9.68	B	Level 5	Poor match, no acquired MSMS
178.1336_6.79	AB	Level 5	Poor match-too few frags
229.1804_6.72	AB	Level 5	Poor match, no formula
294.2057_10.3	AB	Level 5	Poor match, no formula

Table B.11 continued

Feature (m/z_RT)	Site Designation	Confidence	Reason for Level 5 ID
362.2531_8.03	AB	Level 5	poor match, no formula
182.0486_6.06	A	Level 5	No acquired MSMS, No formula
252.1049_9.45	A	Level 5	No match, no formula
288.216_8.24	A	Level 5	Poor match, poor formula match
362.2683_10.37	A	Level 5	poor match, noformula
390.2635_9.92	A	Level 5	Poo rmatch, no formula
642.2989_5.93	A	Level 5	No match, no formula
141.1274_8.72	B	Level 5	Poor match
142.1224_7.78	B	Level 5	No acquired MSMS, No formula
150.1487_7.72	B	Level 5	Poor match, no formula
159.1376_7.53	B	Level 5	Poor match, no formula
162.1484_8.61	B	Level 5	No match, no formula
172.1323_6.4	B	Level 5	No match, no formula
175.1328_8.68	B	Level 5	No match, no formula
178.1434_8.3	B	Level 5	No match, no formula
180.159_7.29	B	Level 5	Isomer of 180.1588_7.09, poor library match
187.0961_7.85	B	Level 5	Poor match, no formula
187.1323_9.41	B	Level 5	No acquried, no alignment of formula
189.0576_6.78	B	Level 5	Poor match, no formula
189.1118_8.71	B	Level 5	No acquired MSMS, No formula
192.1592_6.83	B	Level 5	Poor match, no formula
195.1013_9.94	B	Level 5	Poor match, no formula
195.1373_9.93	B	Level 5	Poor match
203.1273_8.73	B	Level 5	No match, no formula
204.1588_8.77	B	Level 5	No match, no formula
206.138_8.16	B	Level 5	Poor match, no formula
206.1743_7.8	B	Level 5	Poor match, no formula
207.2103_9.86	B	Level 5	No acquired MSMS, No formula
215.1861_5.82	B	Level 5	Poor match
217.143_9.15	B	Level 5	No acquired MSMS, No formula

Table B.12 Masses of interest from the *temp int pos* dataset along with their reason for Level 5 IDs

Feature (m/z_RT)	Site Designation	Confidence	Reason for Level 5 ID
367.1953_10.69	A	Level 5	Poor match, poor formula
151.1141_8.08	AB	Level 5	No match, no formula found
165.0557_7.77	AB	Level 5	Poor match, poor formula
125.0978_8.14	B	Level 5	Poor match, no formula
127.0773_8.2	B	Level 5	No acquired MSMS, no formula
127.1135_8.14	B	Level 5	No acquired MSMS, no formula
131.0716_5.57	B	Level 5	No match, no formula found
141.0927_8.84	B	Level 5	Poor match, poor formula
153.092_9.08	B	Level 5	Poor match, poor formula
155.0716_7.36	B	Level 5	No acquired MSMS, no formula
168.5395_8.14	B	Level 5	No acquired MSMS, no formula
169.0877_8.14	B	Level 5	Poor match, poor formula
169.1236_9.69	B	Level 5	Poor match, poor formula
170.5481_8.16	B	Level 5	No acquired MSMS, no formula
171.1034_8.14	B	Level 5	Poor match, poor formula
173.1181_7.82	B	Level 5	No acquired MSMS, no formula
175.0973_7.44	B	Level 5	Poor match, poor formula
175.0975_7.16	B	Level 5	Poor match, poor formula
181.1243_7.54	B	Level 5	No match, no formula found
183.1029_8.67	B	Level 5	Poor match, poor formula
189.1131_7.62	B	Level 5	Poor match, poor formula
191.1285_7.35	B	Level 5	Poor match, poor formula
193.1241_8.22	B	Level 5	Poor match, poor formula
197.118_9.22	B	Level 5	Poor match, poor formula
199.1335_9.25	B	Level 5	Poor match, poor formula
201.1493_9.58	B	Level 5	Poor match, poor formula
203.1286_8.81	B	Level 5	Poor match, poor formula
205.1076_7.52	B	Level 5	Poor match, poor formula
205.1242_8.44	B	Level 5	Poor match, poor formula
231.1595_7.9	B	Level 5	No acquired MSMS, no formula
255.1235_8.53	B	Level 5	No match, no formula found
261.1338_8.03	B	Level 5	Poor match, poor formula
266.3236_9.93	B	Level 5	No acquired MSMS, no formula
272.2303_10.12	B	Level 5	No acquired MSMS, no formula
277.2162_10.4	B	Level 5	$C_{18}H_{30}O_2$
281.2392_10.97	B	Level 5	No acquired MSMS, no formula
289.2283_10.34	B	Level 5	No acquired MSMS, no formula

Tabel B.12 continued

Feature (m/z_RT)	Site Designation	Confidence	Reason for Level 5 ID
292.2538_10.18	B	Level 5	No acquired MSMS, no formula
294.2637_10.39	B	Level 5	No acquired MSMS, no formula
300.2552_10.6	B	Level 5	Poor match, no formula
317.2594_9.98	B	Level 5	No acquired MSMS, no formula
322.2382_9.88	B	Level 5	Poor match, poor formula
443.28_10.12	B	Level 5	No match, poor formula match
527.3389_10.26	B	Level 5	Poor match, poor formula
593.4775_10.36	B	Level 5	No acquired MSMS, poor formula
621.4371_9.77	B	Level 5	No acquired MSMS, poor formula
623.4523_9.77	B	Level 5	No acquired MSMS, poor formula
659.4748_9.12	B	Level 5	No acquired MSMS, poor formula
717.5341_10.42	B	Level 5	No match, poor formula match
789.2425_6.15	B	Level 5	No match, poor formula match

Table B.13 Features with varying confidence levels detected in all three replictaes from a reference lake. No features detected in *temp int neg* samples.

			Detected in All 3 field replicates?			
Matrix	Feature (ID)	Confidence	CED_A	CED_B	CED_C	CED_D
Grab pos	192.1381_9.54 (DEET)	Level 1				x
	259.2012_6.79 (C13H26N2O3)	Level 4	x			
	268.1915_9.81 (C15H25NO3)	Level 4			x	
	289.2114_7.3 (C14H28N2O4)	Level 4		x		
	352.1574_9.96 (no match)	Level 5			x	
Grab neg	165.0557_6.41 (C9H10O3)	Level 4	x			
	189.0918_8.39 (No match)	Level 5			x	
Temp int pos	134.0712_7.96 (indoxyl)	Level 2		x		
	163.1114_9.15 (2 methoxycinnamaldehyde)	Level 3	x			
	173.117_8.88 (Poor match)	Level 5		x		x
	189.0576_6.78 (antipyrene)	Level 3	x			
	195.1013_9.94 (poor match)	Level 5	x		x	
	206.1743_7.8 (no acquired msms)	Level 5				x
	207.2103_9.86 (acetyleugenol)	Level 3				x
	226.2161_9.68 (C14H27NO)	Level 4	x			
	229.0857_10.27 (oxybenzone)	Level 3				x
	256.1415_5.9 (no match)	Level 5		x		x
	257.1893_10.59 (no match)	Level 5		x		

Table B.13 continued

			Detected in All 3 field replicates?			
Matrix	Feature (ID)	Confidence	CED_A	CED_B	CED_C	CED_D
Temp int pos	263.1093_10.42 (no match)	Level 5			x	
	294.2057_10.3 (Butoxycaine)	Level 3	x			
	295.2262_10.35 (no acquired MSMS)	Level 5				x
	295.2263_9.14 (C18H30O3)	Level 4				x
	299.2573_9.85 (no acquired MSMS)	Level 5		x		
	311.221_10.14 (cis-11-eicosenoic acid)	Level 3				x
	330.2635_9.9 (C18H35NO4)	Level 4				x
	330.2635_9.17 (no match)	Level 5		x		
	330.2636_9.15 (no acquired msms)	Level 5				x
	334.258_8.95 (no match)	Level 5				x
	334.2948_9.79 (C18H39NO4)	Level 4		x		
	348.2741_9.17 (C18H37NO5)	Level 4		x		x
	350.2802_9.18 (C20H35N3O2)	Level 4				x
	360.2743_10.18 (C19H37NO5)	Level 4				x
	362.2531_8.03 (1,5-hexanediol)	Level 3			x	
	364.2691_8.58 (poor match)	Level 5				x
	423.3249_11.18 (C29H42O2)	Level 4		x		
	451.3196_11.18 (C30H42O3)	Level 4		x		
	469.3302_11.18 (poor match)	Level 5		x		
	471.3457_11.88 (C30H46O4)	Level 4	x	x		
	473.361_11.43 (Maslinic acid)	Level 3	x	x		
	506.3472_11.31 (C29H47NO6)	Level 4		x	x	x
	771.2466_8.31 (C33H42N2O19)	Level 4		x		x