8-31-2020

Lakewide and Nearshore Microbial Water Quality Modelling in Lake St. Clair

Mohammad Madani
University of Windsor

Lakewide and Nearshore Microbial Water Quality Modelling in Lake St. Clair

By

Mohammad Madani

A Dissertation

Submitted to the Faculty of Graduate Studies

through the Department of Civil and Environmental Engineering

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy

at the University of Windsor

Windsor, Ontario, Canada

2020

Lakewide and Nearshore Microbial Water Quality Modelling in Lake St. Clair

by

Mohammad Madani

APPROVED BY:

_____

R. Rao Yerubandi, External Examiner
Environment and Climate Change Canada

_____

J. Crossman
School of the Environment

_____

R. Carriveau
Department of Civil and Environmental Engineering

_____

T. Bolisetti
Department of Civil and Environmental Engineering

_____

L.F. Leon, Co-advisor
Department of Civil and Environmental Engineering

_____

R. Seth, Advisor
Department of Civil and Environmental Engineering

July 23, 2020

DECLARATION OF CO-AUTHORSHIP / PREVIOUS PUBLICATION

I. Co-Authorship

I hereby declare that this dissertation incorporates materials that are the result of joint research, as follows:

Chapters 2 was completed under the supervision of my advisor, Dr. Rajesh Seth, and Chapters 3 to 6 were the joined research under the supervision of Dr. Seth and my co-advisor Dr. Luis F. Leon, in collaboration with Reza Valipour and Craig McCrimmon from Environment and Climate Change Canada (ECCC). In all cases, the key ideas, primary contributions, data analysis and writing were carried out by me. The contribution of the co-authors was primarily through advice on the broad research ideas, review of results, participation in scientific discussion, literature review and subsequently in editing the presentation and thesis material.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis. I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

Previous Publication

This dissertation includes two original papers that have been previously published in peer-reviewed journals, as follows:

Thesis Chapter	Authors	Title	Publication Status
2	Mohammad Madani, Rajesh Seth	Evaluating Multiple Predictive Models For Beach Management At A Freshwater Beach On Lake St. Clair In The Great Lakes Region	Accepted, Journal of Environmental Quality
3	Mohammad Madani, Rajesh Seth, Luis F. Leon, Reza Valipour, Craig McCrimmon	Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach.	Published, Journal of Great Lakes Research

I certify that I have obtained written permission from the copyright owner(s) to include the above papers in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

II. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Lake St. Clair is a freshwater lake in the Lake Huron to Erie corridor in the Great Lakes Basin. Millions of people in Canada and the United States rely on that water source for drinking, fishing and recreational purposes. Lake St. Clair's watershed is heavily impacted by human activity, which can result in contamination of its waters by fecal matter of human or animal origin containing waterborne pathogens, and thus pose a direct threat to human health. Common sources of such pollution include combined sewer overflows, wastewater treatment plant bypasses, and agricultural application of manure derived from animal fecal waste. Several such sources are present in Windsor Essex County (WEC), Ontario, Canada, which is located along the southern edge of Lake St. Clair. Two popular public beaches and drinking water intakes are located in the nearshore region adjacent to the southern edge. Fecal microbial pollution is currently monitoring using fecal indicator bacteria (FIB), such as *Escherichia coli* (*E. coli*). Monitoring methods have several limitations including their inability to predict water quality in real-time or in advance, or to identify potential sources of contamination for more effective management. Mathematical models are tools that can be very effective and complementary to monitoring in overcoming its limitations. Model predictions can be real-time or near real-time and also help to identify or exonerate potential sources of microbial pollution.

In the current study, two types of modelling approaches that are commonly being used in the assessment of microbial contamination in beach waters and lakes were investigated: statistical modelling based on multiple linear regression (MLR) and hydrodynamic-ecological modelling. The statistical MLR models developed for Sandpoint Beach in Lake St. Clair showed higher accuracy in the range 64-78%, for predicting both exceedance and non-exceedance of the applicable standard, as compared to 54% accuracy obtained using the current method based on *E. coli* measurements. Amongst the MLR models developed, an increase of about 5-14% in model performance was observed when qualitative sky weather condition was included.

Results with mechanistic structured grid high-resolution AEM3D model developed for Lake St. Clair showed that four major tributaries (Thames, Sydenham, St. Clair and Clinton River) are unlikely to be responsible for the *E. coli* exceedances of provincial guideline

observed at Sandpoint Beach. Amongst the major tributaries, predicted *E. coli* concentrations were dominated by the contribution of St. Clair River for most of Lake St. Clair. The maximum predicted *E. coli* concentration from the combined input of the major tributaries was less than 100 CFU/100 ml for most of the lake and less than 10 CFU/100 ml at Sandpoint Beach. Predicted *E. coli* were significantly affected by varying water temperature and sunlight result in the temporal and diurnal dynamics of microbial water quality in Lake St. Clair. About 12–148% differences in predicted *E. coli* concentrations were observed at six drinking water intakes located in Lake St. Clair when time-variable decay rates were used instead of a constant decay rate. Also, on average nighttime *E. coli* predictions were 21–68% higher at these water intakes, as compared to daytime levels.

Results from the AEM3D model showed that while the flow contribution of eight smaller tributaries in Windsor Essex Region to the lake is insignificant (less than 0.2%), their contribution to the adjacent nearshore region along the southern edge of Lake St. Clair could be quite significant. Within about 1 km from the shoreline of this nearshore region, flow contributions from the small tributaries were estimated in the range between 18-35%, while their contribution to *E. coli* concentration was estimated to be more than 80%. Results with mechanistic unstructured grid TUFLOW-FV/AED2+ lakewide model and with a finer mesh nested model over a 2 km region surrounding Belle River showed differences of up to a factor of four in predicted *E. coli* concentrations at adjacent Lakeview Park West Beach (LP Beach). The differences reduced to a factor of up to 1.3 at nearby Lakeshore WTP intake located about one km away from shore. While the average contribution of the Belle River to *E. coli* concentrations at Lakeshore WTP intake was predicted to be <20%, the contribution increased to >80% when higher concentrations (10-35 CFU/ 100 ml) were predicted. The results also indicate that the construction of the marina may have contributed to some increase in *E. coli* concentrations at LP Beach from the external sources considered. However, construction of a new 150 m jetty in 2018, in place of the 25 m jetty separating Belle River from LP Beach, is expected to reduce the *E. coli* concentrations at LP Beach from the same sources by about 80%.

DEDICATION

پێشکەش بە خۆشەویستم سەمیرا، و کورە نازدارەکانم رایان و ڕێبین

پێشکەش بە دایکم و باوکم

To my love Samira, and my dearest sons Ryan and Rebin

To my parents

Saturday, June 6, 2020

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God for providing me courage, means and blessings to complete my Ph.D. successfully.

I thank my parents and all my family members who have sincerely encouraged and supported me all through. The caring and spirited support that I always receive from my lovely wife, Samira, and my son, Ryan, is immeasurable. The fact that my expecting son, Rebin, motivates me to finish my Ph.D. on time was undeniable. I also thank my parents for their prayers and generous blessings always. All my siblings and other members of my family also deserve my heartfelt thanks.

I sincerely acknowledge my advisor, Dr. Rajesh Seth, for his incredibly valuable mentorship and guidance that has enabled me to accomplish this endeavour. Many thanks for all his kind words, and fatherly advice that I will never forget during my entire life. I thank him for sharing his knowledge and wisdom along with providing adequate technical and financial resources during this journey.

I also take this opportunity to sincerely thank my co-advisor, Dr. Luis F. Leon. I acknowledge him genuinely for being a constant source of inspiration and supporting my work in every aspect. His immense experience in hydrodynamic and water quality modelling has taught me the finer details of different technical and professional skills. I would like to thank both my research collaborators, Dr. Reza Valipour and Mr. Craig McCrimmon for sparing time to minutely review each of my manuscripts and provide strong and valuable feedbacks. Their down-to-earth nature and inspiring words have enabled me to sail through the difficult stages.

I deeply appreciate the feedback provided and patience shown by my committee members Dr. Ram Yerubandi, Dr. Rupp Carriveau, Dr. Jill Crossman and Dr. Tirupati Bolisetti in enhancing the dissertation quality.

I am grateful to the Windsor-Essex County Health Unit (WECHU) in particular Karen Lukic, Philip Wong, Ramsey D'Souza and Victoria Peczulis for providing data and useful feedbacks on Chapter 2. Special thanks to Karina Richters from the City of Windsor for providing us with the high-frequency rainfall data. I am grateful to the Ontario Ministry of

the Environment and Climate Change (MOECC) for the ADCP data that was used for model validation in Chapters 3-6. I thank Environment and Climate Change Canada (ECCC) and Canada Centre for Inland Waters for their support and special thanks to Dr. Tom Edge for providing *E. coli* monitoring data for Thames River and Lake St. Clair used in Chapters 5-6. A special thanks to Katie Stammler from Essex Region Conservation Authority (ERCA) for all her support during the project and providing *E. coli* data for small tributaries. I am very grateful to Eric Anderson from National Oceanic and Atmospheric Administration (NOAA) for providing the bathymetry data of Lake St. Clair and useful comments of availability of water temperature data. I also want to thank Nicole Drumm, Erin Carroll and Kelli Smith from St. Clair Region Conservation Authority and Lilly Snobelen from Chatham-Kent PUC to provide data that helped us in *E. coli* estimation for the Sydenham and St. Clair River. A special thanks to Mr. Mitchell Smith from BMT, developer of TUFLOW-FV software, and Dr. Matthew Hipsey and his team in helping me set up the AED2+ pathogen module and all their related support during the project.

My other fellow graduate students Abdolrazagh Hashemi Shahraki, Vinod Chilkoti, Dylan Verburg, Saranya Jeyalakshmi, Monika Saha, and Israt Jahan also deserve my thanks for holding several technical discussions.

I offer my gratitude to the Ontario Trillium Scholarship and Canada's Natural Sciences and Engineering Research Council (NSERC) Strategic Project Grant (SPG) for providing financial support to my research work. I conclude by thanking everyone who has helped me directly or indirectly for any part of my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS/SYMBOLS

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
ADCP	Acoustic Doppler Current Profilers
AED2+	Aquatic Eco Dynamics
AEM3D	Aquatic Ecosystem Model
AGI	Acute Gastrointestinal Illness
AIC	Akaike's Information Criterion
ANN	Artificial Neural Networks
ANOVA	Analysis Of The Variance
AUROC	Area Under the Receiver Operating Characteristic Curve
BWS	Bathing Water Standard
CAD	Canadian Dollar
CAEDYM	Computational Aquatic Ecosystem Dynamics Model (
CFL	Courant–Friedrichs–Lewy
CFU	Colony Forming Units
CLSM4	A buoy name in Lake St. Clair (*42.471 N 82.877 W*),
CRCM5	Canadian Regional Climate Model
CWR	Centre for Water Research
E. coli	*Escherichia coli*
ECCC	Environment and Climate Change Canada
ELCOM	Estuary and Lake Computer Model
ENT	Enterococci
ERCA	Essex Region Conservation Authority
FDM	Finite Difference Method
FEM	Finite Element Method
FIB	Fecal Indicator Bacteria
Fn	Fourier Norm
GLERL	Great Lakes Environment Research Laboratory
GLSEA	Great Lakes Surface Environmental Analysis
LP Beach	Lakeview Park West Beach
LS_WTP	Lakeshore Water Treatment Plant
LSCM4	A buoy name in Lake St. Clair (*42.465 N 82.755 W*)
M series	Models without qualitative weather data
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
MW series	Models with Qualitative Weather Data Added To M Series
NCEI	National Centres for Environmental Information
NDBC	National Data Buoy Center
NE	Northeast
NGS	Next-Generation Sequencing
NOAA	The National Oceanic and Atmospheric Administration
NRMSE	Normalized Root Mean Square Error
NW	Northwest
OMHLTC	Ontario Ministry of Health and Long-Term Care

PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PM	Persistence Models
PWQMN	Provincial (Stream) Water Quality Monitoring Network
qPCR	Quantitative Polymerase Chain Reaction
R^2	Coefficient Of Determination
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RSD	Relative Standard Deviations
SD	Standard Deviation
SE	Southeast
SEM	Standard Error Of The Mean
SMS	Surface Modeling System
SP Beach	Sandpoint Beach
SP_WTP	Stoney Point Water Treatment Plant
ST	Small Tributaries
SW	Southwest
TBBL	Turbulent Benthic Boundary Layer
UQAM	University of Quebec At Montreal
USD	United State Dollar
USEPA	United States Environmental Protection Agency
USGS	The United States Geological Survey
VIF	Variance Inflation Factor
W series	Models With Qualitative Weather Data
WEC	Windsor Essex County
WECHU	Windsor Essex County Health Unit
WMO	World Meteorological Organization
WS	Skill Performance Value
WTP	Water Treatment Plant
WWTP	Wastewater Treatment Plant

CHAPTER 1 GENERAL INTRODUCTION

1.1 Background

Microbial contamination in the Great Lakes is of particular health concern to more than 30 million people that use its waters for drinking and recreational use. Swimming in contaminated waters may result in gastrointestinal and respiratory diseases caused by pathogens. The most common waterborne pathogens, that include bacteria, viruses, and protozoa, originate from human and animal fecal matter. The risk to get acute gastrointestinal illness (AGI) after swimming in recreational water is about 3-8% (Sanborn and Takaro, 2013). Crowded beaches have higher rates of illness and children are at higher risk as they usually play in the sand and highly contaminated water in the shoreline. In addition to the health effects, there are significant direct and indirect economic impacts associated with fecal pollution of water resources. Activities such as a day trip to a beach can be a significant source of revenue for communities during summer months. Annually, lost revenue in the range of 11.3M to USD 117M is reported from the Great Lakes recreational swimmers for those days when swimming is banned (Rabinovici et al., 2004; Shaikh, 2006). In another study in Ontario, a cost of CAD 1,089 per case of AGI was estimated when over the counter medications, lost patient or parental work time, and costs to the health care provider were included (Majowicz et al., 2006).

1.1.1 Pathogens and Fecal Indicators

The main health risk associated with exposure to recreational water quality hazards is an infection as a result of contact with pathogenic microorganisms. Many kinds of pathogens might end up appearing in water from different sources. Detection of waterborne fecal pathogens is very difficult and costly. For each type of bacterium, virus or protozoan a

different test is required. Hence, it is impractical to monitor water quality for every pathogen on a routine basis. Since most of these pathogens originate from human or animal fecal matter, a more practical alternative is to identify and monitor a more universal marker for fecal pollution. *Escherichia coli* (*E. coli*) and Enterococci fit that bill and are included in the group referred to as fecal indicator bacteria (FIB). These two FIBs are commonly used for routine monitoring of fecal pollution in natural waters (Health-Canada, 2012; USEPA, 2012; WHO, 2017). These indicators are groups of bacteria that are abundant in human and warm-blood animal feces and it is assumed that their concentration is correlated with that of major pathogens. Statistically significant relationships for the probability of occurrence of waterborne pathogens and FIBs have been reported (Payment et al., 2000). Epidemiological studies show that low levels of indicator bacteria such as *E. coli* in beach waters are sufficiently predictive of low gastrointestinal illness rates to be useful in helping to protect the health of beachgoers (Wade et al., 2003).

1.1.2 Guidelines for Recreational and Source Drinking Water Quality

To protect bathers from unexpected adverse water quality due to swimming in polluted water, microbial water quality is typically monitored for FIB by local beach managers. Recreational water quality generally falls under provincial and territorial jurisdiction (Health-Canada, 2012). To maximize beach safety, it is important to monitor water quality routinely and notify users of potential hazards as soon as possible. To test the water, a minimum of five sampling points for each beach should be identified for the beaches with a length of 1 km or less. In Ontario, Canada, the Ontario Ministry of Health and Long-Term Care sets out geometric mean *E. coli* concentration (minimum of five samples) of 100 colony-forming unit (CFU)/100 ml, updated to 200 CFU/100 ml in 2018 and a single-

sample maximum concentration ≤ 400 CFU/100 ml as a safe limit for swimming and bathing at freshwater beaches (Health-Canada, 2012; OMHLTC, 2018). Closures are issued for beaches that have *E. coli* levels ≥ 1000 CFU/100 ml.

In the case of drinking water quality, the highest-priority guidelines are those dealing with microbiological contaminants. Protecting sources of municipal drinking water such as lakes, rivers and well water is important to ensure clean, safe and sustainable drinking water according to the Clean Water Act, 2006 as part of the multi-barrier approach (Health-Canada, 2020). An understanding of fecal contamination occurrence in source waters is essential. It facilitates the selection of the best location with the highest-quality water for drinking-water supply during the construction of water treatment plants. Besides, the determination of the fecal concentrations in the source waters provides a basis for establishing adequate treatment requirements to meet health-based targets within the water safety plan (Ivey et al., 2006).

1.1.3 Monitoring Program Challenge and Limitation

Although *E. coli* is considered the most suitable indicator for recreational water monitoring according to Guidelines for Canadian Recreational Water Quality, its measurement takes 18 to 24 h before results are available. However, water quality in nearshore regions can change over a matter of hours (Boehm et al., 2002; Myers et al., 1998) so concentrations may change between the time of sampling and the reporting of results. Unsafe conditions are frequently announced late due to latencies in the *E. coli* measurement process. This process results in issuing closures based on the previous day's data rather than current water conditions.

Recent advances in molecular biochemistry, genetics, and imaging technology have set the stage for newer and faster methods to complement or replace the growth-based approaches. Developing rapid analytical methods such as quantitative polymerase chain reactions (qPCR) with an average of two hours' laboratory analytical time is a possible solution in which more timely knowledge of water contamination issues would be available to help prevent exposure-related illness. Although these methods may be operationally available, they need higher analytical costs than slower culture-based methods (Setty, 2012). Also, they require a site-specific evaluation before implementation (Campbell and Kleinheinz, 2020). Additional obstacles include requirements for equipment, laboratory and field labour, and limitations in identifying potential sources of fecal pollution that can help with management.

1.1.4 Source of Fecal Contamination

Microbial pollution in a lake can come from a variety of point and non-point sources (Byappanahalli et al., 2015). These include storm drains and rivers, septic system failures, point discharges of treated and untreated sewage, combined sewage overflow (CSO), sanitary sewer overflows (SSO), domesticated animals (manure spreading), birds and wildlife, and direct inputs from swimmers. The main source of fecal contamination in storm drains and rivers are determined by the watershed land use. On the Canadian side of Lake St. Clair, the land use is predominantly agricultural areas and in the western part of the lake, Detroit City is located, hence, the main source of *E. coli* is urban sources such as urban runoff from streets, sidewalks and parking lots (Molina et al., 2014).

1.1.5 Modelling Approaches

Although monitoring for FIBs can be a solution for the detection of the pathogen, it is still unfeasible to experimentally monitor their levels at the high spatiotemporal resolution which is often required in real applications. Therefore, it is an increasingly necessary to combine the FIB measurement with models. These models provide insights for contamination and nutrient tracking (Leon et al., 2012), source tracking (Nevers et al., 2015; Sokolova et al., 2012), nowcasting and real-time forecasting (Boehm et al., 2007; Chan et al., 2013; Francy, 2009; Nevers and Whitman, 2005; Park et al., 2018), dispersion and diffusion (Bonamano et al., 2015; Bravo et al., 2017; Thupaki et al., 2010; Valipour et al., 2018), risk management (Hipsey et al., 2004; Thoe et al., 2015), and examination of different scenarios (Gao et al., 2015; Ge et al., 2012; Vlazaki et al., 2019). Modelling of FIBs could play vital roles in recreational and drinking water management. This includes (i) identifying and quantifying dominant processes or parameters that control fecal contamination dynamics (ii) helping to identify, quantify and track the major sources (iii) assessing the likely impacts of anthropogenic activities on microbial water quality, (iv) examining of different scenarios and natural events such as extreme rainfall, climate change, future catchment management scenarios, etc., as well as in (v) providing a real-time decision-making platform that can be used to protect public health in bathing sites (de Brauwere et al., 2014).

To predict FIB concentration in surface waters two different approaches and models are commonly being used and applied: Statistical models (also known as data-based, regression-based, empirical, or black-box models) and mechanistic models which are also referred to as reactive tracer, process-based, deterministic, physics-based, physiological or

causal models (de Brauwere et al., 2014; Gourmelon et al., 2010; Guisan and Zimmermann, 2000).

1.1.5.1 Statistical Models:

These models link environmental parameters such as rainfall, wind or sunlight and water quality variables such as water temperature, turbidity etc. to FIB concentrations, mostly as fecal coliforms or *E. coli* (Francy et al., 2013; Gourmelon et al., 2010; USEPA, 2010, 2016). In such models, rather than designing equations to reflect a conceptual understanding of the system, the observational data are analyzed to find patterns and relationships that can be used to make predictions, regardless of the causative processes involved (Robson, 2014). They are also known as extrinsic models because any knowledge or hypotheses about the inner structural connectivity of the system are not incorporated in statistical or empirical models (Vlazaki et al., 2019).

Statistical models have been widely used in several studies globally to predict *E. coli* concentration in swimming beaches (Dada and Hamilton, 2016; Feng et al., 2015; Francy et al., 2013; Nevers and Whitman, 2011; Olyphant and Whitman, 2004; Simmer, 2016). In the late 1990s The simplest type of model, namely Rainfall-based alerts, has been used by communities for several years (Francy, 2009). Later, application of predictive models based on variables such as water chemistry parameters and meteorological data are proposed as alternatives or adjuncts to sampling to obtain more timely information (Nevers and Whitman, 2011; S Francy et al., 2013; Simmer, 2016; Zhang et al., 2012). An increased concern for the accurate characterization and forecasting of recreational water quality has created the need for more sophisticated statistical models based on Artificial Neural Networks (ANN) (Jin and Englande, 2006; Mavani et al., 2014), Wavelet (Zhang et al.,

2018), Bayesian Networks (Bertone et al., 2019), and an ensemble of statistical models with the mechanistic model (Hellweger, 2007).

1.1.5.2 Mechanistic Models:

Although statistical models are fast and flexible in the prediction of FIBs and easier to develop with limited data and resources, their limitations include the inability to distinguish between inputs or consider pollution reduction due to advection, transport, or microbial decay rate to provide spatial and temporal distribution. This has spurred interest in development of more realistic models that can describe the processes involved in fate and transport of the nutrients, contaminations and microorganisms (Allan et al., 2018; Oveisy et al., 2014; Paturi et al., 2015; Sokolova et al., 2012). In general, hydrodynamic models provide a broad knowledge about the flow condition, current, water level and scalar mixing by numerically solving the Navier–Stokes equations. Apart from being important in itself, hydrodynamic models also provide a basis to simulate sediment transport, particle tracking and also can be integrated with biochemical and ecological models that describe the fate of microorganisms via the growth and decay of bacteria/viruses depending on environmental conditions (light, temperature and sediment) (Gourmelon et al., 2010; Hipsey et al., 2008; Trolle et al., 2012). Application of hydrodynamic and water quality models is widely used for lakes, reservoirs, rivers, estuaries, and coastal zones to evaluate aquatic management strategies (Romero et al., 2004; Zhang et al., 2017).

Hydrodynamic models have expanded from simplistic tidally averaged one-dimensional models (Luo et al., 2018; Prokopkin et al., 2010; Romero et al., 2004) to fully three-dimensional (3D) models that can resolve wetting and drying, wave-current interaction, and sediment transport (Bocaniov and Scavia, 2018; Chen and Liu, 2017; Hipsey et al.,

2004; Huang et al., 2010; Oveisy et al., 2014; Valipour et al., 2018). Various numerical schemes have been used to solve governing equations of flow dynamics, of which finite difference, finite element or finite volume approaches are the most popular ones (Martyr-Koller et al., 2017). Finite difference methods, that use regular i.e., orthogonal or non-orthogonal curvilinear grids, are sometimes considered to have higher accuracy and efficiency compared with finite element methods. However, unstructured meshes are more suitable for irregular and more complex lake geometries and provide a better domain resolution than regular grids (Martyr-Koller et al., 2017; Morales Marín, 2013). The finite volume methods have advantages over both FDM and FEM as they have combined the best attributes of finite differences (e.g. simple discrete computational efficiency) and finite-elements (geometric flexibility) (Chen et al., 2003). Inclusion of increased complexity and resolution that allowed for the development of more robust mixing routines, higher-resolution domains, and inclusion of detailed circulation processes, has been made possible by the increases in computational resources with time (Ganju et al., 2016). Although mechanistic models require a significant investment of time and resources for development, they are expected to be much more useful and cost-effective in the planning, design, and implementation of long-term assessment and management strategies for surface water resources (de Brauwere et al., 2014).

1.2 Thesis Objectives

The main objective of this dissertation was to develop predictive modelling tools for microbial water quality in Lake St. Clair. The tools were then applied to simulate the temporal change in microbial water quality and identify the relative contribution of major contributing sources at selected public beach locations and drinking water intakes. The

developed models are expected to be useful, complementary and cost-effective additions to monitoring efforts in microbial source tracking, to help with better beach management, reduce economic loss due to beach closures, and reduce risk to human health from drinking water or recreational waters use.

The current dissertation is organized into seven chapters, five of which are original standalone research contributions prepared for referred scientific literature, and presented in Chapters 2 to 6. The dissertation begins with this chapter that provides the background literature on the topic of this dissertation. Chapter 7 summarizes the findings from this dissertation and its engineering significance. The references pertinent to each chapter are listed at the end of the respective chapters.

1.2.1 Chapter 2

Sandpoint Beach is a popular public beach on Lake St. Clair in the City of Windsor, Ontario, Canada. Recreational use advisories for the beach are currently issued based on the microbial water quality (*E. coli*) levels in samples collected two days prior. Complementing *E. coli* monitoring data with easily available or measurable weather and water quality data to develop a near real-time ("NowCast") predictive model for *E. coli* using statistical techniques such as multiple linear regression (MLR) have shown promising results (Dada and Hamilton, 2016; Feng et al., 2015; Francy, 2009; Francy et al., 2013; Kato et al., 2019; Nevers and Whitman, 2005).

Using intensive (5 days per week) water quality and *E. coli* monitoring data collected by the Windsor-Essex County Health Unit (WECHU) at Sandpoint Beach over five years (2014 – 2018) during the summer months, the main aim of Chapter 2 was to conduct a comprehensive assessment of the available data to develop and evaluate multiple MLR

predictive *E. coli* models for Sandpoint Beach. Variables considered include a commonly recorded qualitative weather sky condition information and varying training datasets, which have previously not been included in such modelling efforts. A new approach for integrating multiple performance criteria using Principal Coordinate Analysis (PCoA) for a better and more holistic comparison of the performance of the MLR models for model selection was developed and tested. This research has been accepted for publication in the Journal of Environmental Quality.

1.2.2 Chapter 3

Lake St. Clair is receiving discharges from four major tributaries. While St. Clair River dominates in terms of flow contribution (>98%), the other tributaries (Thames River, Sydenham River and Clinton River) are expected to be a significant contributor to fecal source microbial pollution. It is expected that these major tributaries dominate the microbial water quality of Lake St. Clair but there is less understanding about their extent and potential area of their influence. The impact of these major tributaries on microbial water quality of lake St. Clair was investigated in Chapter 3. Although there have been some studies that examined hydrodynamic and/or nutrient transport modelling (Anderson et al., 2010; Bocaniov and Scavia, 2018; Healy et al., 2007; Holtschlag and Koschik, 2002), to the best of our knowledge there is no study on 3D microbial water quality modelling for Lake St. Clair.

The main objective of Chapter 3 was to evaluate the contributions of major tributaries to fecal microbial pollution of Lake St. Clair and Sandpoint Beach, located on the southern shore. A set of hydrodynamic models with different resolutions, varied from 400 m to 100 m and a nested model with 50 m uniform grid size, were developed and compared using

Aquatic Ecosystem Model (AEM3D) framework. High-resolution 3D hydrodynamic models were developed and validated with field measurements of water temperature and velocity profiles. *E. coli* was modelled as a tracer. Microbial water quality of Lake St. Clair was assessed using backward particle tracking, water age analysis, and study of the potential area of influence. The research has been published in the Journal of Great Lakes Research.

1.2.3 Chapter 4

Lake St. Clair is a source of drinking water to more than 4.5 million people in both the US and Canada. Grosse, Mount Clemens, New Baltimore and Ira Township water treatment plants (WTPs) supply drinking water to residents in the US, while Lakeshore and Stoney Point WTPs supply drinking water to residents in Canada. To be able to prevent waterborne disease outbreaks caused by fecally contaminated drinking water, an accurate assessment of the contribution from different sources to the total fecal contamination at the raw water intake of a drinking water treatment plant is needed (Sokolova et al., 2013). Despite its importance, a comprehensive evaluation of potential sources of fecal contamination and their impacts on drinking water sources in Lake St. Clair is currently lacking. AEM3D modelling framework has the capabilities that can allow for a time-dependent inactivation rate based on temperature, solar insolation and water chemical properties- provided data is available to parameterize the decay model. In Chapter 4 the influence of the four major tributaries (St. Clair, Thames, Clinton, and Sydenham rivers) on the temporal and diurnal microbial water quality at six drinking water intakes was studied. The fate of *E. coli* was modelled using two approaches, i) constant decay rate based on average conditions, and ii) variable decay rate based on water temperature and solar radiation.

1.2.4 Chapter 5

Eight small but seasonally significant tributaries of widely varying drainage areas and hydrologic features discharge into Lake St. Clair along the southern shores. The flow and pollution load from these small tributaries are expected to be negligible in the context of the larger lake, and therefore have been ignored or not have been studied comprehensively in previous hydrodynamic and water quality modelling studies of Lake St. Clair (Anderson et al., 2010; Bocaniov and Scavia, 2018; Healy et al., 2007; Holtschlag and Koschik, 2002; Madani et al., 2020). Chapter 5 focuses on the impact of these local tributaries on the southern edge of the lake. The hydrodynamic component of AEM3D which was developed in Chapters 3 and 4, is coupled with the Computational Aquatic Ecosystem DYnamics Model (CAEDYM) water quality modules to simulate *E. coli* concentration in lake-wide and in the southern shoreline of Lake St. Clair. The relative contribution of the eight small tributaries to *E. coli* concentration and to flow (defined as the relative quantity of water received from each tributary) and the area of their influence along the southern edge of Lake St. Clair on the microbial water quality was studied during summer 2016.

1.2.5 Chapter 6

Chapter 5 results showed that in the nearshore region of approximately ~40 km stretch along the southern shore where the eight tributaries discharge, the contribution of the small tributaries to the *E. coli* concentration was quite different and dominated by one or more of the eight small tributaries. There are reasonable pieces of evidence from our earlier studies to expect that Belle River contributes significantly to the load of *E. coli* delivered to the nearshore of Lake St. Clair. However, there has been a little advance over earlier work in understanding the temporal and spatial degree to which contamination from Belle

River contributes to the level of fecal indicator bacteria observed in Lakeview Park Beach and Lakeshore water treatment plant intakes.

Due to the presence of the Marina which is protected by two breakwaters in the west and east side at the immediate adjacent of the Belle River, the geometry of the location is complex. Complex bathymetry and geomorphic layout have a noticeable effect on the hydrodynamic (flow circulation) and contamination transport, thus should be represented with an appropriate computational mesh (Ganju et al., 2016). A model with an unstructured grid is preferred over the uniform grid because it can accurately resolve the bathymetry in the complex shoreline without the need to build fine grids to the overall domain and hence it has a computational advantage over regular grids (Chen et al., 2003; Martyr-Koller et al., 2017).

The main objective of Chapter 6 was to develop an unstructured high-resolution model (mesh size of about 50-800 m) using the TUFLOW-FV framework and coupled with the pathogen module from Aquatic Eco-Dynamics model (AED2+). This model was used as the base for boundary condition of a nested model with a mesh resolution between 5-10 meters at a zone stretching 4 km alongshore and roughly 2 km onshore-offshore around Belle River mouth. The relative contribution of the Belle River to the microbial water quality at Lakeview Park Beach and Lakeshore water treatment plant was assessed. Several scenarios that reflect the condition during the simulation period (summer 2016) and the recent change in the beach area, construction of a 150-meter jetty, were examined and compared.

1.3 References

Allan, M.G., Hamilton, D.P., Muraoka, K., 2018. A coupled hydrodynamic-ecological model to test management options for restoration of lakes Onoke and Wairarapa, ERI report. Environmental Research Institute, Faculty of Science and Engineering, The University of Waikato: Hamilton, New Zealand.

Anderson, E.J., Schwab, D.J., Lang, G.A., 2010. Real-Time Hydraulic and Hydrodynamic Model of the St. Clair River, Lake St. Clair, Detroit River System. Journal of Hydraulic Engineering 136(8) 507-518.

Bertone, E., Purandare, J., Durand, B., 2019. Spatiotemporal prediction of Escherichia coli and Enterococci for the Commonwealth Games triathlon event using Bayesian Networks. Marine Pollution Bulletin 146 11-21.

Bocaniov, S.A., Scavia, D., 2018. Nutrient Loss Rates in Relation to Transport Time Scales in a Large Shallow Lake (Lake St. Clair, USA—Canada): Insights From a Three-Dimensional Model. Water Resources Research 54(6) 3825-3840.

Boehm, A.B., Grant, S.B., Kim, J.H., Mowbray, S.L., McGee, C.D., Clark, C.D., Foley, D.M., Wellman, D.E., 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. Environmental Science & Technology 36(18) 3885-3892.

Boehm, A.B., Whitman, R.L., Nevers, M.B., Hou, D., Weisberg, S.B., 2007. Nowcasting Recreational Water Quality, In: (Ed.), L.J.W. (Ed.), Statistical Framework for Recreational Water Quality Criteria and Monitoring, pp. 179-210.

Bonamano, S., Madonia, A., Borsellino, C., Stefani, C., Caruso, G., De Pasquale, F., Piermattei, V., Zappala, G., Marcelli, M., 2015. Modeling the dispersion of viable and total Escherichia coli cells in the artificial semi-enclosed bathing area of Santa Marinella (Latium, Italy). Mar Pollut Bull 95(1) 141-154.

Bravo, H.R., McLellan, S.L., Klump, J.V., Hamidi, S.A., Talarczyk, D., 2017. Modeling the fecal coliform footprint in a Lake Michigan urban coastal area. Environmental Modelling & Software 95 401-419.

Byappanahalli, M.N., Nevers, M.B., Whitman, R.L., Ge, Z., Shively, D., Spoljaric, A., Przybyla-Kelly, K., 2015. Wildlife, urban inputs, and landscape configuration are responsible for degraded swimming water quality at an embayed beach. Journal of Great Lakes Research 41(1) 156-163.

Campbell, A., Kleinheinz, G., 2020. Comparison of qPCR and traditional beach monitoring methods at select Great Lakes beaches and the impact on beach management. Lake and Reservoir Management 1-14.

Chan, S.N., Thoe, W., Lee, J.H., 2013. Real-time forecasting of Hong Kong beach water quality by 3D deterministic model. Water Research 47(4) 1631-1647.

Chen, C., Liu, H., Beardsley, R.C., 2003. An Unstructured Grid, Finite-Volume, Three-Dimensional, Primitive Equations Ocean Model: Application to Coastal Ocean and Estuaries. Journal of Atmospheric and Oceanic Technology 20(1) 159-186.

Chen, W.-B., Liu, W.-C., 2017. Investigating the fate and transport of fecal coliform contamination in a tidal estuarine system using a three-dimensional model. Marine Pollution Bulletin 116(1) 365-384.

Dada, A.C., Hamilton, D.P., 2016. Predictive Models for Determination of E. coli Concentrations at Inland Recreational Beaches. Water, Air, & Soil Pollution 227(9) 347.

de Brauwere, A., Ouattara, N.K., Servais, P., 2014. Modeling fecal indicator bacteria concentrations in natural surface waters: a review. Critical Reviews in Environmental Science and Technology 3389(March 2015) 140225123512007-140225123512007.

Feng, Z.X., Reniers, A., Haus, B.K., Solo-Gabriele, H.M., Wang, J.D., Fleming, L.E., 2015. A predictive model for microbial counts on beaches where intertidal sand is the primary source. Marine Pollution Bulletin 94(1-2) 37-47.

Francy, D.S., 2009. Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches. Aquatic Ecosystem Health & Management 12(2) 177-182.

Francy, D.S., Brady, A.M.G., Carvin, R.B., Corsi, S.R., Fuller, L.M., Harrison, J.H., Hayhurst, B.A., Lant, J., Nevers, M.B., Terrio, P.J., Zimmerman, T.M., 2013. Developing and implementing predictive models for estimating recreational water quality at Great Lakes beaches: U.S. Geological Survey Scientific Investigations Report 2013–5166, 68 p., https://doi.org/10.3133/sir20135166.

Ganju, N.K., Brush, M.J., Rashleigh, B., Aretxabaleta, A.L., del Barrio, P., Grear, J.S., Harris, L.A., Lake, S.J., McCardell, G., O'Donnell, J., Ralston, D.K., Signell, R.P., Testa, J.M., Vaudrey, J.M.P., 2016. Progress and Challenges in Coupled Hydrodynamic-Ecological Estuarine Modeling. Estuaries and Coasts 39(2) 311-332.

Gao, G., Falconer, R.A., Lin, B., 2015. Modelling the fate and transport of faecal bacteria in estuarine and coastal waters. Marine Pollution Bulletin 100(1) 162-168.

Ge, Z., Whitman, R.L., Nevers, M.B., Phanikumar, M.S., Byappanahalli, M.N., 2012. Nearshore hydrodynamics as loading and forcing factors for Escherichia coli contamination at an embayed beach. Limnology and Oceanography 57(1) 362-381.

Gourmelon, M., Lazure, P., D., H.-H., Saux, J.C., Caprais, M.P., Guyader, F.S., Catherine, M.H., Pommepuy, M., 2010. Microbial modelling in coastal environments and early warning systems: useful tools to limit shellfish microbial contamination.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135(2) 147-186.

Health-Canada, 2012. Guidelines for Canadian Recreational Water Quality, Third Ed., Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, (Catalogue No H129-15/2012E).

Health-Canada, 2020. Guidelines for Canadian Drinking Water Quality: Guideline Technical Document — Escherichia coli. Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario. (Catalogue No. H129-27/2020E-PDF).

Healy, D.F., Chambers, D.B., Rachol, C.M., Jodoin, R.S., 2007. Water quality of the St. Clair River, Lake St. Clair, and their U.S. tributaries, 1946-2005, Scientific Investigations Report: Reston, VA.

Hellweger, F.L., 2007. Ensemble modeling of E. coli in the Charles River, Boston, Massachusetts, USA. Water Science and Technology 56(6) 39-46.

Hipsey, M.R., Antenucci, J.P., Brookes, J.D., 2008. A generic, process-based model of microbial pollution in aquatic systems. Water Resources Research 44(7).

Hipsey, M.R., Antenucci, J.P., Brookes, J.D., Burch, M.D., Regel, R.H., Linden, L., 2004. A three dimensional model of Cryptosporidium dynamics in lakes and reservoirs: A

new tool for risk management. International Journal of River Basin Management 2(3) 181-197.

Holtschlag, D.J., Koschik, J.A., 2002. A two-dimensional hydrodynamic model of the St. Clair-Detroit River waterway in the Great Lakes basin, Water-Resources Investigations Report: Lansing, MI.

Huang, A., Rao, Y.R., Lu, Y., Zhao, J., 2010. Hydrodynamic modeling of Lake Ontario: An intercomparison of three models. Journal of Geophysical Research: Oceans 115(C12).

Ivey, J., De Loe, R., Kreutzwiser, R., 2006. Planning for source water protection in Ontario. Applied Geography 26(3-4) 192-209.

Jin, G., Englande, A., 2006. Prediction of swimmability in a brackish water body. Management of Environmental Quality: An International Journal.

Kato, T., Kobayashi, A., Oishi, W., Kadoya, S.-s., Okabe, S., Ohta, N., Amarasiri, M., Sano, D., 2019. Sign-constrained linear regression for prediction of microbe concentration based on water quality datasets. Journal of Water and Health 17(3) 404-415.

Leon, L.F., Smith, R.E.H., Malkin, S.Y., Depew, D., Hipsey, M.R., Antenucci, J.P., Higgins, S.N., Hecky, R.E., Rao, R.Y., 2012. Nested 3D modeling of the spatial dynamics of nutrients and phytoplankton in a Lake Ontario nearshore zone. Journal of Great Lakes Research 38 171-183.

Luo, L., Hamilton, D., Lan, J., McBride, C., Trolle, D., 2018. Autocalibration of a one-dimensional hydrodynamic-ecological model (DYRESM 4.0-CAEDYM 3.1) using a Monte Carlo approach: simulations of hypoxic events in a polymictic lake. Geosci. Model Dev. 11(3) 903-913.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020. Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach. Journal of Great Lakes Research.

Majowicz, S., McNab, W., Sockett, P., Henson, S., Doré, K., Edge, V., Buffett, M., Fazil, A., Read, S., McEwen, S., 2006. Burden and cost of gastroenteritis in a Canadian community. Journal of food protection 69(3) 651-659.

Martyr-Koller, R.C., Kernkamp, H.W.J., van Dam, A., van der Wegen, M., Lucas, L.V., Knowles, N., Jaffe, B., Fregoso, T.A., 2017. Application of an unstructured 3D finite volume numerical model to flows and salinity dynamics in the San Francisco Bay-Delta. Estuarine, Coastal and Shelf Science 192 86-107.

Mavani, J., Chen, L., Joksimovic, D., Li, S., 2014. Development and testing of data driven nowcasting models of Beach water quality.

Molina, M., Hunter, S., Cyterski, M., Peed, L.A., Kelty, C.A., Sivaganesan, M., Mooney, T., Prieto, L., Shanks, O.C., 2014. Factors affecting the presence of human-associated and fecal indicator real-time quantitative PCR genetic markers in urban-impacted recreational beaches. Water Research 64 196-208.

Morales Marín, L., 2013. Numerical Modelling of Hydrodynamics and Sedimentation in Upland Lakes and Implications for Sediment Focusing. UCL (University College London).

Myers, D.N., Koltun, G.F., Francy, D.S., 1998. Effects of hydrologic, biological, and environmental processes on sources and concentrations of fecal bacteria in the Cuyahoga River, with implications for management of recreational waters in Summit

and Cuyahoga Counties, Ohio, Water-Resources Investigations Report: Columbus, OH, p. 56.

Nevers, M.B., Byappanahalli, M.N., Phanikumar, M.S., Whitman, R.L., 2015. Fecal Indicator Organism Modeling and Microbial Source Tracking in Environmental Waters, Manual of Environmental Microbiology, pp. 3.4.6-1-3.4.6-16.

Nevers, M.B., Whitman, R.L., 2005. Nowcast modeling of Escherichia coli concentrations at multiple urban beaches of southern Lake Michigan. Water Research 39(20) 5250-5260.

Nevers, M.B., Whitman, R.L., 2011. Efficacy of monitoring and empirical predictive modeling at improving public health protection at Chicago beaches. Water Research 45(4) 1659-1668.

Olyphant, G.A., Whitman, R.L., 2004. Elements of a predictive model for determining beach closures on a real time basis: The case of 63rd Street Beach Chicago. Environmental Monitoring and Assessment 98(1-3) 175-190.

OMHLTC, O.M.o.H.a.L.-T.C., 2018. Operational approaches for recreational water guideline. Queen's Printer for Ontario: Toronto, ON.

Oveisy, A., Rao, Y.R., Leon, L.F., Bocaniov, S.A., 2014. Three-dimensional winter modeling and the effects of ice cover on hydrodynamics, thermal structure and water quality in Lake Erie. Journal of Great Lakes Research 40 19-28.

Park, Y., Kim, M., Pachepsky, Y., Choi, S.-H., Cho, J.-G., Jeon, J., Cho, K.H., 2018. Development of a Nowcasting System Using Machine Learning Approaches to Predict Fecal Contamination Levels at Recreational Beaches in Korea. Journal of Environmental Quality 47(5) 1094-1102.

Paturi, S., Boegman, L., Bouffard, D., Rao, Y.R., 2015. Three-Dimensional Simulation of Lake Ontario North-Shore Hydrodynamics and Contaminant Transport. Journal of Hydraulic Engineering 141(3) 04014082.

Payment, P., Berte, A., Prévost, M., Ménard, B., Barbeau, B., 2000. Occurrence of pathogenic microorganisms in the Saint Lawrence River (Canada) and comparison of health risks for populations using it as their source of drinking water. Can J Microbiol 46(6) 565-576.

Prokopkin, I.G., Mooij, W.M., Janse, J.H., Degermendzhy, A.G., 2010. A general one-dimensional vertical ecosystem model of Lake Shira (Russia, Khakasia): description, parametrization and analysis. Aquatic Ecology 44(3) 585-618.

Rabinovici, S.J.M., Bernknopf, R.L., Wein, A.M., Coursey, D.L., Whitman, R.L., 2004. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. Environmental Science & Technology 38(10) 2737-2745.

Robson, B.J., 2014. When do aquatic systems models provide useful predictions, what is changing, and what is next? Environmental Modelling & Software 61 287-296.

Romero, J.R., Antenucci, J.P., Imberger, J., 2004. One- and three-dimensional biogeochemical simulations of two differing reservoirs. Ecological Modelling 174(1) 143-160.

S Francy, D., Stelzer, E., Duris, J., M G Brady, A., H Harrison, J., E Johnson, H., Ware, M., 2013. Predictive Models for Escherichia coli Concentrations at Inland Lake Beaches and Relationship of Model Variables to Pathogen Detection.

Sanborn, M., Takaro, T., 2013. Recreational water-related illness: office management and prevention. Canadian family physician Medecin de famille canadien 59(5) 491-495.

Setty, K., 2012. Adapting and adopting rapid molecular methods for beach water quality monitoring. Urban Coast 3:48- 53. .

Shaikh, S.L., 2006. Fact Sheet: The value of Chicago beaches, University of Chicago, Chicago, IL. .

Simmer, R.A., 2016. Source determination and predictive model development for Escherichia coli concentrations at F.W. Kent Park Lake, Oxford, Iowa.

Sokolova, E., Astrom, J., Pettersson, T.J., Bergstedt, O., Hermansson, M., 2012. Estimation of pathogen concentrations in a drinking water source using hydrodynamic modelling and microbial source tracking. J Water Health 10(3) 358-370.

Sokolova, E., Pettersson, T.J.R., Bergstedt, O., Hermansson, M., 2013. Hydrodynamic modelling of the microbial water quality in a drinking water source as input for risk reduction management. Journal of Hydrology 497 15-23.

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M.L., Boehm, A.B., 2015. Sunny with a Chance of Gastroenteritis: Predicting Swimmer Risk at California Beaches. Environmental Science & Technology 49(1) 423-431.

Thupaki, P., Phanikumar, M.S., Beletsky, D., Schwab, D.J., Nevers, M.B., Whitman, R.L., 2010. Budget Analysis of Escherichia coli at a Southern Lake Michigan Beach. Environmental Science & Technology 44(3) 1010-1016.

Trolle, D., Hamilton, D.P., Hipsey, M.R., Bolding, K., Bruggeman, J., Mooij, W.M., Janse, J.H., Nielsen, A., Jeppesen, E., Elliott, J.A., Makler-Pick, V., Petzoldt, T., Rinke, K., Flindt, M.R., Arhonditsis, G.B., Gal, G., Bjerring, R., Tominaga, K., Hoen, J.t., Downing, A.S., Marques, D.M., Fragoso, C.R., Søndergaard, M., Hanson, P.C., 2012. A community-based framework for aquatic ecosystem models. Hydrobiologia 683(1) 25-34.

USEPA, 2010. Predictive Tools for Beach Notification, Vol. I: Review and Technical Protocol, EPA-823-R-10-003. U.S. Environmental Protection Agency Office of Water Office of Science and Technology 61.

USEPA, 2012. Recreational Water Quality Criteria, In: 820-F-12-058, E. (Ed.): Office of Water, Washington, DC.

USEPA, 2016. Six Key Steps for Developing and Using Predictive Tools at Your Beach, EPA-820-R-16-001. U.S. Environmental Protection Agency Office of Water Office of Science and Technology available at: https://www.epa.gov/sites/production/files/2016-03/documents/six-key-steps-guidance-report.pdf.

Valipour, R., Rao, Y.R., León, L.F., Depew, D., 2018. Nearshore-offshore exchanges in multi-basin coastal waters: Observations and three-dimensional modeling in Lake Erie. Journal of Great Lakes Research.

Vlazaki, M., Huber, J., Restif, O., 2019. Integrating mathematical models with experimental data to investigate the within-host dynamics of bacterial infections. Pathogens and disease 77(8) ftaa001.

Wade, T.J., Pai, N., Eisenberg, J.N.S., Colford, J.M., Jr., 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environmental health perspectives 111(8) 1102-1109.

WHO, 2017. Guidelines for drinking-water quality: fourth edition incorporating the first addendum. World Health Organization: Geneva.

Zhang, J., Qiu, H., Li, X., Niu, J., Nevers, M.B., Hu, X., Phanikumar, M.S., 2018. Real-Time Nowcasting of Microbiological Water Quality at Recreational Beaches: A Wavelet and Artificial Neural Network-Based Hybrid Modeling Approach. Environmental Science & Technology 52(15) 8446-8455.

Zhang, L., Hipsey, M.R., Zhang, G.X., Busch, B., Li, H.Y., 2017. Simulation of multiple water source ecological replenishment for Chagan Lake based on coupled hydrodynamic and water quality models. Water Supply 17(6) 1774-1784.

Zhang, Z., Deng, Z., Rusch, K.A., 2012. Development of predictive models for determining enterococci levels at Gulf Coast beaches. Water Research 46(2) 465-474.

CHAPTER 2 EVALUATING MULTIPLE PREDICTIVE MODELS FOR BEACH MANAGEMENT AT A FRESHWATER BEACH ON LAKE ST. CLAIR IN THE GREAT LAKES REGION[1]

2.1 Chapter Synopsis

Recreational water quality is currently monitored at Sandpoint Beach on Lake St. Clair using culture-based enumeration of *Escherichia coli* (*E. coli*). Using water quality and weather data collected over four years, several multiple linear regression (MLR)-based models were developed for near-real-time prediction of *E. coli* concentration and were tested using independent data from the fifth year. Model performance was assessed by the determination of metrics such as root mean square errors, accuracy, specificity, sensitivity, and area under the receiver operating characteristic curve (AUROC). Each of the developed MLR models described herein resulted in increased correct responses, for both exceedance and non-exceedance of the applicable standard, as compared to predictions based on *E. coli* measurements (persistence models; using the previous day's *E. coli* concentration) which is the current method being used. AUROC values for persistence models are between 0.5 – 0.6, as compared to >0.7 for all the MLR models described herein. Among the MLR models, model performance improved when qualitative sky weather condition, which is commonly reported but not previously used in similar models, was included. To select the best model, a principal coordinate analysis was used to combine multiple model performance metrics and provide a more sensitive tool for model comparison. While models developed using two, three, and four years of monitoring data provided reasonable

[1] This Chapter was accepted to publication in Journal of Environmental Quality: Madani, M. and Seth, R. (2020), Evaluating multiple predictive models for beach management at a freshwater beach on lake St. Clair in the great lakes region. J. Environ. Qual., Accepted Author Manuscript https://doi.org/10.1002/jeq2.20107

performance, the model developed using the most recent two-year data was marginally better. Thus, data from the most recent two years is likely sufficient as a training dataset for updating the MLR model for Sandpoint Beach in the future.

2.2 Introduction

Annually, more than 120 million cases of gastrointestinal disease and more than 50 million cases of respiratory diseases are reported globally that are caused by swimming in waters contaminated with pathogenic fecal bacteria and viruses originating from domestic wastewater sources (Shuval, 2003). Furthermore, swimming beach closures can have substantial financial impacts. Annually, an estimated $11.3M to $117M USD in revenues are lost from Great Lakes recreational swimmers for those days when swimming is prohibited in error (Rabinovici et al., 2004). To protect bathers from swimming in polluted waters, microbial water quality is typically monitored using fecal indicator bacteria (FIB) such as *Escherichia coli*.

Although *E. coli* is considered the most suitable indicator organism for recreational water quality monitoring according to Guidelines for Canadian Recreational Water Quality (Health-Canada, 2012), its measurement using culture-based methods requires 18 to 24 hours before results are available. At beaches where high FIB levels are typically seen to persist for a day or longer, the use of the results for the following day based on this "persistence model" may be valid (USEPA, 2010). The persistence models regulate today's swimming condition with the previous day's *E. coli* measurement. However routine monitoring of FIB levels at most beaches is only conducted once per week. In addition, water quality at many nearshore waters next to beaches has been shown to change over a matter of hours (Boehm et al., 2002; Myers et al., 1998) so bacterial concentrations may

change between the time of sampling and the reporting of results even for the next day. Thus, the microbial safety of water for recreational use often cannot be determined in a timely manner using the "persistence model". Better solutions are needed that can provide near-real-time estimates ("NowCasts") to beachgoers within a few hours of sample collection (typically 4 hours or less; the shorter the better).

Newer rapid analytical methods (such as quantitative real-time Polymerase Chain Reaction (PCR), or qPCR) are being pursued as a possible approach to address this problem (Byappanahalli et al., 2018). However, challenges such as the high cost of establishing and operating analytical laboratory techniques, and the sophistication and technical challenge of new procedures relative to current culture-based methods must be overcome before such methods could become feasible alternatives (Dorevitch et al., 2017; Griffith and Weisberg, 2011). Modelling, in conjunction with laboratory experiments and field observations, can provide an alternative method to improve our understanding of the fate and transport of pathogenic contaminants in waterways and provide predictive decision-making support for effective public health management (Pachepsky et al., 2018).

An approach that has shown promising results is to use collected FIB monitoring data and complement them with common weather data and easily measurable water quality data to develop a predictive NowCast model to predict future FIB levels using mathematical techniques such as logistic regression (Aranda et al., 2015; Thoe et al., 2014), classification tree (Thoe et al., 2015a), physical descriptive models (He et al., 2016), machine learning (Park et al., 2018), and multiple linear regression (MLR) (Dada and Hamilton, 2016; Feng et al., 2015; Francy, 2009; Kato et al., 2019; Nevers and Whitman, 2005). More recent applications of MLR models have successfully improved model predictions for inland

recreational lakes (Bedri et al., 2016; Dada and Hamilton, 2016; Francy et al., 2013b; Nevers and Whitman, 2005; Nevers and Whitman, 2011; Olyphant and Whitman, 2004; Shively et al., 2016), coastal beaches (Boehm et al., 2007; Thoe et al., 2012) and reservoirs (Francy et al., 2013b) by including other water quality and hydro-metrological parameters that are easily or routinely measured, as well as by using better model optimization techniques made possible by advancements in computing.

MLR models are unique and must be developed specifically for each location to which they will be applied. Given the differences in data availability and the multitude of water quality and hydro-metrological parameters that may be relevant in each scenario, further learning and improvement in MLR models may still be possible through new applications. Quantitative weather-related variables such as wind speed, wind direction, wave height, Julian day and barometric pressure available from local weather monitoring stations have been routinely observed to be significant factors in MLR based predictive models for *E. coli* concentration at recreational beaches (Francy et al., 2013b; Thoe et al., 2014). Commonly recorded qualitative weather sky condition information (e.g., clear, cloudy, rainy) is widely reported but has not been included in such modelling efforts.

A minimum of two years of daily monitoring data with 100 or more data points is recommended as a training set for the development of MLR-based models (Francy et al., 2013b; USEPA, 2010). The benefit of developing models with varying training sets produced by data from longer monitoring periods has not been investigated. For testing and comparing model performance using a testing dataset, various statistics such as adjusted R^2, model accuracy (is the percentage of correct advisory predictions), sensitivity (percent correct exceedances or true positive), and specificity (percent correct non-

exceedances or true negative) have been used (see Appendix 2-1). While model sensitivity has the greatest implications for public health, the other statistics are also useful measures of model performance to consider when assessing the merits of posting beach advisories.

Sandpoint Beach is a popular public beach on Lake St. Clair in the City of Windsor, Ontario, Canada. The Windsor-Essex County Health Unit (WECHU) collected intensive (5 days per week) water quality and *E. coli* monitoring data at Sandpoint Beach over five years (2014 – 2018) during the summer months to aid in the development of statistical models.

The study described herein had four objectives. The first objective was to use the *E. coli* data collected over four years (2014 – 2017) and weather data provided by local weather stations over the same period to develop MLR-based NowCast models to predict microbial water quality (*E. coli* concentration) at Sandpoint Beach. Data and their transformations selected for use in MLR model development were selected based on best practices established in the literature (Francy et al., 2013a; Francy and Darner, 2006; USEPA, 2016). This includes identifying the potential explanatory variables and their transformations, creation of new variables, model evaluation and model selection. The second objective was to develop and compare the performance of MLR models that either include or do not include qualitative sky weather information commonly reported by weather stations. The third objective was to develop and compare the performance of MLR models using training datasets collected over 2 to 4 years from 2014 to 2017. The fourth objective was to examine the usefulness of the Principal Coordinate Analysis (PCoA) for a better and more holistic comparison of the performance of the MLR models developed using an independent testing

dataset (collected in 2018). The developed MLR models were also compared against two persistence models which represent the incumbent methodology.

2.3 Methods

2.3.1 Study Site:

Lake St. Clair is a 1,110 km^2 freshwater lake in the Lake Huron to Erie corridor in the Great Lakes Basin. It has several public beaches and access points in both the United States and Canada that are relevant for public health impact through recreational water use. Sandpoint Beach in Windsor, Ontario, Canada is located on the southern edge of Lake St. Clair (Figure 2-1). It is a popular beach for recreational use during the summer months for around 400,000 residents in the Windsor-Essex Region. Windsor is located at the same latitude as Northern California and it is the most humid city in the region. The average daily temperature reaches above 10 °C for about 200 days of the year. The warmest month of the year is July, with an average temperature of 22.3 °C during the study period (2014-2018). Precipitation is distributed throughout the year, though more tends to fall during the summer months with a mean and maximum monthly rainfall of 90 mm and 179 mm respectively. The safe *E. coli* limit for swimming and bathing at freshwater beaches in Ontario, set by the Ontario Ministry of Health, was updated to a geometric mean concentration of ≤ 200 colony-forming unit (CFU)/100 ml in 2018. Previously the limit was 100 CFU/100 ml (OMHLTC, 2018). Frequent exceedances of this threshold have been reported for many beaches in the Windsor-Essex Region, including Sandpoint Beach (McPhedran et al., 2013).

2.3.2 *E. coli* and Water Quality Data:

The WECHU collected water quality and *E. coli* monitoring data five days per week from Sandpoint Beach over five years during the summer months (2014 – 2018) to aid in the development of statistical models. Water samples were collected at five locations (1 point per ~35 meters) and analyzed for *E. coli* content within 4−6 hrs after collection in the laboratory using the Ontario Ministry of Health and Long-Term Care guidelines using membrane filtration method (OMHLTC, 2018). From 2014 to 2018, there were 65, 58, 63, 54, and 64 days of sampling each year respectively. Five samples were collected per day and the geometric mean *E. coli* concentration was calculated and reported. Values of 10 and 1000 were used in the calculation for results reported as <10 CFU/100 ml and > 1000 CFU/100 ml respectively. There was no reported value of less than 10 CFU/100 ml but 29 out of 240 values (12.1%) during the training period (2014-2017) and 7 out of 64 values (11%) during testing (2018) was reported as 1000 CFU/100 ml. More information about the censored data is included in Appendix 2-4. Water temperature was also measured at the site for every sample, and a randomly selected sample was analyzed in the laboratory for turbidity using Hach 2100Q portable turbidimeter.

To check the similarity of the data of each year, analysis of the variance (ANOVA) was used followed by Student-Newman-Keuls post-hoc test to check their differences in means, similar to Whitman and Nevers (2008). Also, non-parametric Mann-Kendall trend tests and Durbin–Watson tests were performed to find a trend and autocorrelation behaviour among the data respectively. Mann-Whitney test was used to compare daily *E. coli* values of each month with data from the other months. The critical p-value was set at 0.05.

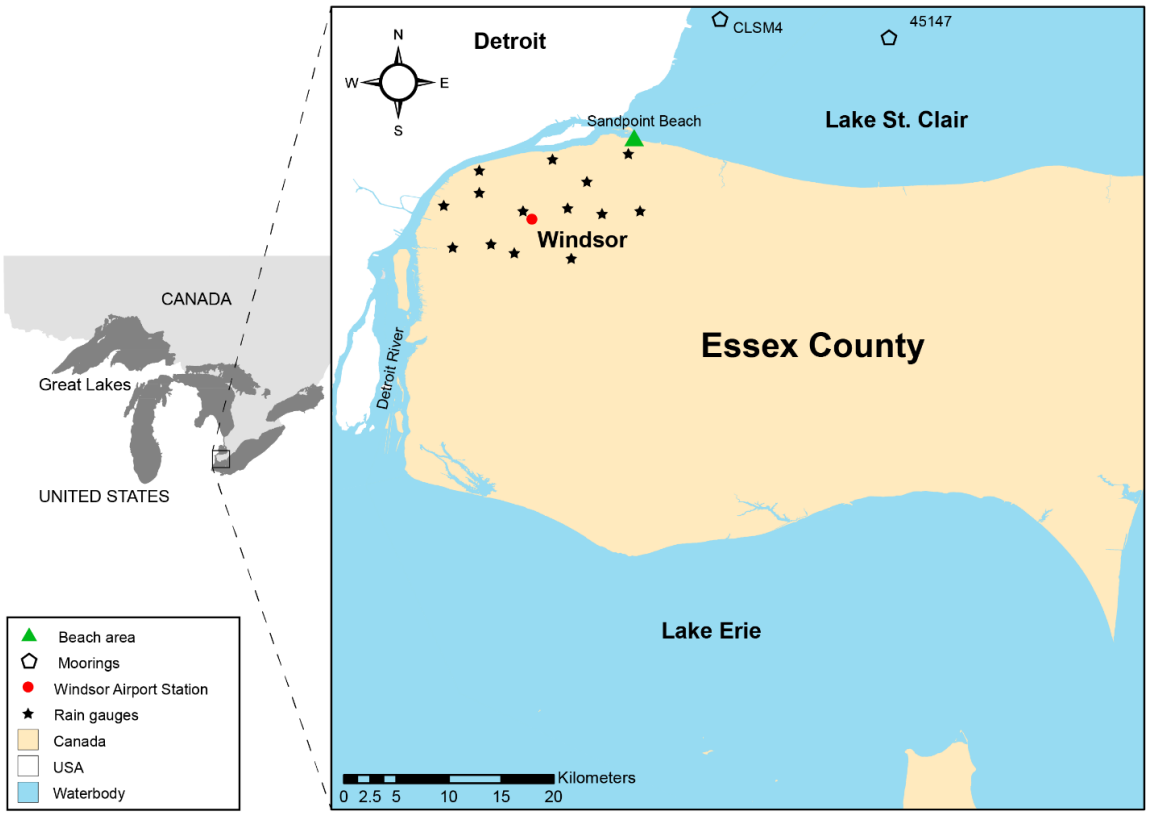

Figure 2-1: Map of Lake St. Clair and study site in Sandpoint Beach (green triangle). Location of Windsor Airport Station, 14 gauging stations, and moorings locations are identified by the red circle, black stars, and pentagon symbols respectively.

2.3.3 Metrological and Environmental Data:

Quantitative and qualitative hourly weather condition data were obtained from the nearest Environment Canada stations, including Windsor Airport (Station ID=4716 and 52838), buoy 45147 (42.430 N 82.680 W), and National Data Buoy Center (NDBC) weather stations (CLSM4 (42.471 N 82.877 W) (http://www.ndbc.noaa.gov/) in the middle and east part of Lake St. Clair (see Figure 2-1). Data included air temperature ($^{\circ}C$), daily rainfall (mm), and hourly measurements of wind speed (m/s), wind direction (degree), and wave height (m). 15-minute time interval rainfall data were obtained from Windsor's 14 gauging stations across the city (see Figure 2-1). Data collected at each station was converted to

hourly rainfall and averaged across all the stations. Hourly qualitative data based on their definition (see: http://climate.weather.gc.ca/glossary_e.html) were recorded at Windsor Airport station. For example, sky condition reflects the observation of total cloud amount, and is reported as the amount (in tenths) of cloud covering the dome of the sky, and is categorized as "Clear" (0 tenths), "Mainly clear" (1 to 4 tenths), "Mostly cloudy" (5 to 9 tenths), and Cloudy (10 tenths). In the case of precipitation, any form of water (liquid or solid) that falls from clouds is classified as drizzle, rain, moderate rain, heavy rain, snow, snow pellets, and thunderstorm. This information provides an hourly qualitative description of the weather and sky conditions in the region.

2.3.4 Persistence Models:

The WECHU conducts beach water quality monitoring at nine public beaches at least once each week (on Mondays) from June to September to assess the bacterial (*E. coli*) counts in the water and compare the results with provincial standards. Geometric mean < 200 CFU/100 ml indicates a low health risk and the beach is open, while anything greater than 200 and up to 999 poses a possible risk and swimming is not recommended. Closures are issued for beaches that have *E. coli* levels ≥ 1000 CFU/100 ml, and the beach water is resampled on Thursday of the same week. Beach water quality results from Monday's sampling are updated every Wednesday during the summer. If the beach was resampled, results are made available by the end of Friday to enable the opening of the beach over the weekend if the levels drop to within acceptable range. Based on the current monitoring approach described above, a simple model termed the "persistence model #1" was developed and used as the baseline for application of the predictive models to the current situation (Shively et al., 2016). In this model, *E. coli* concentration from Wednesday to

Monday is assumed to be the value reported on Wednesday (for the sample collected on Monday), unless the beach is resampled in between. To demonstrate the importance of predictive models and compare them to monitoring programs that only rely on measurements, another model based on the previous day's *E. coli* was developed. In this model (termed "persistence model #2"), results from samples collected five days a week (Monday to Friday) were used, as were available for the data used in the current study. As it takes two days for the results to be prepared and reported to the public, *E. coli* values as per this method are assumed to be the same as two days ago, or the most recent reported value. This model can be expressed mathematically as:

$$Y_i = \begin{cases} M_{i-2} & i = 3-7 \\ M_5 & i = 1,2 \end{cases} \quad (1)$$

where Y_i and M_i are the modelled and measured *E. coli* concentrations of day i respectively and where the value in subscript represents the day of the week, numbered from 1 (Monday) to 7 (Sunday).

2.3.5 MLR Model Descriptions

The MLR model for predicting *E. coli* concentrations can be expressed as:

$$\log_{10} Ecoli = \beta_o + \sum_{i=1}^{n} \beta_i V_i + \varepsilon \quad (2)$$

where $\log_{10} Ecoli$ is the log10-transformed *E. coli* concentration, n is the number of explanatory variables, V_i and β are i th explanatory variable and corresponding regression coefficient respectively, and ε is the residual error. *E. coli* data used in model development were log10-transformed before model development to meet parametric assumptions of

equality of variances and normal distribution over the wide range of expected values (Francy et al., 2013b). Based on available water quality and meteorological data and the review of related literature (Bedri et al., 2016; Brooks et al., 2016; Dada and Hamilton, 2016; Feng et al., 2015; Nevers and Whitman, 2011; Thoe et al., 2014; USEPA, 2010), the following is the list of explanatory variables chosen for MLR model development. For water quality and beach condition parameters: number of birds (NBirds), day of the year (DOY), turbidity (Turb), water temperature at the time of sampling (Wtemp), and for meteorological data: Rain combined based on 8h, 12h, 24h, 48h and 72h antecedent cumulative rainfall (RainCombine), 10 hours' antecedent wind direction (WDRanked), 10 hours' antecedent wind speed and its components parallel and perpendicular to the shoreline (WSPAVec and WSPOVec respectively), daily air temperature (Atemp), and wave height (WaveH). Detailed descriptions of data and variables are presented in Appendix 2-2. Multi-collinearity between some variables, such as wind speed components, is expected and was avoided during the model selection process.

Water quality parameters and meteorological data were not transformed. These parameters were considered for MLR M series (M1 to M5; Table 2-1) models. In addition to the meteorological data stated above, additional qualitative weather information in the form of categorical sky conditions and atmospheric phenomena are commonly reported in Canada as per the recommended standards set by the World Meteorological Organization (WMO) (ECCC, 2019). In addition to the parameters considered for the M models, the qualitative weather parameter (WeatherRank) was included for the development of MLR W series (W1 to W5; Table 2-1) models. To create this explanatory variable out of qualitative data and use them to construct MLR W series models, a genetic algorithm along with one-hot

encoding approach was applied, as follows. First qualitative weather data are classified into ten unique categories. Quantitative data that can be used for the building of MLR models were generated using one-hot-encoding. It returns a vector for each unique entries of the categorical data. Each vector contains only one '1' while all other values in the vector are '0'. A computer code was written in MATLAB (Mathworks, Natick, MA) to use a genetic algorithm along with a one-hot encoding approach to develop and select the best MLR model under a given set of conditions. Population selection was based on stochastic uniform function. Crossover and mutation rates were kept as default values of 0.8 and 0.01 respectively. Optimization was terminated when the average change in the fitness value was less than 10^{-6}. More details about the preparation of qualitative weather data and model building are presented in Appendix 2-3. Significance was set at $P = 0.05$ unless otherwise stated.

The best model was selected based on criteria that maximized the R^2 value, minimized the root mean square error (RMSE), and minimized Akaike's Information Criterion (AIC). Five M and W MLR models were established using different training periods: 2014-2015 (M1 and W1), 2014-2016 (M2 and W2), 2014-2017 (M3 and W3), 2015-2017 (M4 and W4), and 2016-2017 (M5 and W5). Monitoring data from the year 2018 was used for model testing and comparison. In addition, to isolate and directly evaluate the value added by the qualitative weather data to the MLR models, we built MW series models with the same architecture as the M series models except that the qualitative weather data was added as the 8^{th} variable.

2.3.6 Model Selection and Metrics:

Model building was limited to seven variables (see Appendix 2-5). To create the best model, an exhaustive search (all possible combinations through all variables) was performed (Feng et al., 2015). Thus with 17 variables, we ended up with 19,448 models for M series. Inclusion of qualitative weather data, as the 18th variable, added 12,376 W series models. Model performance was examined by determination of metrics such as R^2, RMSE, AIC, accuracy, sensitivity (ability to predict *E. coli* exceedances of 200 CFU/100 ml or true positive), and specificity (ability to predict *E. coli* lower than the threshold of 200 CFU/100 ml or true negative). Each model was checked for multi-collinearity based on the Variance Inflation Factor (VIF) as suggested by Cyterski et al. (2013). Any model containing an explanatory variable with a high degree of correlation with others (as measured by a large VIF value > 5) was removed from consideration during model selection (See Appendix 2-6). For model performance, both sensitivity and specificity are important. By plotting the true positive rate (sensitivity) against its false positive rate (1 - specificity) for a given model, known as receiver operating characteristic (ROC) curve, a single value for its area under the ROC curve (AUROC) can be obtained (Morrison et al., 2003). The advantage of AUROC (as opposed to simply reporting specificity and sensitivity) is that the AUROC is independent of the decision threshold. The AUROC values vary between 0 – 1. A model with AUROC of less than 0.5 indicates reciprocal classification (i.e., suggests below when actually above and vice-versa), which should be corrected. An AUROC equal to 0.5 (i.e. coinciding with the diagonal) shows a random classification model. Models with AUROC of 0.5 – 0.6 are considered a fail, 0.6 – 0.7

considered poor, 0.7 – 0.8 considered good, 0.8 – 0.9 considered very good, and 0.9 – 1 considered excellent (Holtschlag et al., 2008; Tape, 2007).

In this study, AUROC values were used for comparing the performance of different models. AUROC is independent of the decision threshold and therefore has an advantage over the traditional approach of reporting specificity and sensitivity. However, AUROC did not take into account all aspects of model performance evaluation such as R^2, R^2-adjusted, and RMSE. Thus, in addition to AUROC analysis, a more comprehensive comparison of the models using PCoA was conducted. Chi-squared distance is used to generate the dissimilarity matrix. The analyses and plots were performed using a self-written MATLAB script and we used XLSTAT trial version (https://www.xlstat.com/en/) to check for the accuracy of the code in performing PCoA analysis. Detailed description of this analysis is presented in Appendix 2-9.

2.4 Results and Discussion

2.4.1 *E. coli* Trend Analysis

Figure 2-2 shows a descriptive plot of the *E. coli* concentration during the training (2014-2017) and testing period (2018). Results from ANOVA and Student Newman Keuls post hoc test show that *E. coli* data are significantly different in means, with a trend decreasing from 2014, when there was a high mean, to 2017, when the mean and median *E. coli* data were both low (see Figure 2-2c).

In addition, *E. coli* data for year 2017 had the lowest spread, while 2014 data had the highest. During the training period 2014-2017, out of 240 daily values, there were 131 *E. coli* exceedances greater than 100 CFU/100 ml (~ 54.6%), 87 exceedances greater than 200 CFU/100 ml (~ 36.2%), and 29 *E. coli* exceedances greater than 1000 CFU/100 ml

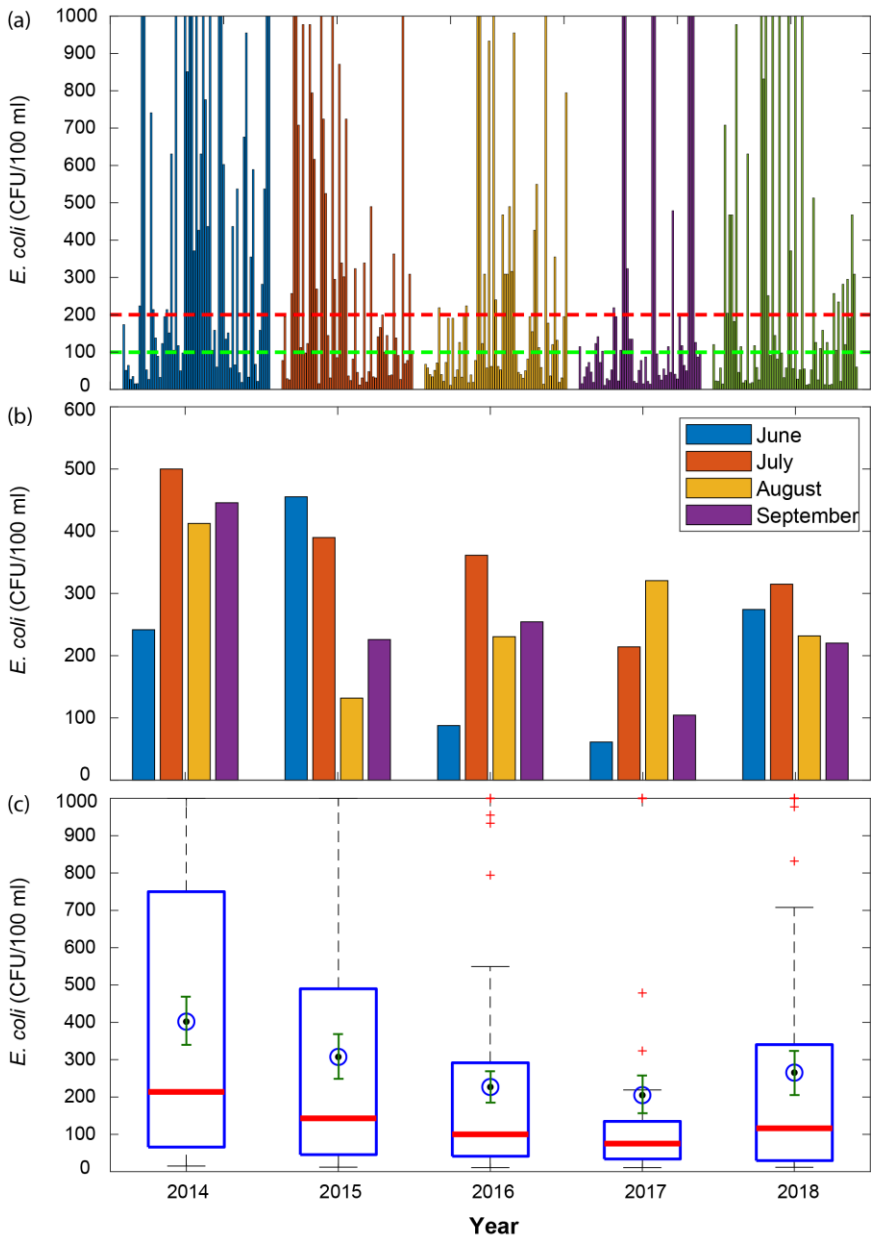

Figure 2-2: *E. coli* data from the 2014-2017 training period and the 2018 testing period: a) 5-day per week samples b) monthly average c) box plot that shows median (red line) and geometric mean (black dot inside a blue circle, and error bar), and spread of the data (the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Whiskers ('+' symbol) show the most extreme data points not considered outliers. Green and red dotted lines in a) show the safe limit for *E. coli* concentration according to the Ontario Ministry of Health

(~12.1%) (see Figure 2-3a). Non-parametric Mann-Kendall trend tests reveal that there is

no trend in the time series (Figure 2-2a) and monthly mean data (Figure 2-2b) (P value of 0.28 and 0.32 respectively). Additionally, Durbin–Watson statistic reveals that selected variables were not auto-correlated. These checks are important as they help ensure the

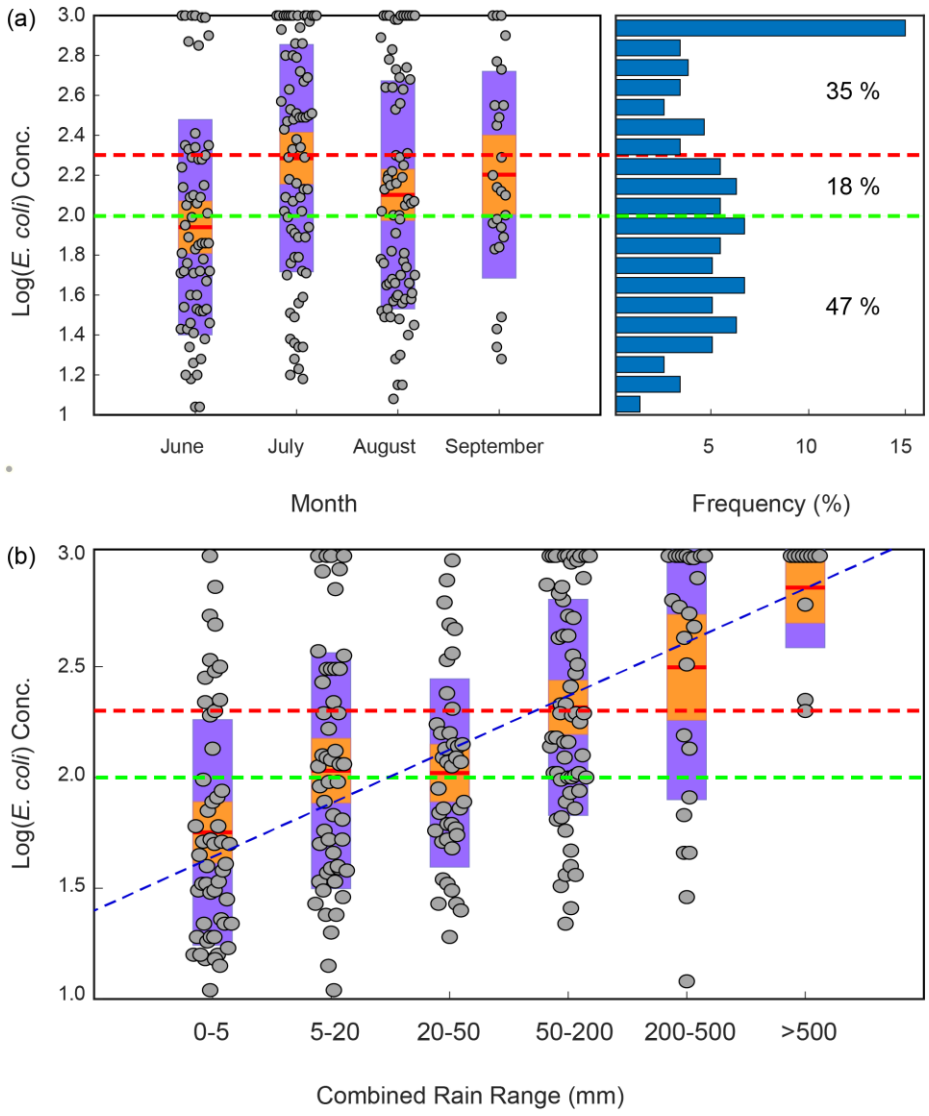

Figure 2-3: *E. coli* variation with a) month of the year and b) combined rainfall range during the training period 2014-2017. Points are laid over a 1.96 standard error of the mean (SEM) (95% confidence interval) in orange and a 1 standard deviation (SD) in the purple area. The horizontal bar chart in a) shows the distribution of the *E. coli* data and frequency of occurrence (%). Green and red dotted lines show the safe limit for *E. coli* concentration according to the Ontario Ministry of Health.

serial-correlation condition is met and the data is suitable for MLR analysis (Boehm et al., 2007; Ge and Frick, 2007).

The monthly variation of *E. coli* for the period of 2014-2017 is shown in Figure 2-3a. Although the spread of data for all months is very similar, results from the Mann-Whitney test reveal that July month *E. coli* data is significantly higher than June, August and September ($p = 0.001$, $p = 0.02$ and $p = 0.017$ respectively). A trend of increasing combined rainfall amount from past 8, 12, 24, 48, and 72 hours can be observed in Figure 2-3b. A combined value of 50 mm or higher is strongly correlated with the exceedance of *E. coli* data from 200 CFU/100 ml.

Table 2-1: Performance, statistics, and accuracy of the statistical MLR models without qualitative weather information (M series), MLR models with qualitative weather information (W series), and persistence models (PM#1 and PM#2) developed. The information shown in every model is from the testing year 2018.

Name	Training Period	TP*	TN†	FP‡	FN⁂	R^2	R^2Adj	RMSE	AIC	Accu.	Specif.	Sensit.	AUROC
M1	2014-2015	11	33	8	12	0.225	0.213	0.292	187	69	80	48	0.73
M2	2014-2016	9	32	9	14	0.102	0.088	0.346	273	64	78	39	0.73
M3	2014-2017	7	37	4	16	0.065	0.050	0.251	363	69	90	30	0.70
M4	2015-2017	9	37	4	14	0.071	0.056	0.341	240	72	90	39	0.72
M5	2016-2017	10	37	4	13	0.129	0.115	0.342	151	73	90	43	0.75
W1	2014-2015	18	30	11	5	0.173	0.159	0.380	176	75	73	78	0.78
W2	2014-2016	17	33	8	6	0.137	0.124	0.384	257	78	80	74	0.78
W3	2014-2017	16	32	9	7	0.182	0.168	0.391	334	75	78	70	0.76
W4	2015-2017	15	32	9	8	0.098	0.084	0.449	232	73	78	65	0.77
W5	2016-2017	16	34	7	7	0.128	0.114	0.416	142	78	83	70	0.79
PM#1	-	9	26	10	19	0.001	-0.017	0.620	-	54	72	32	0.57
PM#2	-	6	27	12	17	0.021	0.004	0.614	-	53	69	26	0.53

* TP: True positive; when the modelled and observed *E. coli* levels are both above the bathing water standard (BWS) of 200 CFU/100 ml
† TN: True negative; when the modelled and observed *E. coli* levels are both below the bathing water standard (BWS) of 200 CFU/100 ml
‡ FP: False positive; when the modelled *E. coli* levels are above BWS but the observed *E. coli* levels are below BWS. (Type I error)
⁂ FN: False negative; when the modelled *E. coli* levels are below BWS but the observed *E. coli* levels are above BWS. (Type II error)

2.4.2 MLR Model Testing and Comparison

The two simplistic persistence models (PM) – #1 and #2 – only use the *E. coli* monitoring data. Results presented in Table 2-1 show that the AUROC values for PM 1 and 2 are between 0.5 – 0.6, as compared to >0.7 for all the MLR models developed. RMSE obtained by all the MLR models range from 0.29 – 0.44 log CFU/100 ml, which is lower than or within the range of 0.4 – 0.5 log CFU/100 ml for MLR models reported in a review by Thoe et al. (2014). Thus, all the M and W series MLR models developed have lower RMSE compared to persistence models 1# and #2 with RMSE 0.62 and 0.61 log CFU/100 ml respectively. For the M series MLR models (without the qualitative weather data), the best AUROC value of 0.75 was obtained with the M5 model that used the most recent two years (2016-2017) preceding the testing year 2018 as the training period. AUROC values ranging between 0.70 – 0.73 for the remaining M models (M1 – M4) were also good and had relative standard deviations (RSD) of 2.7%, a value not far off from that of the M5 model.

Including the qualitative weather data, higher AUROC values ranging between 0.76 – 0.79 were observed for W1-W5 models. The highest AUROC value of 0.79 was again obtained with the model (W5) that used the most recent two years (2016-2017) preceding the testing year 2018 as the training period. False positives or false negatives are site-specific and often also related to the number of exceedances observed. For instance, Francy et al. (2003) reported false positive and negative values for six Ohio beaches in the range of 0-27.4% and 2.6-26.3% respectively. During the training period, false positives and negatives in the current study were within the range for both the M series models (6.3-14.1% and 18.8-25% respectively), as well as the W series models (10.9-18.8% and 7.8-12.5% respectively). More detailed discussion on false positive and false negative results is presented in

Appendix 2-11. According to Table 2-1, the number of Type II errors (or false negatives) in all the W series models (between 5 to 8 cases) is lower than M series models (between 12 to 16 cases). Figure 2-4 shows the Type I and II errors for the W5 model with both the training (2016-17) and testing (2018) datasets. During the training period, there were more Type II and Type I errors (20 vs. 6), whereas during the testing period they were even. That is probably because of the unbalanced number of observations that fall in the two categories of lower and higher than 200 CFU/100 ml. Out of 117 values in the training period of model W5, 87 values (~ 75%) are lower than 200 CFU/100 ml. During the testing period, the data is more balanced (~ 64% are lower than 200 CFU/100 ml). The relatively few Type I and II errors for the model explain the high AUROC value observed. As with the M series models, with relative standard deviation (RSD) of 1.2%, the AUROC values for

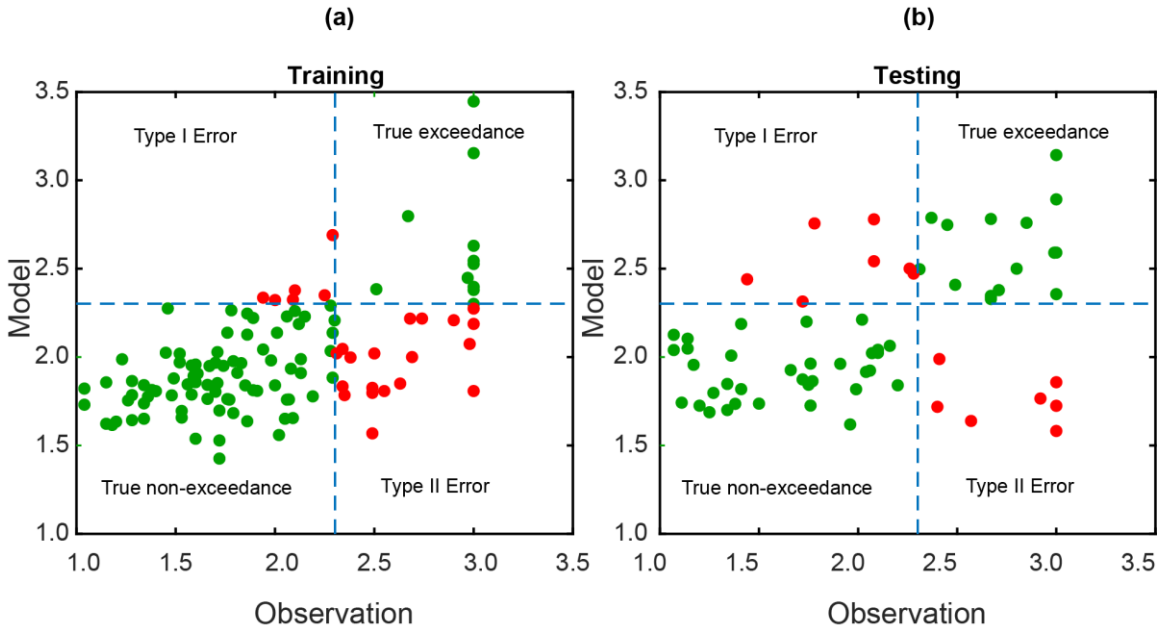

Figure 2-4: Observed vs predicted log10- *E. coli* concentrations of model W5. a) shows the training and b) shows testing observation vs model and status of false positives (Type I error), false negatives (Type II error), and the exceedance and non-exceedance based on the threshold of 200 CFU/100 ml *E. coli* concentration (blue dash line).

the other W series models (W1 – W4) were not much different from the W5 model. However, the higher AUROC values observed for the W series models as compared to the M series models were statistically significant (Mann-Whitney test, P = 0.012).

MW series models were obtained by adding the qualitative weather parameter as the eight variable to the M series models, and their performance was compared with the M series models for a direct assessment of the value of adding the qualitative weather parameter. Details of the comparison are presented in Appendix 2-7. Results show that in 100% of the models for R^2, 98-100% models for RMSE, and 91-100% for AIC, the addition of quantitative weather data improved model performance as compared to the corresponding M series model with similar model architecture.

Similarly, models with similar model structure but different training period were compared for both the M and W series models. Details of the comparison are presented in Appendix 2-8. For the M-series, models built with 2 years of data in training period (2016-2017), showed higher R^2, lower RMSE, and lower AIC than the corresponding models built with 3-(2015-2017), and 4-(2014-2017) years, for 87.7%, 96.1%, and 100% of the models respectively. Similar results were observed with the W series models where models built with 2 years' data (2016-2017), showed better performance measured by higher R^2, lower RMSE, and lower AIC than the corresponding models built with 3- (2015-2017), and 4-(2014-2017) years, for 99.9%, 100%, and 100% models respectively.

While AUROC values are important indicators of model performance, comparing model performance including the other metrics presented in Table 2-1 should yield even better results. Such a comprehensive comparison can be made possible using PCoA, and the results of the MLR and persistence models developed in this study are compared and

presented in Figure 2-5. Point X on the graph represents the ideal model with 100% accuracy (R^2, sensitivity, specificity and AUROC equal to 1 and RMSE = 0). The shorter the distance between "X" and the developed model, the better the model is.

It is clear from Figure 2-5 that the conclusions drawn from PCoA are very similar to those based on the AUROC values in Table 2-1. Both the persistence models are located farther than the M and W series models from the "X", and thus have the lowest predictive

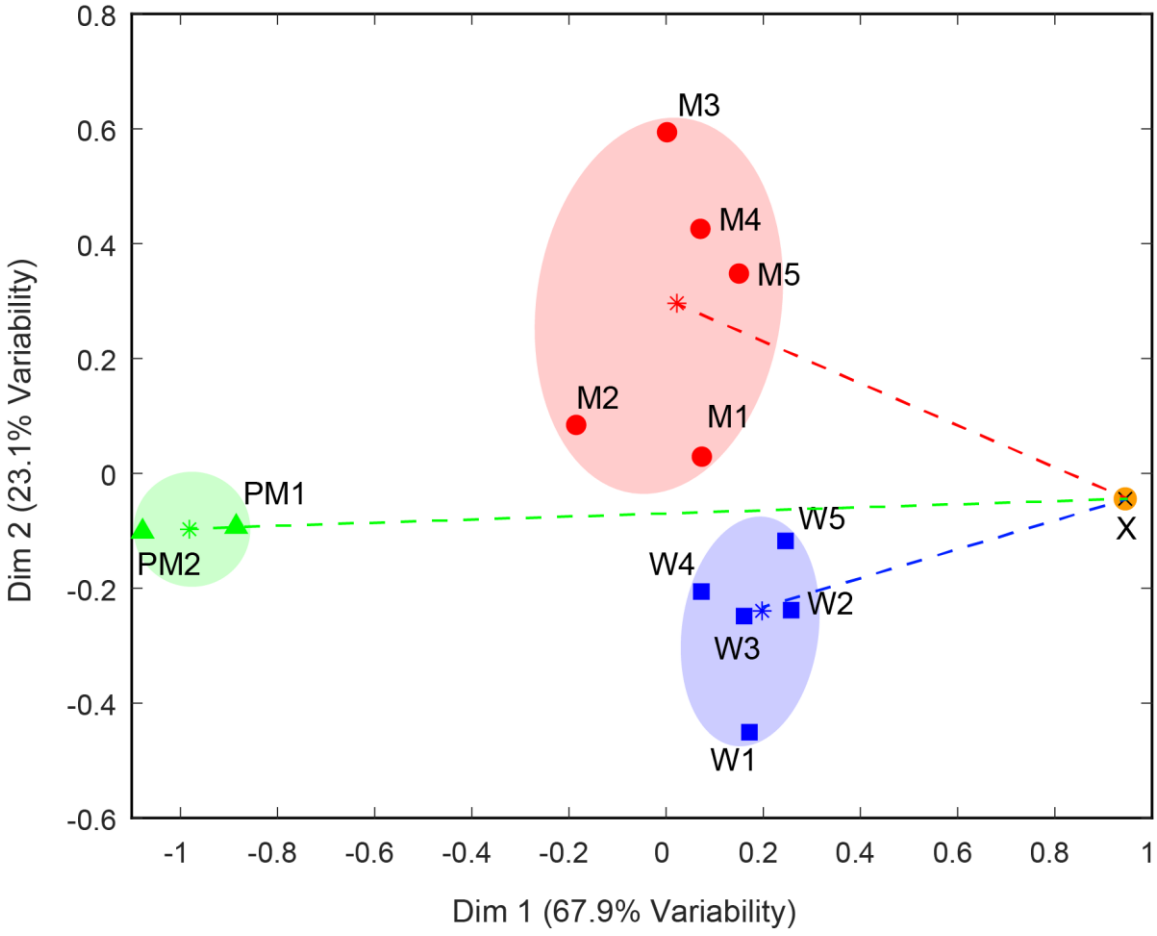

Figure 2-5: Principal Coordinate Analysis (PCoA) of the different models' performance. MLR M series are depicted as red circles, MLR W series are plotted as blue squares and persistence models are depicted as green triangles. The ellipses show the cluster of the models and the star inside each ellipse shows the mean of all the models in that cluster.

capability. Importantly, all the W series models have better predictive capability than any of the M series models except for M1 and M5 that show more similarity to the ideal model compared to W1 and W4. This shows that the inclusion of the qualitative weather data significantly improved the predictive capability of the model. Among the W series models, although the AUROC values are similar, the W5 model does emerge as the superior model as compared to the others (W1-W4 models).

For the W series models, the W5 Model had the highest AUROC value at 0.79 and thus was deemed to be the best model. However, two other models (W1 and W2) had very similar AUROC values at 0.78. PCoA analysis using AUROC values as well as other metrics such as R^2, sensitivity, specificity, and RMSE still showed W5 to be the best model based on the shortest distance from the ideal model (Figure 2-5). This shows that PCoA analysis using AUROC and other metrics used in the current study may be a more sensitive and thus better tool for model comparison than AUROC values alone. For instance, comparing the W5 Model with the best M series model (M5), the AUROC value improved from 0.75 to 0.79, but PCoA clearly separated the two models. As another example, according to AUROC, M1 is much lower than W1 however, PCoA analysis shows that including all the parameters into account M1 performed better than W1. While model specificity was slightly lower (83% as compared to 90%), model sensitivity for the W5 Model (70%) was much higher than that of the M5 Model (43%). Furthermore, in the PCoA graph, model W5 and W1 are located far apart which means they are not as similar as they are presented based solely on the AUROC results in Table 2-1.

The most significant explanatory variables in the five W series models were compared, and the results are presented in Appendix 2-10. Six variables (qualitative weather data

(WeatherRank), wind direction (WDRanked), wave height (WaveH), wind speed moving perpendicular to the shoreline (WSPOVec), air temperature (Atemp), and turbidity (Turb)) were common among three or more models. The qualitative weather data variable and wind direction were significant for all five models. The inclusion of the qualitative weather data is what differentiates the W and M series models, and its significance in all five models explains the better predictive capability of the W series over the M series models. The *E. coli* data >200 CFU/ 100 ml were segregated from data <200 CFU/100 ml for 2018, and the wind pattern for those days is presented in Figure 2-S3 (see Appendix 2-12). Figure 2-S3 shows that the most frequent *E. coli* exceedances were observed when the prevailing winds were from between NNW and NNE. Given the location and orientation of Sandpoint Beach (Figure 2-1), this indicates that the *E. coli* might originate from sediment resuspension, beach sand, or microbial pollution from watershed sources E to NE from the beach. However, Figure 2-S3 and the differences among the explanatory variables between the various models highlights the limitations of statistical models in identifying potential sources of microbial pollution responsible for *E. coli* exceedances observed in beach waters.

Of all the MLR models developed, W5 has the shortest distance to the location of the ideal model "X" (see Figure 2-5) and thus has the best predictive capability. W5 is, therefore, the MLR model recommended for use at Sandpoint Beach among all the MLR models developed in this study. W5 is based on two-year data (2016-2017) immediately preceding the testing year (2018). The results suggest that for Sandpoint Beach, it is best to develop a new model every year based on preceding two-year data. Similar findings and suggestions have also been made previously in the literature (Francy et al., 2013b; Thoe et

al., 2014, 2015b). However, the intensive (5 days per week) water quality and *E. coli* monitoring conducted by WECHU during the summer over five years (2014 – 2018) was done to assist in the development of statistical models and has now been discontinued. The results for Sandpoint Beach also show that models W1 to W4 developed based on data collected during the preceding 5 years (2014-2017), were still reasonably good. Thus, the recommended W5 MLR model may be expected to give reasonable predictions for more than two years provided there are no significant changes in microbial loadings from watershed-affected beach water quality at Sandpoint Beach. It is recommended that at least two years of the intensive (5 days per week) water quality and *E. coli* monitoring over summer months be conducted every few years to confirm the validity of, or development of a new MLR model.

2.5 Conclusion

Current beach postings at Sandpoint Beach on Lake St. Clair rely on the persistence model, which only uses *E. coli* monitoring data collected as part of the beach monitoring program. Our results show that both persistence model #1 (one sample per week) and persistence model #2 (five samples per week) performed poorly, with AUROC values between 0.5 – 0.6, and sensitivity of 32% or less. Using the same data with some additional easily and rapidly measurable water quality data (as part of the beach monitoring program) and weather data from local weather monitoring stations that can be available in near real-time, all the MLR models developed in the current study performed much better than the persistence models, with AUROC values between 0.7 – 0.8.

The inclusion of qualitative weather data improved the performance of the MLR models. Qualitative weather condition is commonly reported by weather stations and is

recommended for consideration in similar future MLR modelling exercises. MLR models developed using the most recent two-year monitoring data performed better than those developed with three- and four-year data. This suggests that two-year monitoring data should be sufficient as a training dataset for Sandpoint Beach in the event that there is a need for updating the model in the future. Similar analysis at other beaches may be useful for determining the adequacy of the training period. PCoA, as presented, seems to be an effective tool for model selection based on multiple performance criteria. Based on this, among the MLR models developed in the current study, the W5 model is recommended for use by WECHU for NowCasting microbial water quality at Sandpoint Beach over the persistence model #1 that is currently used.

Results presented in this study demonstrate that the application of MLR models supports the monitoring tools' capacity to predict, with reasonable accuracy, *E. coli* concentrations at Sandpoint Beach. However, due to the complexity of *E. coli* fate in the environment as well as model accuracy limits and the inherent high spatial and temporal variability of measured *E. coli* concentrations, inconsistencies between measured and predicted *E. coli* concentrations may still occur. We suggest that, because developing and using predictive models is a dynamic process based on continued data collection via existing beach-monitoring programs, model performance and accuracy should be evaluated at the end of each beach season to increase or maintain their predictive power. Any significant decreases in performance would indicate that environmental or microbial loadings that affect microbial water quality have changed, and the development of an updated model may be necessary.

2.6 Acknowledgements

The authors are grateful to the Windsor-Essex County Health Unit (WECHU) for providing the daily *E. coli*, water temperature and turbidity data at Sandpoint Beach collected over five years' summer months (2014 – 2018). We would also like to thank Karen Lukic, Philip Wong, Ramsey D'Souza and Victoria Peczulis from the WECHU for their review and comments on the manuscript. Special thanks to Karina Richters from the City of Windsor for providing us with the high-frequency rainfall data. This research received financial support from Canada's Natural Sciences and Engineering Research Council (NSERC) Strategic Project Grant (SPG) and was performed while Mr. Madani (first author) held an Ontario Trillium Scholarship (OTS).

2.7 Supplemental Material

A detailed description of the metrics, explanatory variables, model selection and PCoA analysis are presented in Appendix 2.

2.8 References

Aranda, D., Lopez, J.V., Solo-Gabriele, H.M., Fleisher, J.M., 2015. Using probabilities of enterococci exceedance and logistic regression to evaluate long term weekly beach monitoring data. Journal of Water and Health 14(1) 81-89.

Bedri, Z., Corkery, A., O'Sullivan, J.J., Deering, L.A., Demeter, K., Meijer, W.G., O'Hare, G., Masterson, B., 2016. Evaluating a microbial water quality prediction model for beach management under the revised EU Bathing Water Directive. Journal of Environmental Management 167 49-58.

Boehm, A.B., Grant, S.B., Kim, J.H., Mowbray, S.L., McGee, C.D., Clark, C.D., Foley, D.M., Wellman, D.E., 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. Environmental Science & Technology 36(18) 3885-3892.

Boehm, A.B., Whitman, R.L., Nevers, M.B., Hou, D., Weisberg, S.B., 2007. Nowcasting Recreational Water Quality, In: (Ed.), L.J.W. (Ed.), Statistical Framework for Recreational Water Quality Criteria and Monitoring, pp. 179-210.

Brooks, W., Corsi, S., Fienen, M., Carvin, R., 2016. Predicting recreational water quality advisories: A comparison of statistical methods. Environmental Modelling & Software 76 81-94.

Byappanahalli, M.N., Nevers, M.B., Shively, D.A., Spoljaric, A., Otto, C., 2018. Real-Time Water Quality Monitoring at a Great Lakes National Park. Journal of Environmental Quality 47(5) 1086-1093.

Cyterski, M., Brooks, W., Galvin, M., Wolfe, K., Carvin, R., Roddick, T., Fienen, M., Corsi, S., 2013. Virtual Beach 3: User's Guide. United States Environmental Protection Agency.

Dada, A.C., Hamilton, D.P., 2016. Predictive Models for Determination of E. coli Concentrations at Inland Recreational Beaches. Water, Air, & Soil Pollution 227(9) 347.

Dorevitch, S., Shrestha, A., DeFlorio-Barker, S., Breitenbach, C., Heimler, I., 2017. Monitoring urban beaches with qPCR vs. culture measures of fecal indicator bacteria: Implications for public notification. Environmental health : a global access science source 16(1) 45-45.

ECCC, 2019. Manobs—Manual of Surface Weather Observations, Eight Edition, Chapter 7 Sky condition standards. Monitoring and Data Services Directorate, Meteorological Service of Canada, Gatineau, QC, Canada, Catalog Number: En56-238/2-2018E-PDF.

Feng, Z.X., Reniers, A., Haus, B.K., Solo-Gabriele, H.M., Wang, J.D., Fleming, L.E., 2015. A predictive model for microbial counts on beaches where intertidal sand is the primary source. Marine Pollution Bulletin 94(1-2) 37-47.

Francy, D.S., 2009. Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches. Aquatic Ecosystem Health & Management 12(2) 177-182.

Francy, D.S., Brady, A.M.G., Carvin, R.B., Corsi, S.R., Fuller, L.M., Harrison, J.H., Hayhurst, B.A., Lant, J., Nevers, M.B., Terrio, P.J., Zimmerman, T.M., 2013a. Developing and implementing predictive models for estimating recreational water quality at Great Lakes beaches: U.S. Geological Survey Scientific Investigations Report 2013–5166, 68 p., https://doi.org/10.3133/sir20135166.

Francy, D.S., Darner, R.A., 2006. Procedures for developing models to predict exceedances of recreational waterquality standards at coastal beaches: U.S. Geological Survey Techniques and Methods 6–B5, 34 p.

Francy, D.S., Gifford, A.M., Darner, R.A., 2003. Escherichia coli at Ohio bathing beaches: Distribution, sources, wastewater indicators, and predictive modeling. US Department of the Interior, US Geological Survey.

Francy, D.S., Stelzer, E.A., Duris, J.W., Brady, A.M.G., Harrison, J.H., Johnson, H.E., Ware, M.W., 2013b. Predictive models for Escherichia coli concentrations at inland lake beaches and relationship of model variables to pathogen detection. Applied and Environmental Microbiology 79(5) 1676.

Ge, Z., Frick, W.E., 2007. Some statistical issues related to multiple linear regression modeling of beach bacteria concentrations. Environmental Research 103(3) 358-364.

Griffith, J.F., Weisberg, S.B., 2011. Challenges in Implementing New Technology for Beach Water Quality Monitoring: Lessons From a California Demonstration Project. Marine Technology Society Journal 45(2) 65-73.

He, C., Post, Y., Dony, J., Edge, T., Patel, M., Rochfort, Q., 2016. A physical descriptive model for predicting bacteria level variation at a dynamic beach. Journal of Water and Health 14(4) 617-629.

Health-Canada, 2012. Guidelines for Canadian Recreational Water Quality, Third Ed., Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, (Catalogue No H129-15/2012E).

Holtschlag, D.J., Shively, D., Whitman, R.L., Haack, S.K., Fogarty, L.R., 2008. Environmental factors and flow paths related to Escherichia coli concentrations at two beaches on Lake St. Clair, Michigan, 2002–2005, Scientific Investigations Report: Reston, VA.

Kato, T., Kobayashi, A., Oishi, W., Kadoya, S.-s., Okabe, S., Ohta, N., Amarasiri, M., Sano, D., 2019. Sign-constrained linear regression for prediction of microbe concentration based on water quality datasets. Journal of Water and Health 17(3) 404-415.

McPhedran, K., Seth, R., Bejankiwar, R., 2013. Occurrence and predictive correlations of Escherichia coli and Enterococci at Sandpoint beach (Lake St Clair), Windsor, Ontario and Holiday beach (Lake Erie), Amherstburg, Ontario. Water Quality Research Journal of Canada 48(1) 99-110.

Morrison, A.M., Coughlin, K., Shine, J.P., Coull, B.A., Rex, A.C., 2003. Receiver Operating Characteristic Curve Analysis of Beach Water Quality Indicator Variables. Applied and Environmental Microbiology 69(11) 6405.

Myers, D.N., Koltun, G.F., Francy, D.S., 1998. Effects of hydrologic, biological, and environmental processes on sources and concentrations of fecal bacteria in the Cuyahoga River, with implications for management of recreational waters in Summit and Cuyahoga Counties, Ohio, Water-Resources Investigations Report: Columbus, OH, p. 56.

Nevers, M.B., Whitman, R.L., 2005. Nowcast modeling of Escherichia coli concentrations at multiple urban beaches of southern Lake Michigan. Water Research 39(20) 5250-5260.

Nevers, M.B., Whitman, R.L., 2011. Efficacy of monitoring and empirical predictive modeling at improving public health protection at Chicago beaches. Water Research 45(4) 1659-1668.

Olyphant, G.A., Whitman, R.L., 2004. Elements of a predictive model for determining beach closures on a real time basis: The case of 63rd Street Beach Chicago. Environmental Monitoring and Assessment 98(1-3) 175-190.

OMHLTC, O.M.o.H.a.L.-T.C., 2018. Operational approaches for recreational water guideline. Queen's Printer for Ontario: Toronto, ON.

Pachepsky, Y.A., Allende, A., Boithias, L., Cho, K., Jamieson, R., Hofstra, N., Molina, M., 2018. Microbial Water Quality: Monitoring and Modeling. Journal of Environmental Quality 47(5) 931-938.

Park, Y., Kim, M., Pachepsky, Y., Choi, S.-H., Cho, J.-G., Jeon, J., Cho, K.H., 2018. Development of a Nowcasting System Using Machine Learning Approaches to Predict Fecal Contamination Levels at Recreational Beaches in Korea. Journal of Environmental Quality 47(5) 1094-1102.

Rabinovici, S.J.M., Bernknopf, R.L., Wein, A.M., Coursey, D.L., Whitman, R.L., 2004. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. Environmental Science & Technology 38(10) 2737-2745.

Shively, D.A., Nevers, M.B., Breitenbach, C., Phanikumar, M.S., Przybyla-Kelly, K., Spoljaric, A.M., Whitman, R.L., 2016. Prototypic automated continuous recreational

water quality monitoring of nine Chicago beaches. Journal of Environmental Management 166 285-293.

Shuval, H., 2003. Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. Journal of Water and Health 1(2) 53-64.

Tape, T.G., 2007. The area under an ROC curve: University of Nebraska Medical Center: accessed April 2020 at http://gim.unmc.edu/dxtests/roc3.htm.

Thoe, W., Choi, K.W., Lee, J.H.-w., 2015a. Predicting 'very poor' beach water quality gradings using classification tree. Journal of Water and Health 14(1) 97-108.

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M.L., Boehm, A.B., 2014. Predicting water quality at Santa Monica Beach: Evaluation of five different models for public notification of unsafe swimming conditions. Water Research 67 105-117.

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M.L., Boehm, A.B., 2015b. Sunny with a Chance of Gastroenteritis: Predicting Swimmer Risk at California Beaches. Environmental Science & Technology 49(1) 423-431.

Thoe, W., Wong, S.H.C., Choi, K.W., Lee, J.H.W., 2012. Daily prediction of marine beach water quality in Hong Kong. Journal of Hydro-Environment Research 6(3) 164-180.

USEPA, 2010. Predictive Tools for Beach Notification, Vol. I: Review and Technical Protocol, EPA-823-R-10-003. U.S. Environmental Protection Agency Office of Water Office of Science and Technology 61.

USEPA, 2016. Six Key Steps for Developing and Using Predictive Tools at Your Beach, EPA-820-R-16-001. U.S. Environmental Protection Agency Office of Water Office of Science and Technology available at: https://www.epa.gov/sites/production/files/2016-03/documents/six-key-steps-guidance-report.pdf.

Whitman, R.L., Nevers, M.B., 2008. Summer E-coli Patterns and Responses along 23 Chicago Beaches. Environmental Science & Technology 42(24) 9217-9224.

CHAPTER 3 THREE DIMENSIONAL MODELLING TO ASSESS CONTRIBUTIONS OF MAJOR TRIBUTARIES TO FECAL MICROBIAL POLLUTION OF LAKE ST. CLAIR AND SANDPOINT BEACH [1]

3.1 Chapter Synopsis

The presence of high levels of *E. coli* in Lake St. Clair is a significant concern for millions of people, in Canada and the United States, who rely on that water source for drinking, fishing and recreational purposes. A combination of mathematical modelling and monitoring techniques in the lake can provide an efficient and cost-effective framework for the management of microbial pollution of beach waters, as well as serving as a timely reporting tool to communicate associated human health risks from recreational use. In this paper, a high-resolution 3D hydrodynamic model is developed and applied to assess the flow and microbial contribution of major Lake St. Clair tributaries during the Summer of 2010. The model skill in reproducing water temperature is in good agreement with the observed data (γ^2, NRMSE, R^2 and WS values of 0.12, 0.37, 0.88 and 0.96, respectively). Assuming *E. coli* input estimates to be conservative, the model results show that the maximum predicted fecal concentrations from the combined input of the major tributaries to be <100 CFU/100 ml for most of the lake. The corresponding maximum at Sandpoint Beach was < 10 CFU/100 ml. High dominant flow with low *E. coli* input from the St. Clair River and microbial decay due to residence time in the lake are largely responsible for the results obtained. The results evidently indicate that the four major tributaries are unlikely

[1] This Chapter was published as a journal article: Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020. Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach. Journal of Great Lakes Research. https://doi.org/10.1016/j.jglr.2019.12.005

responsible for the observed *E. coli* exceedances of the Ontario safety guidelines for recreational activity at Sandpoint Beach.

3.2 Introduction

Escherichia coli (*E. coli*) is a common inhabitant of the intestines of humans and other warm-blooded animals and can eventually leak to the natural environment. Fecal matter contains high numbers of *E. coli* including pathogenic organisms from sick individuals. Epidemiological studies have shown a strong correlation between *E. coli* and waterborne disease outbreaks from the recreational use of beach waters (Fewtrell and Kay, 2015; Field and Samadpour, 2007; Sokolova, 2011; USEPA, 2010). Canadian Drinking Water Quality Guidelines (Health Canada, 2012) and the United States Environmental Protection Agency criteria (USEPA, 2012) use *E. coli* as an indicator of microbial water quality for drinking and recreational purposes. Particularly, in Ontario, Canada, the Ontario Ministry of Health and Long-Term Care sets out *E. coli* concentration of 100 colony-forming unit (CFU)/100 ml, updated to 200 CFU/100 ml in 2018 (OMHLTC, 2018), as a safe limit for swimming and bathing at freshwater beaches. There are several public beaches on Lake St. Clair where microbial water quality standards for *E. coli* for recreational use are frequently exceeded during the summer months (McPhedran et al., 2013), A review of more than one hundred years (1900-2010) of ecological and socioeconomic characteristics of Lake St. Clair concluded that the major issue previously was waterborne illnesses due to contaminated drinking water, but currently contaminated recreational waters and coastal pollution are the main concerns (Baustian et al., 2014).

Lake St. Clair is a relatively shallow water body surrounded by large urban areas and receives discharges from many tributaries along its southern shore. Its beaches are

especially susceptible to high microbial contamination. Major sources of microbial contamination of beaches and resulting *E. coli* exceedances might include fecal pollution from combined sewer overflow, sanitary sewer overflow, stormwater flows and direct input from birds and other animals feces. These sources are dynamic over time and have multiple entry points into watersheds and tributaries that feed into the lake, making it very difficult to identify specific sources of *E. coli* exceedances at beaches (He et al., 2016; Simpson et al., 2002). Current culture-based methods used for *E. coli* monitoring are unable to support management effectively. Next-generation sequencing (NGS) based monitoring methods show a lot of promise both for helping in identifying potential sources of fecal pollution as well as reducing the time required to obtain analytical results, but there are still limitations on the quantitative interpretation of NGS data (Tan et al., 2015).

Additional tools to aid in the monitoring and management of microbial pollution at beaches are mathematical models used to predict contaminant fate and transport in recreational waters. If properly validated, such models can have numerous benefits over conventional methods, as they are very cost-effective and do not require continuous field sample collection and analysis. Model predictions can be real-time which could potentially help reduce or eliminate both false positive and negative results. Models can also help to identify or exonerate potential sources of microbial pollution responsible for *E. coli* exceedances at a given beach. Nevertheless, monitoring data will still be required for possible model upgrades, as well as for periodic testing and calibration of the model to account for changing environmental conditions.

Fate and transport of microbial pollution in a lake is largely influenced by the hydrodynamics, the microbial load (from tributaries), and the die-off (decay) of the

microbes which depend on the ecological conditions in the lake. The hydrodynamics in a lake environment are affected by meteorological forcing at the air/water interface (where physical processes such as heat, kinetic energy, momentum, and matter occurred), the lake bathymetry and the lake bottom friction. Expectations are that a 3D unsteady-state model will adequately predict the spatial and temporal variability in hydrodynamics within the lake. Recent studies have shown the benefits of applying unsteady-state 3D hydrodynamic models with biogeochemical lake processes for management studies in the Great Lakes (Bocaniov et al., 2016; Bocaniov et al., 2014; Hamidi et al., 2015; Leon et al., 2011; Leon et al., 2012b; Oveisy et al., 2014; Valipour et al., 2016) and particularly for Lake St. Clair (Anderson and Schwab, 2011; Anderson et al., 2010; Bocaniov and Scavia, 2018).

Bocaniov and Scavia (2018) further extended their 3D hydrodynamic model to simulate nutrient mass balance in the lake and investigate the loss rate of nutrients and their correlation with transport time scale. In a limited regional study, Holtschlag et al. (2008) used the 2D-RMA2 Hydrodynamic model and reverse particle-tracking analyses for 10 selected events to identify the Clinton River and Clinton Cutoff Canal (Spillway) as potential sources of *E. coli* bacteria for two beaches (Memorial and Metropolitan beaches) on the U.S. side of Lake St. Clair. Microbial fate processes such as deactivation (based on three key environmental factors: light intensity, salinity and water temperature), sedimentation, resuspension, grazing etc. are factors that have the potential to influence microbial distributions in the aquatic environment (de Brauwere et al., 2014; Liang et al., 2017; Liu et al., 2006). Many studies have used *E. coli* to model the microbial distribution in lakes (Bonamano et al., 2015; Chan et al., 2013; Liu et al., 2006; Sokolova et al., 2012). However, no such approaches have been applied for a more comprehensive evaluation of

the impact of fecal microbial pollution on microbial water quality in Lake St. Clair or its beaches.

For the current study, the Aquatic Ecosystem Model (AEM3D), based on the ELCOM-CAEDYM model developed at the Centre for Water Research (CWR), University of Western Australia (Hodges, 2000), is used as the modelling platform. The objective of this study is to estimate the contributions of major tributaries to fecal microbial pollution of Lake St. Clair and Sandpoint Beach, located on the southern shore. A high-resolution 3D hydrodynamic model was developed and validated with field measurements of water temperature and velocity profiles. Assuming constant microbial loads from the tributaries, the potential lake-wide *E. coli* influenced zones are delineated to identify those places with higher exposure risks to incoming riverine microbial contamination.

3.3 Methods and Materials

3.3.1 Study Area:

Lake St. Clair (42.2956-42.6901 N- Latitude, 82.4119-82.9258 W-Longitude) is a water body that forms part of the binational boundary water between Canada and United States in the Great Lakes Basin (Figure 3-1). Its surface area is roughly 1,114 km^2 and 59% of the local watersheds that drain to Lake St. Clair are on the Canadian side (8,988 km^2), while the remainder (6,317 km^2) is on the U.S. side (Baustian et al., 2014). The St. Clair River, which drains Lake Huron and is the main inlet of Lake St. Clair, has a drainage area of about 576,013 km^2 and delivers water at a rate of about 5,200 m^3/s (98.2% of total inflow) (Holtschlag et al., 2008). The remaining water inflow to the lake (1% of total lake inflow) is mainly from the three next largest tributaries: Thames River and Sydenham Rivers in Ontario, Canada and Clinton River in Michigan, United States. Negligible flow from about

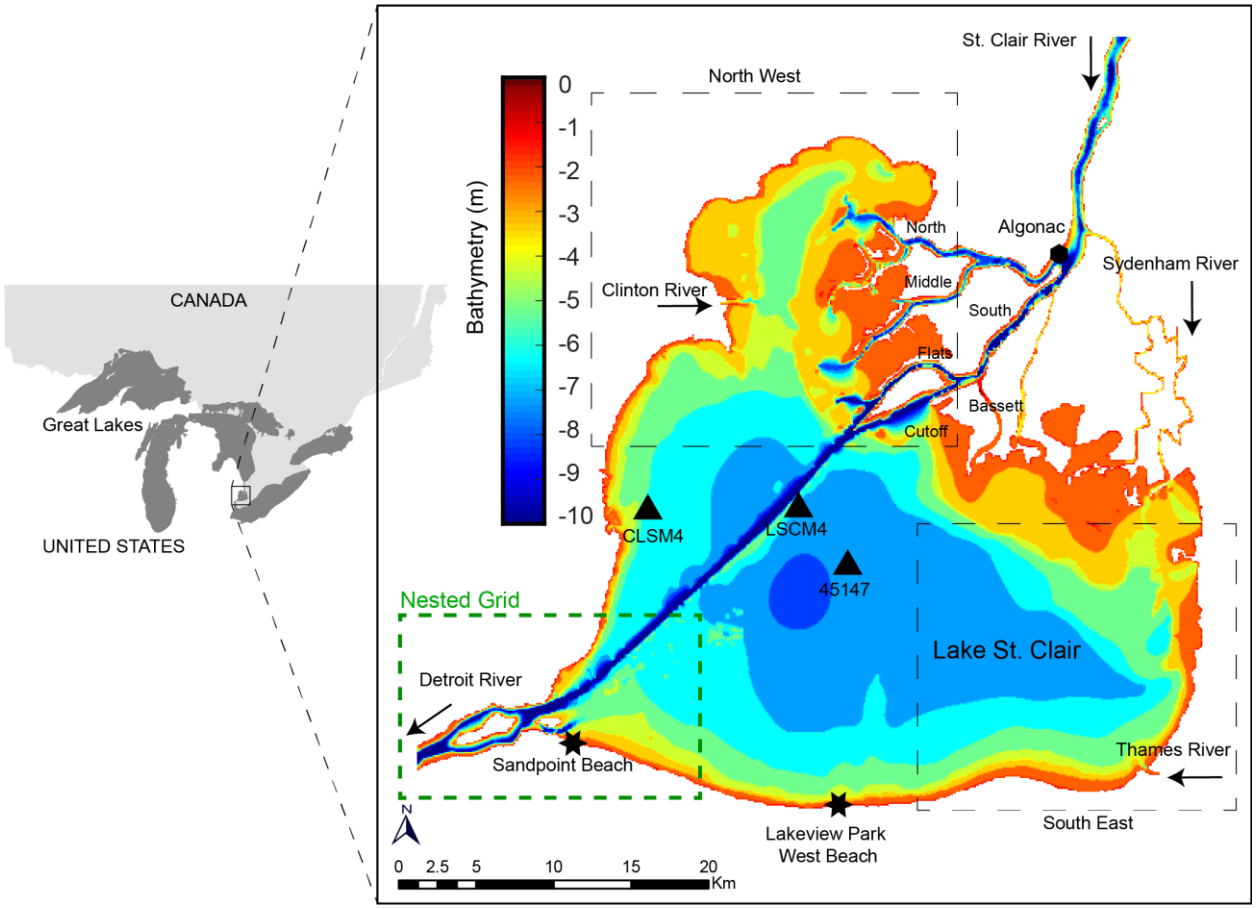

Figure 3-1: Map of Lake St. Clair and its bathymetry. Triangle symbols show the locations of the moorings (CLSM4 (*42.471 N 82.877 W*), LSCM4 (*42.465 N 82.755 W*) and 45147 (*42.430 N 82.680 W*)). Location of Sandpoint Beach and area of nested model (see Model Description and Analysis Approach section) are identified by black stars. Bathymetric data are in meters and all model tributaries are labelled on the map.

10 smaller tributaries located in the south of the lake is not considered in this study. The Detroit River (discharge rate of about 5,270 m^3/s) is the only natural outlet from the lake. Evaporation and withdrawals for water supply are associated with other water losses.

Water Quality and Data Collection:

Lakeview Park West Beach in Lakeshore, and Sandpoint Beach in Windsor, are two Canadian public beaches along the Lake St. Clair's southern shoreline. According to the Windsor-Essex County Health Unit (WECHU) report and field measurements in 2010,

both the Lakeview Park West and Sandpoint beaches have had incidents involving high bacterial counts (McPhedran et al., 2013). Sandpoint Beach is a popular beach on the Canadian side of Lake St. Clair. A review of weekly data collected by the WECHU during the summer months over a decade (2005-2014) showed that the Ontario Ministry of Health and Long-Term Care (OMHLTC) standard of 100 CFU/100 ml was exceeded more than 50% of the time (Spalding, 2014).

More intense daily *E. coli* data was collected at the same beach over 30 consecutive days during late Summer 2010 (August 10 – September 9) by McPhedran et al. (2013). The data shows the *E. coli* concentrations to be quite dynamic with several exceedances of the OMHLTC standard. Reasons for the *E. coli* exceedances and the dynamic variation observed are largely unknown. Inputs from St. Clair, Clinton, Thames and Sydenham rivers account for >99% of all flows into Lake St. Clair, while Thames and Clinton are dominant amongst contributors from urbanized watersheds. Thus the four combined rivers are expected to also be the dominant contributors to the inputs and budget of all water quality parameters in Lake St. Clair, including fecal microbial pollution.

3.3.2 Bathymetry and Forcing Data:

The bathymetry, sourced from National Geophysical Data Center (https://www.ngdc.noaa.gov/), has depths referenced to a generic datum of 176.784 meters (580 feet) and is a rectangular grid with cell sizes of ~70 m by ~90 m that was processed into coarser grid sizes (100 m, 200 m and 400 m uniform grids). Meteorological forcing data sets include wind direction, wind speed, air temperature, humidity, atmospheric pressure, rainfall, and solar radiation. Data is from the nearest stations (Windsor A. Station ID=4716; Windsor Riverside Station ID=4715) and moored surface buoys (Buoy (West Erie) 45005 (41.677

N 82.398 W); Buoy 45147 (42.430N 82.680W); LSCM4 (42.465N 82.755W); CLSM4 (42.471N 82.877W)). Buoy 45147, which is maintained by Environment and Climate Change Canada (ECCC), was used to fill out missing wind direction and wind speed data and for model validation of water temperature. Solar radiation data were obtained from the Canadian Regional Climate Model (CRCM5) provided by the University of Quebec at Montreal (UQAM) (Huziy and Sushama, 2017). (See Table 3-S1 for more information)

3.3.3 Inflows and Boundary Conditions:

Flow and temperature data of the four major tributaries to the lake (Thames, Sydenham, St. Clair, and Clinton rivers) and the outflow (Detroit River) were obtained from the nearest gauged stations as follows. St. Clair River daily flow data is available at Port Huron station (02GG014) from 2009 to 2012. Since this station is upstream of the model boundary, flow discharged from other major tributaries such as Black River, Pine River and Belle River is also added. Note that the tributary stations tend to be upstream of their outlets so prorating based on drainage area may be required. Smaller tributaries such as Deer Creek (US side), Cray Creek, and Baby Creek (Canadian side) are not included. Gaps in daily observed St. Clair River water temperature, obtained from the National Data Buoy Center (NOAA) station at Algonac (Station ID: 9014070), were filled using regression with daily air temperature in Sarnia. The best regression was obtained with a 35-day average air temperature, which seems reasonable as the water from a large lake like Lake Huron would change slowly with air temperature.

For the Sydenham River, flow from the most downstream flow stations on it's two main branches, Sydenham River at Florence (station 02GG003, drainage area 1150 km^2) and Bear Creek below Brigden (02GG009, drainage area 536 km^2), are extrapolated to the

watershed outlet using the drainage area ratio method (Emerson et al., 2005). Note that the Florence and Brigden stations combined drainage area is 1686 km^2, which is 63% of watershed outlet near Wallaceburg and is at the borderline for using the drainage area ratio method. Sydenham water temperature was available at Wallaceburg (115 observations from 2002-2015) from the Provincial Water Quality Monitoring Network (PWQMN) (station 04002701702). The best regression of observed water temperature was obtained with 7-day average air temperatures (R^2=0.94) from the Sarnia climate station (Station ID=44323) and was used to fill the missing values.

Thames River flow data is available at Thamesville (02GE003), which is approximately 77% of the total drainage area of Thames River. Flow rates were extrapolated to the watershed outlet using the drainage area ratio method (Emerson et al., 2005). Thames River water temperature was also available at Thamesville (73 observations from 2002-2014) from the PWQMN (station 04001305802). Using air temperature from London WSC climate stations, regression curves were fitted with the existing water temperatures; the resulting best regression was with 3-day average air temperatures (R^2=0.89); this regression curve was then used to predict the water temperatures for missing days.

Flow and water temperatures for the Clinton River were obtained from the United States Geological Survey (USGS) data web site (USGS station 04165500, Clinton River at Moravian Drive). Daily streamflow of the Detroit River is available from 2009 to 2014 at Fort Wayne station (02GH015) and used directly. Satellite-derived lake-wide daily time series of water surface temperature in 2010 was gathered from NOAA (Coast Watch Great Lakes Surface Environmental Analysis, GLSEA: https://coastwatch.glerl.noaa.gov/ threads) and was used to initialize the model and for model validation.

3.3.4 Model Description:

In the current study, a 3D unsteady numerical model of Lake St. Clair was set up for a five-month simulation period (May-September) of 2010 using the AEM3D hydrodynamic modelling framework. The first month is considered as warm-up period and results are not shown here. AEM3D solves hydrostatic Reynolds-averaged Navier–Stokes equations, based on a semi-implicit scheme which is adapted from a three-dimensional model known as TRIM approach (Cheng et al., 1993), for heat and momentum transfer on a Cartesian Arakawa C-grid with a fixed z-coordinate finite difference mesh (Hodges et al., 2000). It uses the scalar transport equations to simulate spatial and temporal variations of mass, temperature and salinity distributions (Hodges and Dallimore, 2006). Proposed models have been applied extensively to study lake processes and for biogeochemical and management studies in the Great Lakes (Anderson and Schwab, 2011; Bocaniov et al., 2016; Bocaniov et al., 2014; Leon et al., 2011; Leon et al., 2012b; Oveisy et al., 2014; Valipour et al., 2016; Zhao et al., 2012). AEM3D is selected in this study because it was proven that it is able to provide a spatially explicit simulation of the hydrodynamics and tracer simulation in a large lake. Furthermore, AEM3D has been well documented (Hodges and Dallimore, 2006) and tested previously in the Great Lakes system by the authors (Leon et al., 2012b; Valipour et al., 2016; Valipour et al., 2018).

To resolve Lake St. Clair geometric data, three structured, uniform grids (400 m, 200 m, 100 m) were used. Eight vertical layers with 2-meter thicknesses each were considered for all the models. In order to better resolve the bathymetry and study the detailed current, river distribution, temperature structure and flushing time of nearshore areas of Sandpoint Beach, a 50 m nested-mesh was considered to allow for a finer grid configuration in the

area of interest using an approach similar to Rao and Sheng (2008) (see Table 3-S2 of Appendix 3).

Flow and scalar concentration from the coarse grid was used as boundary conditions for the nested grids. A smaller fixed time-step (60 s) was required in the nested model to satisfy the Courant-Friedrichs-Lewy (CFL) condition. Information transfer between the coarse and fine-grids in the nested model is a one-way data transformation with results of the coarse grid model interpolated to the boundary of the fine grid. All the models were started from rest and initialized with surface temperatures from GLSEA data. The daily temperature profiles (assumed vertically uniform) measured in the last week of April 2010 were averaged and linearly interpolated to the grid points as initial conditions of the basin (see Figure 3-1).

For bottom boundary condition, two options were tested; i) turbulent benthic boundary layer (TBBL) and ii) constant drag coefficient on all surfaces (drag-all). TBBL was selected because it re-produced the currents obtained from Acoustic Doppler Current Profilers (ADCP) well in Lake St. Clair compared to the drag all option (See Figure 3-S2 in Appendix 3). Meteorological forcing was applied uniformly on the entire free surface of the lake. In the wind-mixed layer, the momentum input of the wind is modelled using the stress boundary condition at the free surface. The wind stress is a function of wind speed at a 10-meter elevation (U_{10}). The drag coefficient at the surface and in the bottom was assumed to be constant. Because all biochemical processes are temperature dependent, it is important that the models properly simulate the surface and sub-surface temperature structure in large lakes (Leon et al., 2012a). In this regard, a comprehensive heat and energy balance model was implemented that considers non-penetrative components of long-wave

radiation, sensible heat transfer, evaporative heat loss, and penetrative shortwave radiation across the free surface (Hodges and Dallimore, 2006).

3.3.5 Analysis Approach for *E. coli*:

We modelled contribution to fecal microbial pollution of Lake St. Clair and Sandpoint Beach by the major tributaries using *E. coli* as the indicator for fecal pollution. The assessment of the contribution was done in two parts. First, separate model simulations were carried out for each tributary for the entire 2010 summer simulation period with a constant inlet *E. coli* concentration of 100 CFU/100 ml to determine its fractional contribution to fecal microbial pollution. Second, conservative estimates of microbial pollution loads were used for all tributaries and model simulations were carried out for the entire 2010 summer period to simulate the combined effect, distribution of microbial contamination and relative contribution of all the tributaries on microbial water quality.

Microbial (*E. coli*) concentrations were simulated assuming first-order decay, with decay rate estimates from relevant field-based studies. However, *E. coli* is also associated with sediment transport and bed mobilization, especially during storm events (Droppo et al., 2011). Upon receiving the water to the lake environment other physical processes such as sediment transport and resuspensions, as well as *E. coli* inactivation by UV, grazing and die-off due to water temperature and other chemical properties of water (pH, dissolved oxygen etc.) may influence the fate and transport of *E. coli* (Brookes et al., 2004). Riverine inputs get diluted with the lake-wide water, which also is involved in the process of *E. coli* reduction. The combined influence of all factors affecting the decay rate is complex and expected to be site-specific and variable with time. No model, that can universally define

Table 3-1: *E. coli* decay rate reported in the literature for different locations and temperature conditions.

#	Decay (d^{-1})	Temperature	Location	Reference
1	0.78-1.28	20 °C (August)	Lake Rådasjön – Sweden (microcosms)	(Sokolova et al., 2012a)
2	1.417	18 °C	Hamilton Bay, Lake Ontario, Canada	(Crane and Moore, 1986)
3	0.62-0.75	Sunny and Variables days (45 cm and 90 cm)	Lake Michigan at 63rd St. Beach, US	(Whitman et al., 2004)
4	0.55-1.23	28.8-30.0 °C	Lake Weija - West Africa	(Ansa et al., 2011)
5	0.83	summer	Polishing ponds-municipal sewage (Field)	(Toms et al., 1975)
6	0.5-0.72	Dark Death rate (10-35°C)	Onondaga Lake, NY	(Auer and Niehaus, 1993)

its variability, is currently available. However, there are some simplified decay rate formulas.

Good agreement and wider applicability have been shown with simplified lumped parameter first-order decay rates that represent the net effect of all the factors. In the current study, measured first-order decay rates were selected based on the literature (Table 3-1). The selected values vary between 0.5-1.7 d^{-1} and are similar to the range of *E. coli* decay rates used in similar freshwater modelling studies (Chan et al., 2015; Liu et al., 2006).

E. coli was modelled as a tracer transport in AEM3D. Tracer transport in AEM3D uses an explicit approach and has an advective CFL such that $\mathrm{CFL_a} = u\,\Delta t/\delta x < 1$ is required (Hodges and Dallimore, 2006). Tracer concentration is advected using a conservative ULTIMATE QUICKEST discretization that allows it to exceed the CFL condition for velocity, without producing numerical instability (Hannoun et al., 2006). The tracer transport module implemented in this study includes the processes of advection, dilution due to mixing, and constant decay rate. Model simulations were carried out with four decay rates (0.5 d^{-1}, 0.7 d^{-1}, 1.4 d^{-1}, 1.7 d^{-1}) to span the range of measured values presented in Table 3-1.

Table 3-1 summarizes the studies that reported *E. coli* decay rates. In addition, summarized data from 12 studies (total of 49 samples) were used to find the natural mortality ("dark death") rates as a function of temperature. Mortality rate at dark and 20 °C was calculated as 0.482 d^{-1} which is close to our lower range; however, it does not account for solar inactivation, which increases the decay rate. In reality, many other processes such as sediment settling and resuspension, grazing, etc. may also influence the concentrations of *E. coli* in the water but here all are included in the simple lumped parameter first-order decay rate model. During the simulation period, lake-wide mean monthly water temperatures of July and August are the warmest (23.9 °C and 23.8 °C, respectively) and June and September are the coolest (19.8 °C and 20.3 °C, respectively). Based on the correlation proposed by Sokolova et al. (2012), the decay rate for warm and cool months are 1.46 and 0.85 d^{-1}, respectively. These values are within the range that is selected based on previous literature.

Water age simulation was carried out to understand how its spatial variation in the different regions of the lake and, especially, along the edges can affect the fate and transport of the *E. coli*. Water entering the lake from the tributaries was considered as "zero" age. A detailed description of the water age simulation method was proposed by Anderson et al. (2011). Water age is an aggregated measure of the time elapsed, since entry into the lake, for the water to reach the location of interest. In AEM3D, water age was calculated based on a similar definition and approach. Using a scalar simulation and assuming water to be conservative (neither created nor destroyed), the governing advection-diffusion transfer equations were then applied to calculate the water age.

To calculate the local flushing time, the approach presented by Zhao et al. (2012) was followed in the nested model for the southwestern region of Lake St. Clair (see Figure 3-1). The flushing time is defined as the time necessary for a conservative substance concentration to decrease to 1/e (~ 0.37) of its initial concentration. Further details are provided in Appendix 3-2.

3.3.6 Approach for Model Validation:

To validate the modelled water temperature, the Great Lakes Coast Watch product called GLSEA was used. GLSEA provides gridded (~1,400 m) mean surface temperatures and it largely reflects offshore conditions. GLSEA grids that overlap the shoreline are removed from the GLSEA water temperature estimate. For three days selected in June, July, and August 2010, the simulated temperature is re-gridded to the same resolution of GLSEA and aggregated daily.

To define the zones mainly affected by the dispersion of bacteria that comes from the major tributaries into Lake St. Clair, and their potential effect on the bathing area of Sandpoint Beach, the Microbiological Potential Risk Area (MPRA) is used. It is defined as the lake area over which the *E. coli* concentration is greater than or equal to 1 % of the concentration measured at the mouth of the rivers discharged to the Lake (Bonamano et al., 2015; Bonamano et al., 2016). A tracer with a continuous concentration of 100 is discharged into the lake from each tributary and 90th percentiles of its concentration at the discharge areas during the entire summer were calculated. The lowest *E. coli* decay rate (k= 0.5 d^{-1}) was used in the model to find the maximum possible area of potential risk in the lake.

The skill of the model is measured by the variance of the simulation errors divided by the variance of the observations (henceforth denoted by γ^2). In general, the smaller the γ^2, the

higher the skill of the model. Also, the goodness of fit between observations and model output temperatures was quantified in terms of the estimated root mean square error (RMSE), normalized RMSE (NRMSE), and correlation coefficient. There are several methods for calculation of NRMSE to quantify the agreement between model output and observation (Acosta et al., 2015; Bravo et al., 2017; Hamidi et al., 2015; Trolle et al., 2014). Trolle et al. (2014) derived NRMSE by dividing the RMSE by the variable sample mean, for Chlorophyll_a, total nitrogen, total phosphorus, dissolved oxygen, and temperature. Here, the method proposed by Acosta et al. (2015) which calculate NRMSE by dividing the RMSE by the range of data ($x_{max} - x_{min}$) is followed. To further analyse the model performance, skill value (WS) proposed by Chen and Liu (2017) is used. WS of 1.0 represents perfect performance while values in the range of 0.65-1.0, 0.5-0.65, 0.2-0.5, and <0.2 indicate excellent, very good, good, and poor performance, respectively.

3.4 Results and Discussion

Time series of Lake St. Clair daily river inflows and outflow from June to September in the summer of 2010 are shown in Figure 3-2. Data showed that 98% of the lake inflow is from the St. Clair River with an average flow of 5,169 m^3/s (minimum 4,564 and maximum 5,611 m^3/s). The Thames River contributes about 1% of the inflow with an average of 29.4 m^3/s during the simulation period. Two peaks were observed in the Thames River flow data of 249.1 m^3/s and 108.8 m^3/s during the thunderstorm events of June 05 and July 23 (daily precipitation of 44 and 42 mm/d), respectively. Sydenham River, with an average flow rate of 6.6 m^3/s, experienced peak flow of 68.8 m^3/s during the June 5 event. In the same manner, Clinton River had an average flow rate of 9.8 m^3/s with a peak of 100.2 m^3/s. The contribution of Thames, Sydenham and Clinton rivers average inflows are not

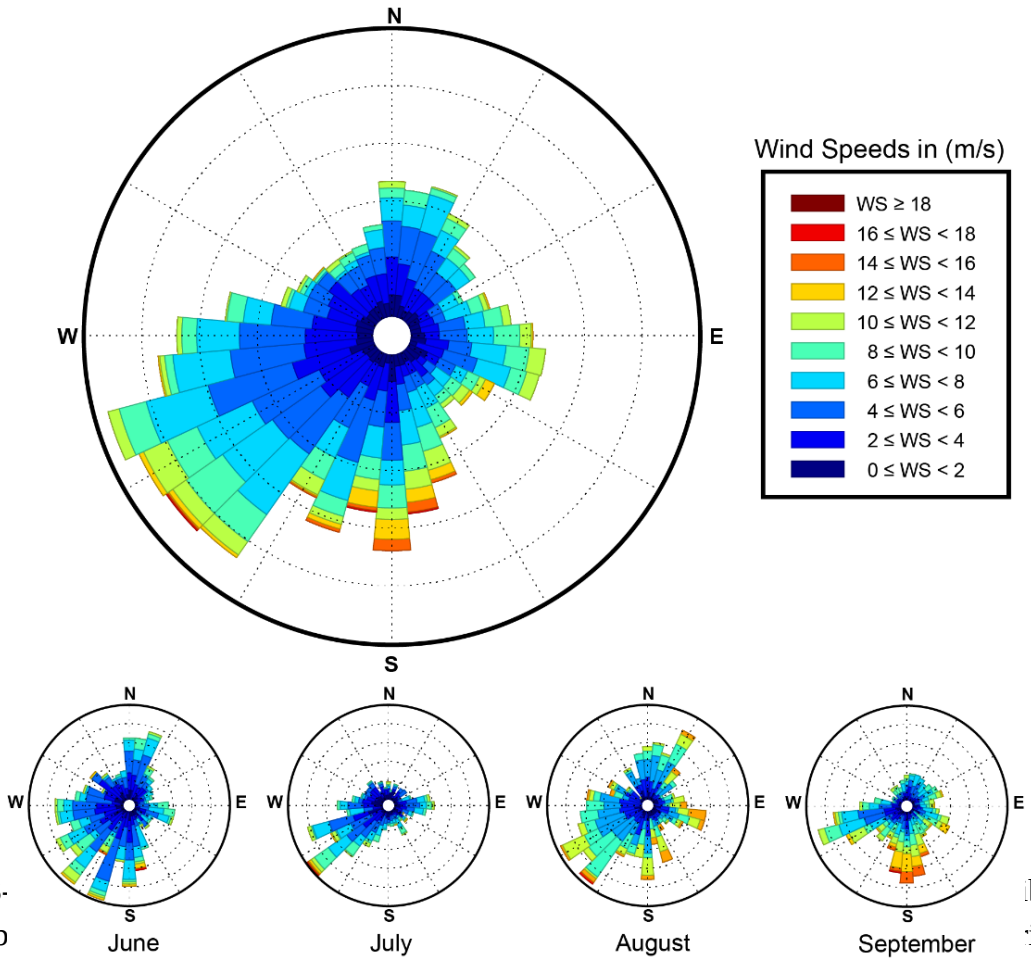

Figure 3- butaries, used as b ibutaries so it is plotted on the right Y-Axis in different scale.

Figure 3-3: Wind rose illustration of wind speed and direction for 2010 simulation period and for each month separately

comparable to the St. Clair River but in the case of big storm events, these tributaries can increase in flow 10 times from their average flow and their contribution increases accordingly.

In the case of nutrient contribution, the Thames, Sydenham and Clinton rivers are more important, contributing ~23% of 2009 and ~14% of 2010 total phosphorus load (Bocaniov and Scavia, 2018). This shows that they can affect the Lake nutrient cycle and they can possibly have major impacts on the microbial contamination of the Lake as well. The higher flow of the St. Clair River is able to assimilate incoming dissolved and suspended

nutrients and also microbial loads from its tributaries with little effect upon its overall water quality.

The daily mean wind speed and direction for the summer of 2010 at Windsor Airport was used as the forcing input for the model (Figure 3-3). The dominant wind direction is southwest. During September, high wind speeds from the south affect the lake, while in July, SW seems to be the only wind direction. Also, wind speed tended to be higher in August and September than during June and July. The predominant wind forcing is from the southwest (SW) direction. In September, very strong south-winds are common in the nearshore region of the south basin because of temperature differences between the lake and the air moving over it. This forms gyres in the nearshore area that usually start in the middle of the lake and continued all along the edge, and end at the Detroit River. The dominant wind direction of west or southwest is most likely to promote transport of surface water from the upstream of the mouth of the Detroit River into the beach area. Unless a strong east or northeast wind changes the situation, the dominant current along the shoreline from Lakeview Park West Beach to the mouth of Thames River, is on a west-east direction.

3.4.1 Model Validation

The developed 3D model was validated against a comprehensive dataset collected in Lake St. Clair and from previous studies and performed to a reasonable degree of accuracy. Validation of the flow distribution in the St. Clair River channels, thermal structure of the lake, and lake hydrodynamics are presented below.

3.4.1.1 Flow Distribution in the St. Clair River Channels

Flow distribution in the St. Clair River channels changes as a function of water levels and wind stress (Anderson et al., 2010), and eventually affects the path and proportion of the cooler water that comes from the higher latitude of Lake Huron. Due to the ecological impacts that circulation and water temperature have on the lake, it is important to accurately simulate the flow distribution in the St. Clair River channels. Inflow water temperatures (Figure 3-2b) show a seasonal trend, but there are more fluctuations in the small tributaries compared to the St. Clair River. Water from the St. Clair River, has lower temperatures compared to the other tributaries and affects the whole water thermal structure in the lake because of its dominant flow contribution.

Our simulation results reveal that the average discharge distribution in the channels is similar to the previous field measurements and modelling studies (North=34.79%, Middle=20.09%, Flats=17.25%, St. Clair Cutoff=24.7%, Bassett=3.15% (Anderson et al., 2010; Bolsenga and Herdendorf, 1993; Holtschlag and Koschik, 2001, 2002a, b; Schwab et al., 1989)). Current patterns in Lake St. Clair are dominated by the wind conditions (Schwab et al., 1989) which affect water temperature by changing heat transfer coefficients and evaporation heat loss. In addition, water temperature has been influenced by the variation in flow distribution in the St. Clair River's channels.

3.4.1.2 Thermal Structure of Lake St. Clair

Although no calibration and parameter adjustment was done in developing the model, its validation was done carefully because the adequate evaluation of the hydrodynamic model (and its thermal structure, which is an indirect way of validating model performance with respect to hydrodynamics) with field measurement is essential for having confidence in the

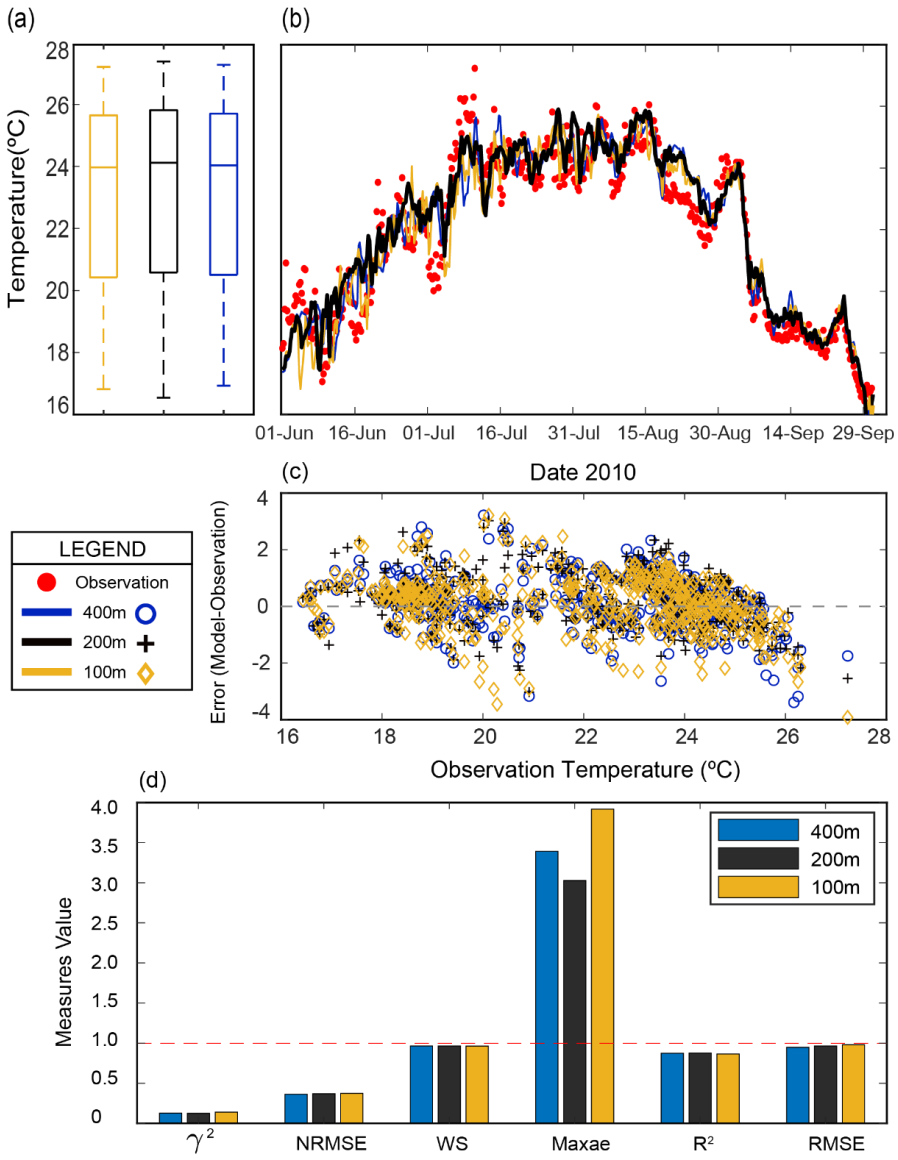

Figure 3-4: Simulated water temperature calibration results for each model grid: (a) boxplot (b) time series comparison to observed (c) model error versus observed (d) model performance measures.

accuracy of subsequent water quality simulations. In addition, water temperature plays a crucial role in the life cycle and habitat distribution of aquatic species and hence its accurate simulation is necessary. Three uniform grids (400 m, 200 m, 100 m) were used to simulate water temperature and the results were compared with buoy measurements in the middle of the lake and satellite imagery of the entire lake. Also, model performance is assessed at

Sandpoint Beach using the nested 50 m grid model to compare water temperatures with field measurements. To summarize, we validated the performance of the model at three different levels: i) using buoy measurement in the middle of the lake. ii) using lake-wide temperature predictions and comparing model results for three select day in each of June, July and August 2010 with values provided by GLSEA through analysis of satellite-derived data (Figure 3-7), and iii) nearshore temperature predictions by comparing model results with one month of daily water temperatures recorded at Sandpoint Beach in August – September 2010 by McPhedran et al. (2013) (Figure 3-8).

Figure 3-4 shows model results comparison with buoy LSCM4 water temperature data and the evaluation statistics. Simulations for the study period show the RMSE between observed and simulated surface water temperature to be around 1 °C, which is less than 5% of the mean time-averaged surface water temperature for the simulation period. In the left panel box plot the model output results are not significantly different ($P>0.05$). The 200 m grid simulation has a higher median and spread of temperatures. As can be seen in Figure 3-4c, the 200 m grid also has the lowest maximum absolute error. Even though a finer grid model may provide for higher domain coverage in the edges, it does not necessarily predict the field data with greater accuracy and, as can be seen in Figure 3-4c, it may even be characterized by very high absolute errors at some locations. Comparing γ^2, NRMSE, R^2 and WS for all three grids show that there are no statistical differences observed in selecting the higher resolution grid. Furthermore, because of the nature of uniform grid sizes, moving from the 200 m to 100 m grid size increases the computational time from 5.5 hours to around 42 hours on a single-core Intel® Xeon® Processor E3-1245 v6, 8M cache, 3.70 GH.

Generally, to meet the CFL stability condition, increasing the model resolution from 400 m to 100 m or less requires reducing the time step. However, in our study, except for the 50 m grid size, which needs a 60 second time step, a 300 second time step worked for all the other models. Applying the nested modelling approach can control the problem of computational expense to a certain extent (Leon et al., 2012). We abandoned the idea of reducing the grid size lower than 100 m for the entire lake-wide model due to the high

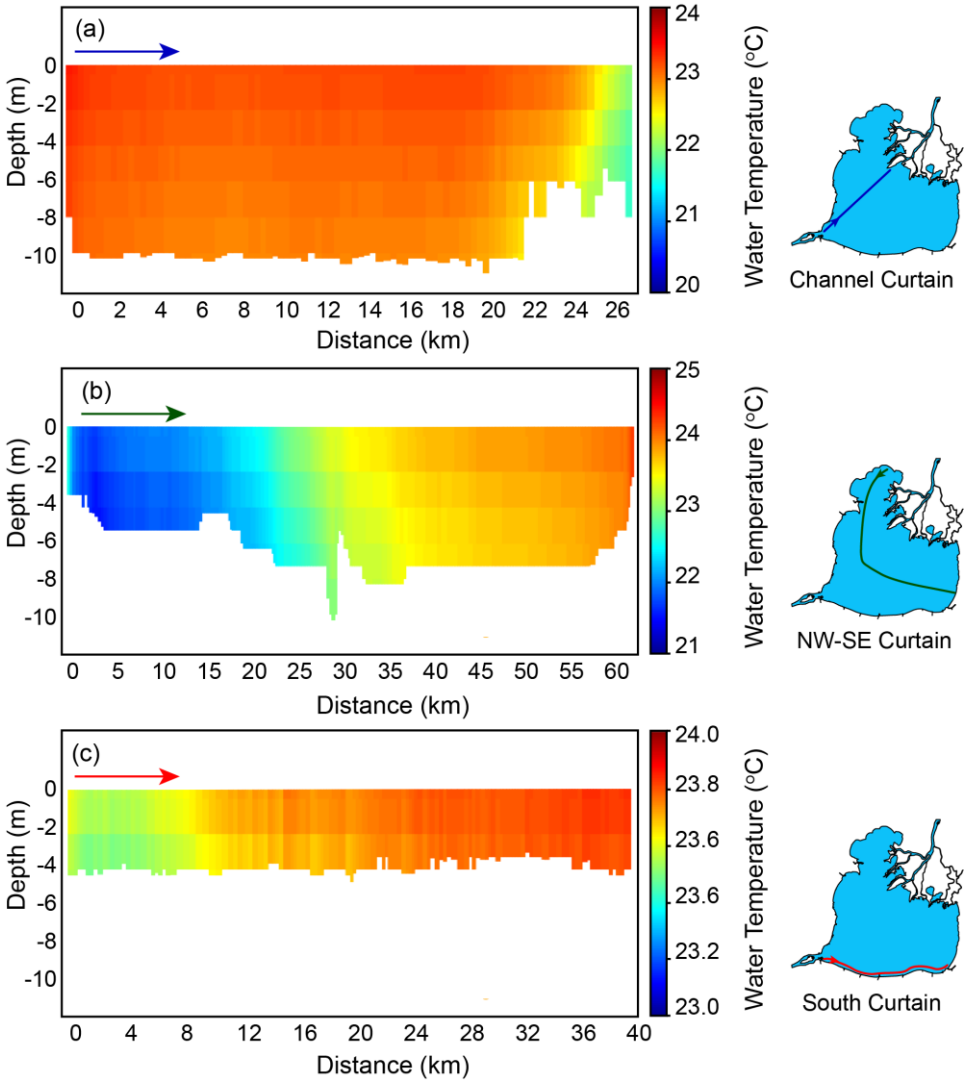

Figure 3-5: Curtain plot to show the stratification along three cross sections at different locations of Lake St. Clair for mean water temperature during the simulation period (June-September 2010) (a) along the channel (b) from North West to South East (c) along the south shoreline. (The temperature range are different for each subgraph)

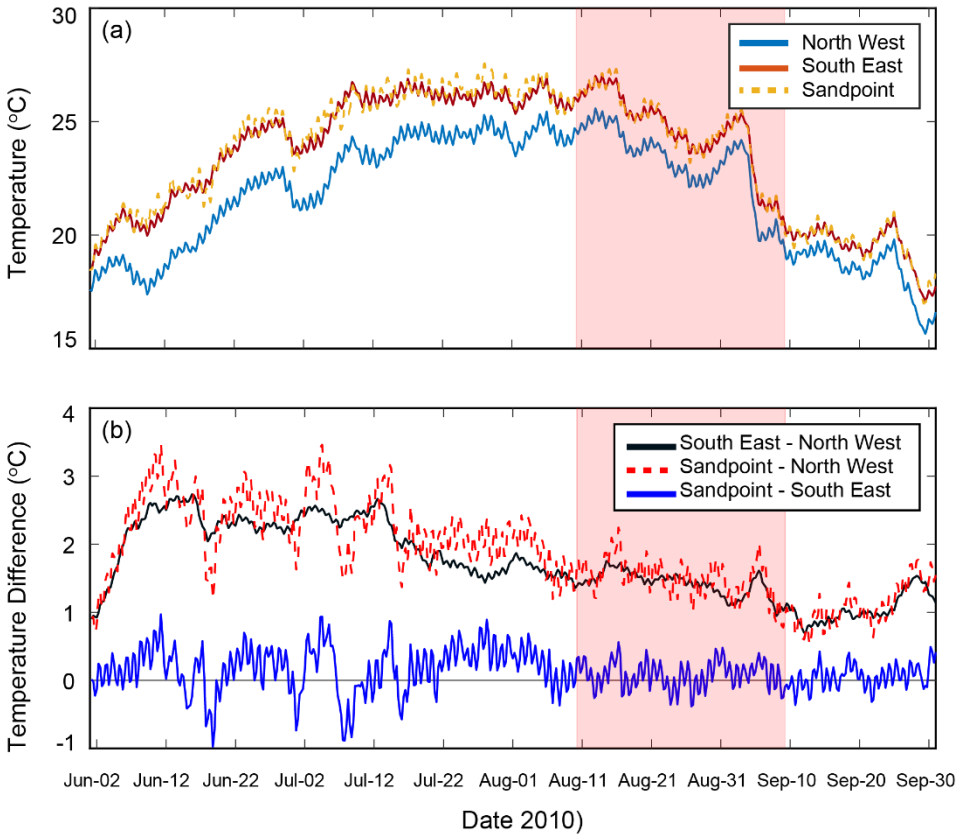

Figure 3-6: (a) Time series of simulated water temperature at different zones (North West, South East, and location of Sandpoint Beach) of Lake St. Clair and (b) the water temperature differences between North West, South East, and Sandpoint Beach. Red shade shows one-month study period of McPhedran et al. (2013)

computational time, compounded by the reduction in the time step forced by the higher resolution grid. It takes 12 days to run the model for the entire lake with a 50 m grid size and a 60 second time step.

The time series in Figure 3-4a show that all the models are able to predict both small- and large-scale fluctuations in water temperature for the observed period, especially during rapid weather changes in late August and early September. In the end, the 200 m grid model was selected as the base configuration to cover our study area. Model-predicted temperature values were in good agreement with the observed data with values for γ^2,

NRMSE, R^2 and WS of 0.12, 0.37, 0.88 and 0.96, respectively. Other measures are presented in Figure 3-4c.

Model outputs were used to understand the circulation and temperature profiles in Lake St. Clair. Based on the simulated temperature, the vertical distribution (Figure 3-5) of water temperature consistently varied less than 1°C in Lake St Clair (~95% of the time). Measured profiles of temperature and water density were not available in Lake St. Clair

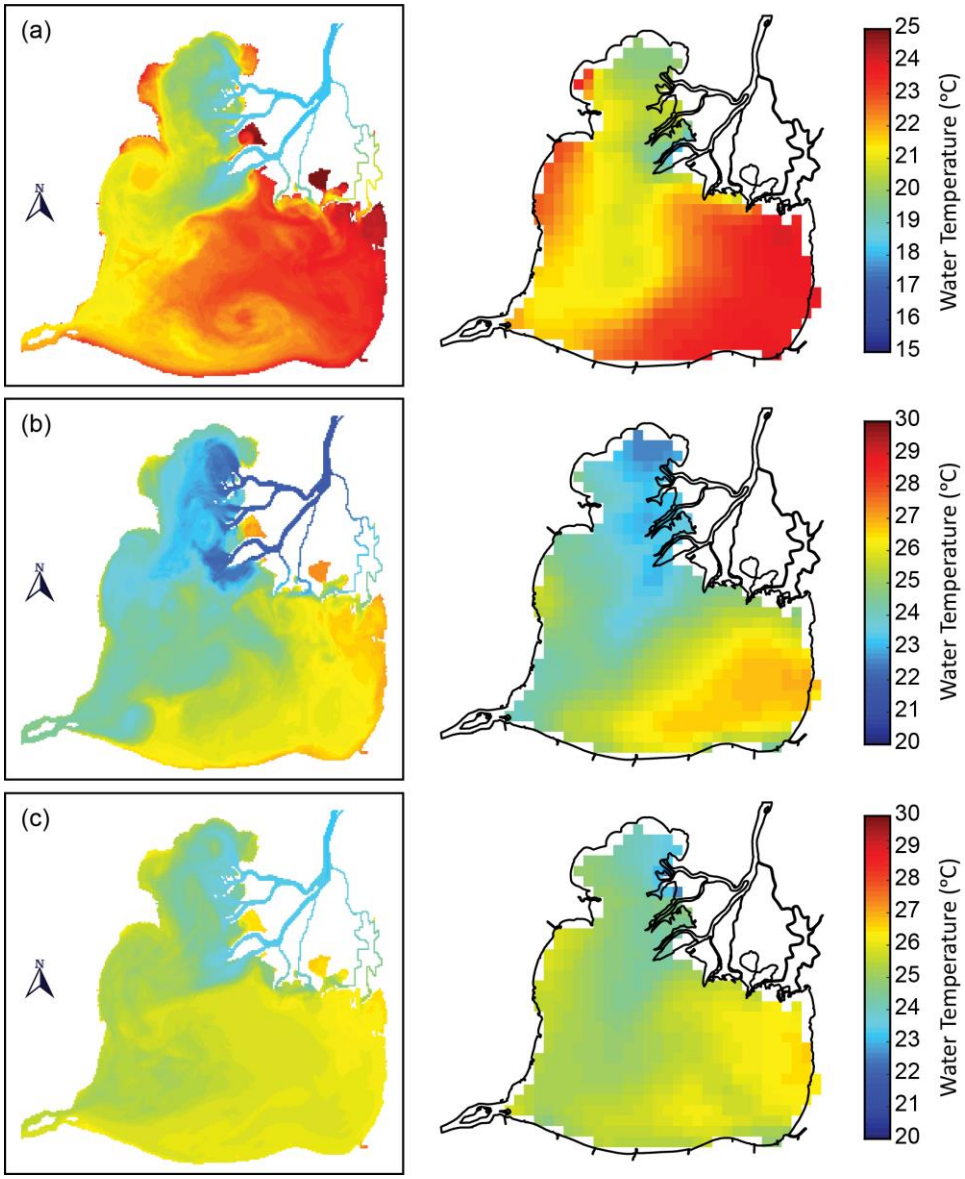

Figure 3-7: Comparison of model (left) vs. GLSEA data (right) on (a) June 16, (b) July 20 and (c) August 23, 2010. (Note each subgraph has a different temperature range (colorbar))

during the summer of 2010 for comparison, but a similar analysis to that by Loewen et al. (2007) and Ackerman et al., (2001), using densimetric Froude number (F_{rd}), was conducted (see Appendix 3-4). The results reveal that despite the small water temperature variation, weak stratification can be developed in parts of Lake St. Clair, but it will have a negligible influence on the local hydrodynamics, in particular, in the beach regions. These findings are consistent with those for the shallow western basin of Lake Erie, where stratification was shown to occur during summer months (Loewen et al. 2007; Boegman et al., 2008).

Three curtain cross section outputs are presented to show the vertical temperature profile in different zones of the lake. The curtains from northwest to southeast (Figure 3-5b) represent the lake-wide profile that clearly shows the difference between the top (U.S. side) and bottom (Canada side) water temperatures. Figure 3-6a shows simulated temperatures and Figure 3-6b shows the differences between Northwest and Southeast basins and Sandpoint Beach (see Figure 3-1). In Figure 3-6b, simulation results show Northwest water temperature is usually 1-3 °C cooler than the Southeast region, probably because of dominant cool water from the St. Clair River channels. The previous study period of McPhedran et al. (2013) is highlighted with a transparent red shade in Figure 3-6b. Within this period, the lake water temperature is uniform and the average temperature difference between Northwest and Southeast is around 1.5 °C. The maximum temperature difference between Northwest to Southeast is during June and early July when cool water still enters from the St. Clair River to the northwestern part of the lake. During August and September, water temperature in all the rivers is mostly similar (see Figure 3-2), which leads to less fluctuation in lake water temperature in the different zones.

Accuracy of GLSEA data has been evaluated in many previous studies. For example, Schwab et al. (1999) observed excellent agreement between in-situ water temperatures recorded at eight offshore buoys across the Great Lakes and GLSEA temperatures measured at the buoy locations (mean absolute error (MAE) were less than 0.5 °C for all buoys). Daily comparison of modelled surface temperature and GLSEA data for the entire simulation period shows very good agreement. WS ranged from 0.81 to 0.92, RMSE from 0.79 to 1.55 °C, and NRMSE from 0.32 to 0.63. Also, good agreement between present model results and GLSEA is evident for each of the three days selected in June, July, and August 2010, both by visual comparison (Figure 3-7) and by examining values of various statistical error indices. Good agreement between model results and nearshore water temperatures recorded by (McPhedran et al., 2013) can be observed in Figure 3-8 with RMSE of 1.27 °C, NRMSE of 0.11, and WS of 0.94.

3.4.1.3 Lake Hydrodynamics

The Thames, Sydenham, St. Clair, and Clinton rivers are major tributaries to Lake St. Clair. While St. Clair River dominates in terms of flow contribution (>98%), the other tributaries are expected to be a significant contributor to fecal source microbial pollution due to large human settlements in their watersheds. It is expected that the survival of this pollution and its spatial and temporal variation within the lake to be quite variable due to variations in water retention time, circulation and hydrodynamics. Hence, it is very important to have a good knowledge of the hydrodynamic processes that control water transport in this aquatic environment.

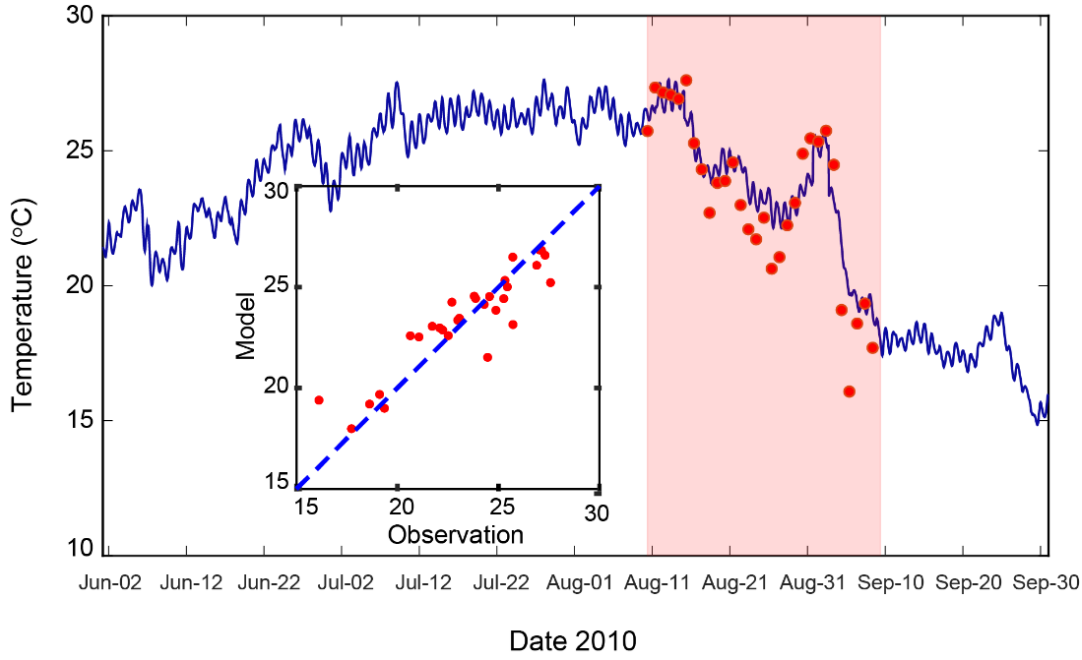

Figure 3-8: Temperature simulation (solid line) at Sandpoint Beach and field measurements (red dots (McPhedran et al., 2013)). Inner graph shows model vs. observation.

Figure 3-9a shows the average spatial distribution of water age in Lake St. Clair. The average retention time of the lake is approximately 9 days. The northwest area of the lake, dominated by the North and Middle channels, has the lowest retention time of less than 5 days. On the southeast side of the lake, near the Thames River, is a zone with very high water age (20-25 days). As well, the water age at the south shoreline is also high (11-28 days) as shown in the Sandpoint Beach time series of water age (Figure 3-9a sub-graph). These findings are in very good agreement with the simulated mean water age for period June 1 to October 1, 1985 by Anderson and Schwab (2011). Based on their results, water age along the shipping channel was found to be less than 5 days and along the southern boundary of Lake St. Clair was the highest mean age of around 25-30 days. The study of the water age inside the lake is very important because dispersion and decay rates are intrinsically related to the amount of time that species stay in the system before die off.

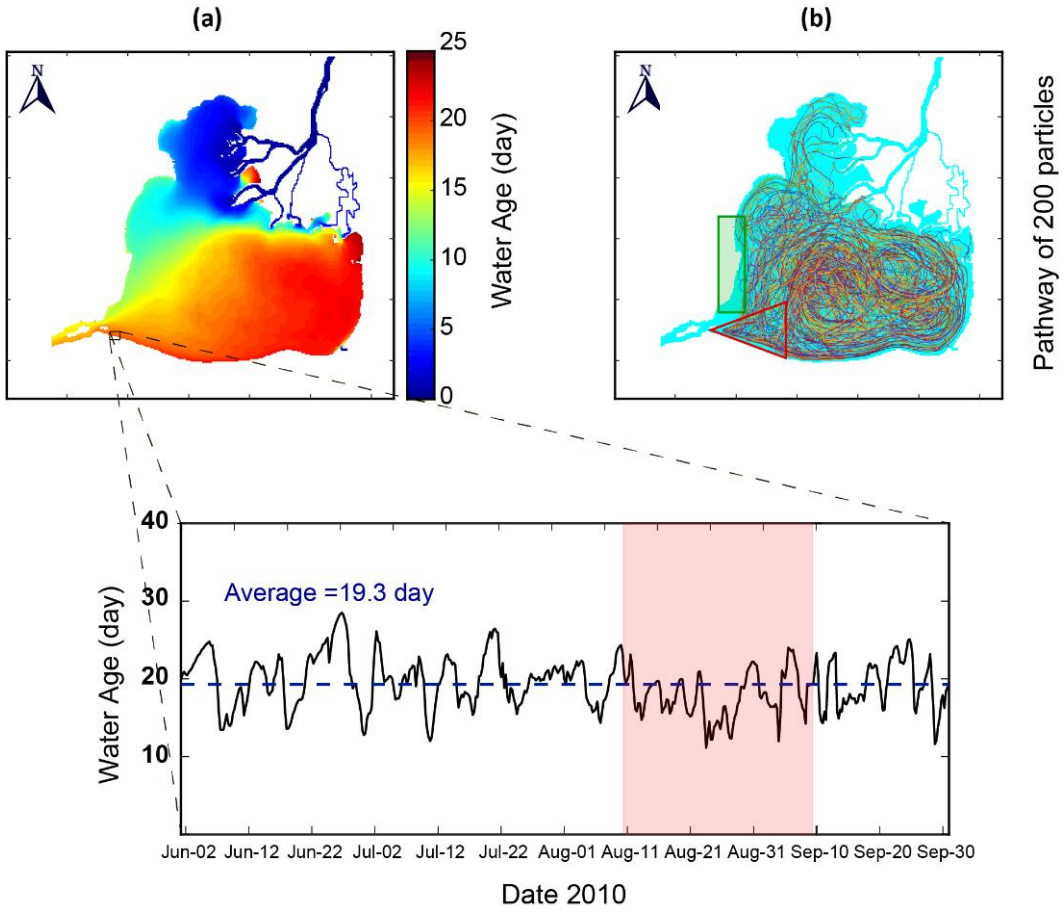

Figure 3-9: (a) Mean water age in Lake St. Clair (June to September 2010) and time series of water age that passed from the location of Sandpoint Beach (black line) and its average during the simulation period (19.3 days). (b) Backward particle tracking of 200 particles from Sandpoint Beach; rectangle shows the area on the US shore with few or no particles passing through; and the triangle shows the direction of particles.

Figure 3-9b shows the backward tracking of 200 particles released at Sandpoint Beach over the simulation period. This figure shows the general overview of the path of the water that reaches Sandpoint Beach. The three main observations from Figure 3-9b are i) particle paths show that all the water reaching Sandpoint Beach is in a narrow direction from east of the beach (range between 60 to 120 degrees from the North, indicated with a red triangle in Figure 3-9b), ii) The Cut-off, Flat and North channels of St. Clair River are the major inputs bringing water to Sandpoint Beach, and iii) the majority of the particles go along the

south shoreline (with high water age; see Figure 3-9a), taking less time to reach the beach because of fast flushing water which goes in a straight path with a high velocity along the shoreline (See Figure 3-10). The fact that all the water reaching Sandpoint Beach comes from the same direction is important because it shows that the nearshore regions on the U.S. side (indicated with the green box in Figure 3-9b) generally did not affect the water quality at Sandpoint Beach, so these areas are not of much concern.

Due to the lack of ADCP data during the 2010 simulation period, we tested the performance of the model with the same configuration during the summer of 2016, when ADCP data was available, together with initial conditions, forcing and inflow data for 2016 (provided

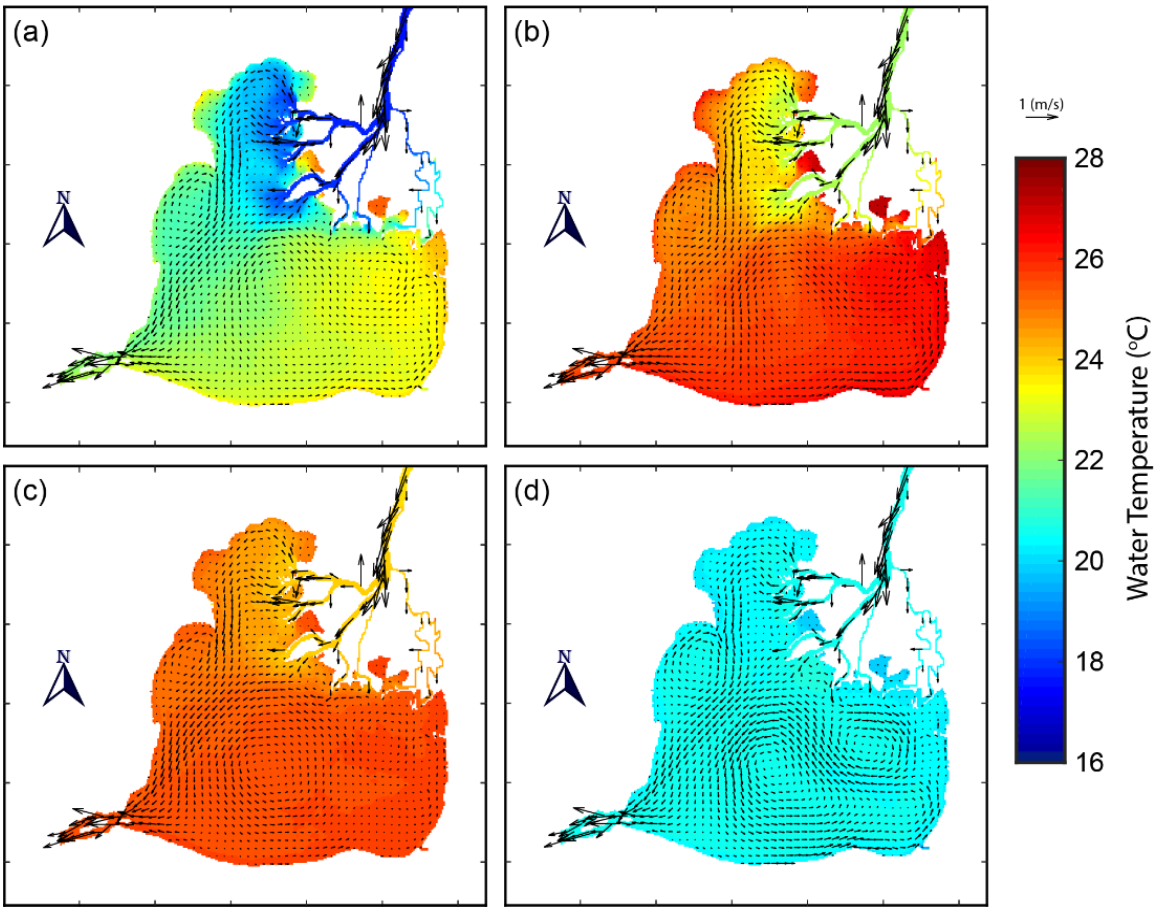

Figure 3-10: Monthly variation of the depth-averaged water circulation pattern (quiver plot) and water temperature (colormap) for (a) June, (b) July, (c) August and (d) September 2010.

by Environment and Climate Change Canada). We include these results and additional validation in Appendix 3-3. The simulated currents were reasonable and of the same order of magnitude as the ADCP data. The simulated west-east velocity component range was from -0.23 to 0.12 m/s (ADCP range from -0.18 - 0.10) and south-north component of the velocity range was from -0.12 to 0.10 m/s (ADCP range from -0.10 - 0.06). Overall, during 2016, the hydrodynamic model provided a reasonable output of currents with a low RMSE (and NRMSE) value of 0.055 m/s (0.276) and 0.049 m/s (0.449) for south-north and west-east components, respectively (see Figure 3-S2).

Simulation results for the monthly variation of depth-averaged circulation patterns and water temperature for the summer of 2010 presented in Figure 3-10 are consistent with known circulation patterns in Lake St. Clair (Anderson and Schwab, 2011; Anderson et al., 2010; Schwab et al., 1989). The dominant wind directions during our simulation period are south and southwest for all four months (See Figure 3-3). Overall, the simulated monthly averaged circulation pattern matched very well with the results of Anderson and Schwab (2011) and Schwab et al. (1989) in which the constant 10 m/s south wind and southwest wind directions were applied. Three features of the circulation pattern show high degree of similarity between the current study and those previous works including i) similar counter-clockwise gyre in the west part the lake, ii) flow direction along the east shoreline of the lake from the south to north direction, and iii) near the point where Lakeview Park West Beach is located (see Figure 3-1) the flow direction changed from east-west (in west part of the lake) to west-east (in east part of the lake).

Monthly variation of depth-averaged water temperature is expected because of variation in air temperature and changes in water temperature of tributaries especially that from St.

Clair River (see Figure 3-2). Except for the vicinity of Sandpoint Beach, the dominant along-shore velocity component for all the months is from west to east direction, which means that any watershed inputs from the south shore of the lake between Sandpoint Beach and Thames River tend to follow isobaths through the alongshore advective exchange with a much lower interaction with the main lake due to the smaller south-north (or nearshore-offshore) current component. This is due to the dominant W-SW wind direction in the region during summer (see Figure 3-3). Nearshore circulation is not significantly affected by the large-scale lake-wide circulation patterns except when gyres are formed from time to time. Thus, incomplete mixing especially in the nearshore regions may be expected leading to localized water-quality impairment, particularly downstream from tributary confluences. Near the mouth of the Thames River in the southeastern part of the lake, the depth-averaged south-north and west-east velocity components result in a counter-

clockwise circulation that pushes the Thames River inputs up and to the middle of the lake, allowing for a greater retention time and corresponding decay of microbial pollution. This circulation is also responsible for the longer average water age time in the southeastern part of the lake (Figure 3-9a).

3.4.2 Contribution of Major Tributaries to the Fecal Microbial Pollution of Lake St. Clair

Calculations of the fractional contribution to fecal microbial pollution of each tributary was the first part of the study analysis as mentioned in the model description. Presented in

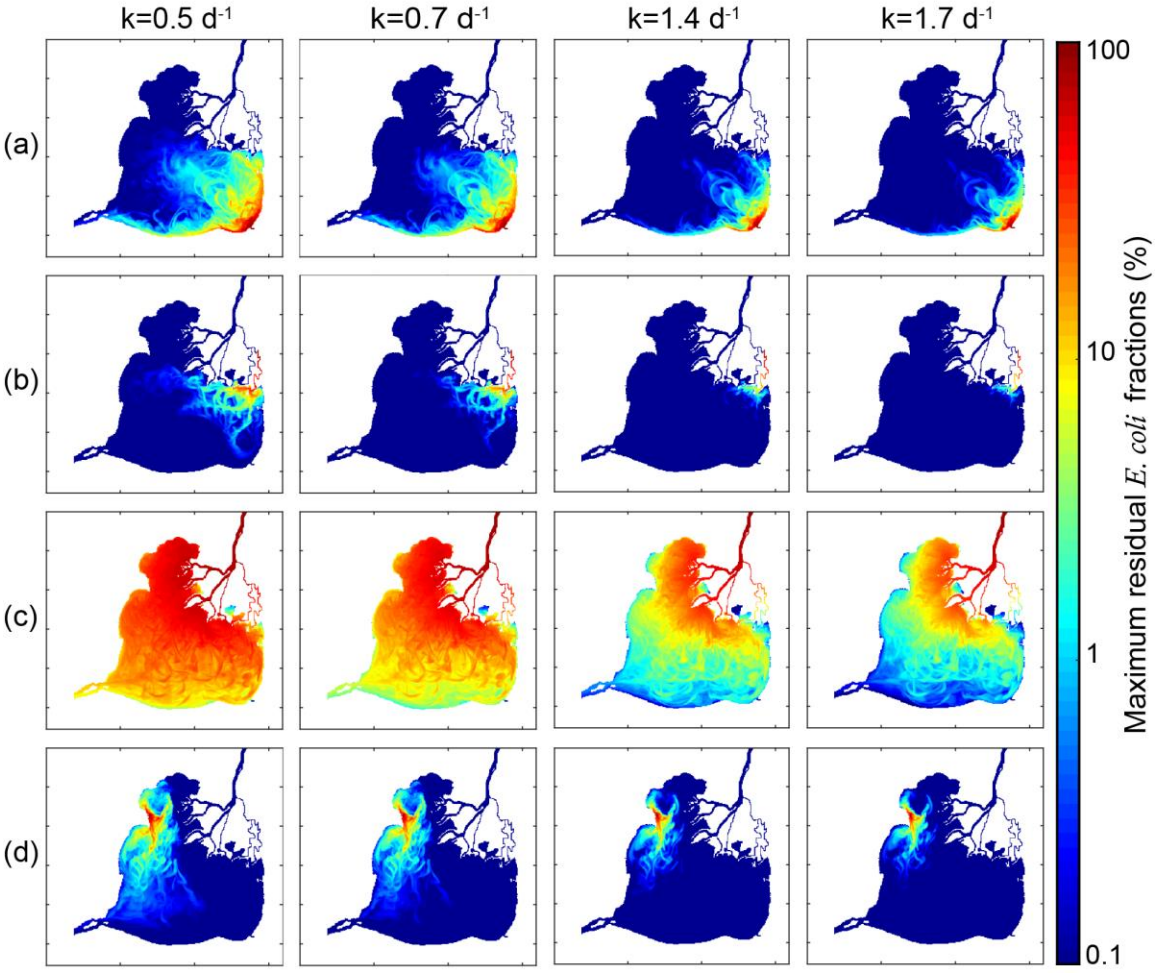

Figure 3-11: Spatial variations of maximum residual *E. coli* fractions at varying k values for (a) Thames River, (b) Sydenham River, (c) St. Clair River, and (d) Clinton River. Inlet *E. coli* concentration for each tributary was set to 100 CFU/100 ml.

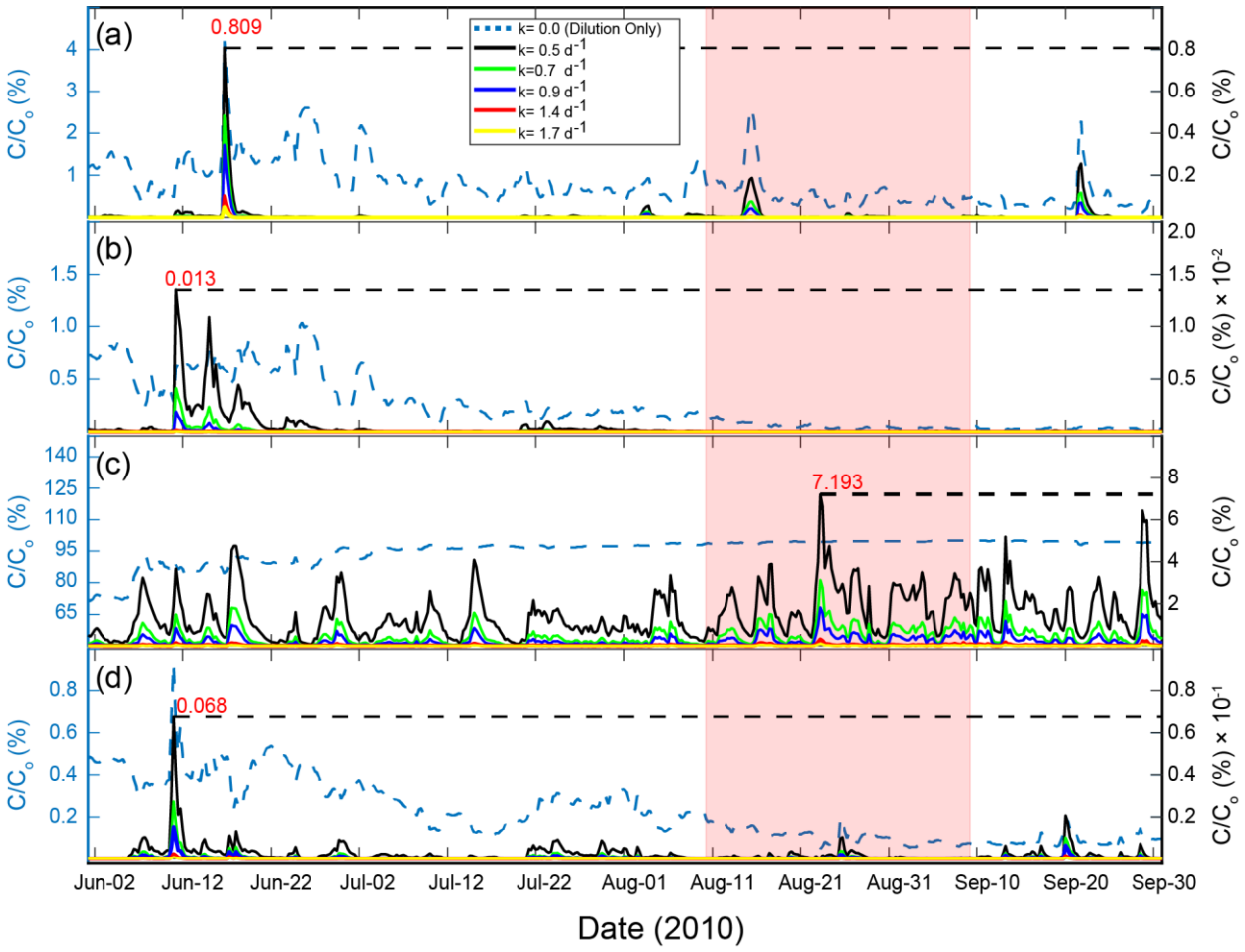

Figure 3-12: Temporal variation of *E. coli* residual fraction (right Y-axis) at Sandpoint Beach at varying k values for (a) Thames, (b) Sydenham, (c) St. Clair, and (d) Clinton rivers. Inlet *E. coli* concentration for each tributary was set to 100 CFU/100 ml. Left Y-axis shows *E. coli* residual fraction assuming k=0 (conservative) to quantify the contribution of flow dilution alone.

Figure 3-11 are the spatial variations of maximum residual *E. coli* fraction at Sandpoint Beach for varying k values for each tributary. Figure 3-12 (right Y-axis) shows the temporal variation of residual *E. coli* fraction for each tributary at varying decay rates. In Figure 3-12, the left Y-axis shows the expected temporal variation of residual *E. coli* fraction at k=0, which essentially quantifies the effect of flow dilution. Maximum fractions of *E. coli* that can reach the area near Sandpoint Beach are around 0.81, 0.01, 7.2 and 0.07 % of their inputs. From Figure 3-12, the minimum reduction factors due to dilution at

Sandpoint Beach of *E. coli* inputs from the Thames, Sydenham and Clinton River by the dominant flow from St. Clair River were estimated to be 24, 97 and 108, respectively.

The difference between model simulations at k=0 and any other k value then represents the contribution of decay alone to the reduction in *E. coli*. The minimum reduction factors in *E. coli* concentrations at Sandpoint Beach due to decay at the lowest decay coefficient of 0.5 d^{-1} for the Thames, Sydenham, St. Clair, and Clinton rivers were estimated to be 5.2, 76.8, 13.9, and 13.7, to give combined minimum total reduction factors of 125, 7450, 13.9, and 1480 for the same rivers. If higher decay rates are selected, then these factors would increase proportionately. This means that even if inlet *E. coli* concentrations were as high as 74,500 and 14,800 CFU/100 ml in the Sydenham and Clinton rivers, then the resulting maximum expected *E. coli* concentrations at Sandpoint Beach from each is expected to be 10 CFU/100 ml, which is 5% of the current Ontario safety guideline for recreational beach water use for swimming. This is consistent with Figure 3-11, which shows that even under conditions resulting in maximum *E. coli* concentrations, apart from a local region near their point of entry, the impact of fecal pollution from the Sydenham and Clinton rivers on microbial water quality in Lake St. Clair is quite limited, particularly in the region near the southern edge of Lake St. Clair where Sandpoint Beach is located. In addition, Figure 3-12 shows that the flow contribution from Thames River at Sandpoint Beach is higher than that from Sydenham or Clinton rivers, and ranged mostly between 1 to 4%. However, the residual fraction of *E. coli* concentrations was much lower, even at the lowest decay rate of 0.5 d^{-1}, with maximum, 99th percentile, and mean of 0.81%, <0.22%, and 0.01% respectively, corresponding to reduction factors of 125, >454, and 10,000.

This is due to the combination of dilution (due to much larger flow contribution from St. Clair River) and decay of *E. coli*, which is a function of water residence time. For the same flow contribution, if the residence time is longer, the resulting decay in *E. coli* concentrations is higher. In the same manner, for the St. Clair River, which dominated the flow contribution to Sandpoint Beach at >97%, the residual fraction of *E. coli* concentrations were again much lower with maximum, 99th percentile, and mean of 7.2%, <5.6%, and 1.5%, respectively, at the lowest decay rate of 0.5 d^{-1}. The corresponding reduction factors were 14, >18, and 67. The high and variable reduction factors are due to the high and variable water age. As shown in Figure 3-9a, the average water age in the southern and eastern parts of Lake St. Clair is >15 days and ~ 19 days at Sandpoint Beach. Such high water age results in longer periods that *E. coli* can remain in the environment, and which can decay to lower concentrations. Knowing that most of the water travels from the southern region of the lake with such high water age (Figure 3-9), results in low concentrations of microbial contamination. The St. Clair River is very important because of its dominant flow and total *E. coli* load inputs into the lake but high water age in the southern and eastern parts of the lake determines the microbial concentration at Sandpoint Beach.

For the second part of the contribution assessment, we obtained estimates of microbial pollution concentration for all tributaries by using a hydrograph-scaled loading approach to simulate their impact on microbial water quality of Lake St. Clair and Sandpoint Beach. In the case of the St. Clair River, in general, the overall microbial contamination is lower than that of many of the tributaries discharging to the lake. For instance, the annual reports for water quality of The Lambton Area Water Supply System (LAWSS) show that the

quality of water intake from St. Clair and Lake Huron was very good with the maximum detected *E. coli* concentration in raw water of 10 CFU/100 ml in 2010 and 5 CFU/100 ml in 2009. Based on *E. coli* data collected (from monitoring locations 21MICH_WQX-740402 and 21MICH_WQX-740404) from the National Water Quality Monitoring Council, the estimate of the *E. coli* concentration in 2010 summer months was obtained. Most of the values are very low with the range of *E. coli* between 1 CFU/100 ml to 134 CFU/100 ml.

Healy et al. (2007) presented the microbial characteristic of the Clinton River in much more detail. Data collected ranged from 10 to 10,000 CFU/100 ml with a median of 203 and 90th percentile of around 1,000 CFU/100 ml. To estimate the *E. coli* concentration of the Clinton River the hydrograph-scaled correlation based on 53 samples from 1989 to 2014 was obtained and used. The estimated values for the simulation period are in the range of 182 CFU/100 ml to 1100 CFU/100 ml.

There is no current *E. coli* data from the Lower Thames Valley Conservation Authority (LTVCA) stations but Thames River water quality at City of London is monitored regularly with values ranging from 110 to 3,300 CFU/100 ml in 2010 (ZEAS, 2010). Correlation between discharge flow and 138 samples of *E. coli* measurements at the mouth of the

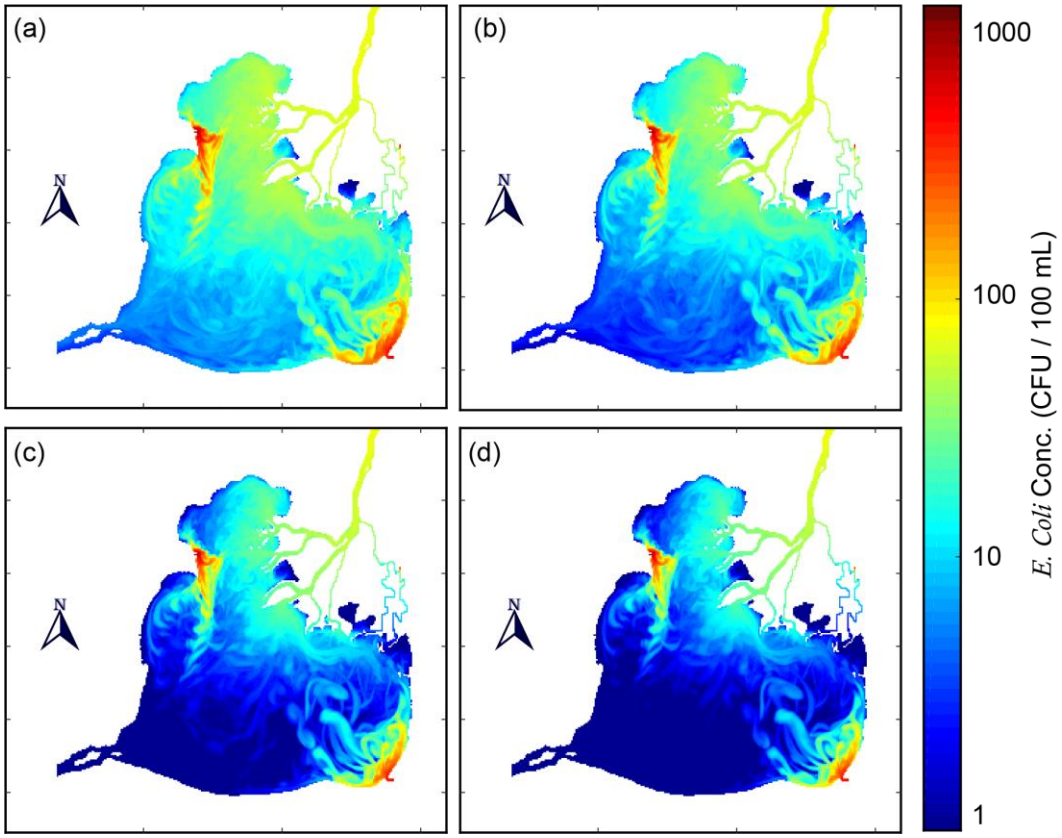

Figure 3-13: Spatial variation of maximum *E. coli* concentration at Lake St. Clair considering a scaled based hydrograph estimation of *E. coli* concentration for each tributary, for different decay rates (a) k=0.5 d^{-1}, (b) k=0.7 d^{-1}, (c) k=1.4 d^{-1} and (d) k=1.7 d^{-1}.

Thames River from 2016 to 2018 (unpublished results; Tom Edge 2018) was used to estimate the *E. coli* concentration for the simulation period.

In the case of the Sydenham River, a monitoring site in Wallaceburg is routinely tested for bacteria. The *E. coli* concentration at Wallaceburg is reported to be very high (during the summers of 2000 – 2008, 91% of the sampling had counts above the guidelines 100 CFU/100 ml). No detailed data was available for Sydenham River *E. coli* concentration, hence due to the proximity and similarity of the Thames River watershed with Sydenham River watershed, the same correlation for Thames River is applied to estimate its *E. coli*

concentration. The estimated values for the simulation period are in the range of 10 CFU/100 ml to 3,850 CFU/100 ml with mean *E. coli* concentration of 1,492 CFU/100 ml. Simulation results with first-order decay rates varying between 0.5 – 1.7 d^{-1} are illustrated in Figure 3-13. The results show that when using the estimated *E. coli* concentrations in the Thames, Sydenham, and Clinton rivers and the lowest decay rate of 0.5 d^{-1}, the

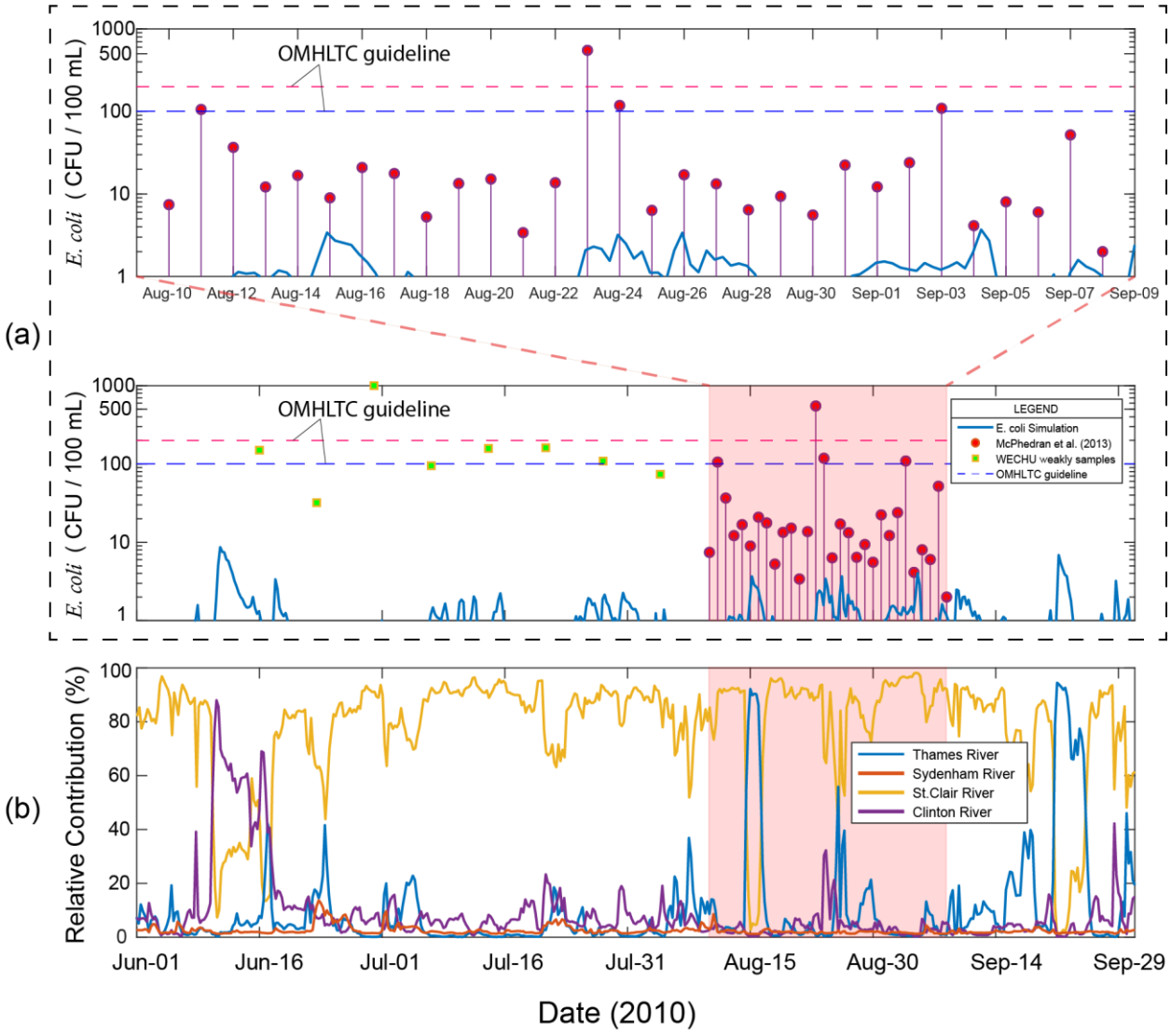

Figure 3-14: (a) Time series of simulated (solid line) and measured (red dots: and green squares WECHU weekly data) *E. coli* concentration (one-month simulation was zoomed for better visualization) (b) Relative contribution of each tributary to simulated *E. coli* concentrations at Sandpoint Beach (decay rate of k=0.5 d^{-1}).

maximum predicted *E. coli* concentrations in most of Lake St. Clair were <100 CFU/100 ml. A low *E. coli* zone in the southwestern part of the lake and near the location of Sandpoint Beach can be observed. In addition, distinct areas with maximum predicted *E. coli* concentrations >100 CFU/100 ml (orange and red color) can be observed in the eastern and northwestern part of the lake near the confluences of the Thames, Sydenham and Clinton rivers with Lake St. Clair, representing regions of greater impact or risk from microbial pollution from these tributaries.

Time series of simulated *E. coli* concentration at Sandpoint Beach for the Summer of 2010 with the lowest decay rate of 0.5 d^{-1} is presented in Figure 3-14a. The simulations are compared against one month of daily monitoring data (McPhedran et al., 2013) and weekly monitoring data collected by Windsor Essex County Health Unit (WECHU). The simulation shows the dynamics of predicted *E. coli* concentration with several peaks. However, even with the lowest decay rate of 0.5 d^{-1}, the maximum predicted *E. coli* concentration at Sandpoint Beach from the combined input of the major tributaries is less than 10 CFU/100 ml. The contributions of individual tributaries to *E. coli* concentrations at Sandpoint Beach are expected to be dependent on the input concentration, flow contribution and the time taken by that flow to reach Sandpoint Beach (water age). From Figure 3-14a and based on the results presented in Figure 3-12a, all peaks > 1 CFU/100 ml were dominated by the contributions of St. Clair River, except for three peaks that Thames River contributed. This can be explained by the fact that about 98% of the water comes to Sandpoint Beach from the St. Clair River with an average water age of more than 9 days (except for some occasions when the water age is 4-5 days). The peaks occur when unique weather conditions cause the water from St. Clair River to reach Sandpoint Beach faster

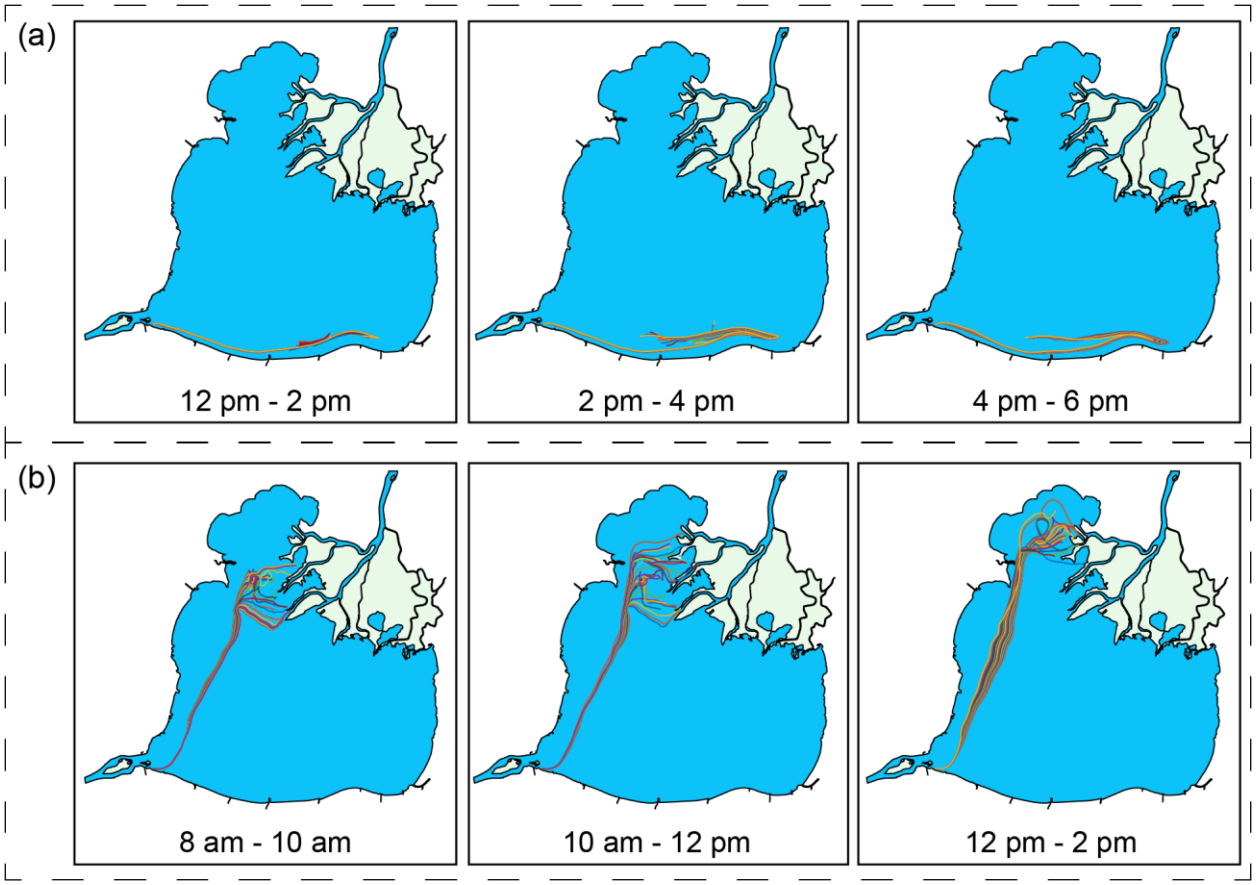

Figure 3-15: Backward particle tracking on two different days (a) June 16th (12pm-6pm) and (b) August 23rd (8am – 2pm); Each graph shows backward tracking of 24 particles released in 5 minute intervals during the mentioned hours.

than usual, as shown by the backward particle tracking for one such event on August 23rd in Figure 3-15b.

With an input *E. coli* concentration of 100 CFU/100 ml and a decay rate of 0.5 d^{-1}, when the water ages are 9 and 5 days, this will result in *E. coli* reductions of about 99% and 92%, respectively, which explains why the contributions of the St. Clair River to *E. coli* concentrations at Sandpoint Beach is reduced to about 7 CFU/100 ml or lower. For the Thames River, except for June 16th, Aug. 15th and Sept. 22nd, when the relative contribution of the Thames River is higher than 70% (See Figure 3-14b), the contributions to *E. coli* concentrations at Sandpoint Beach were calculated to be < 1 CFU/100 ml. This is due to a

combination of dilution by the much larger flow from the St. Clair River and decay resulting from the time taken for the flow to reach Sandpoint Beach. As shown with Figure 3-12, dilution alone can result in a minimum reduction factor of 24 (96% reduction) for *E. coli* concentration. On June 16th, Aug. 15th, and Sept. 22nd, unique weather conditions cause the water from the Thames River to more quickly reach Sandpoint Beach along the southern edge of Lake St. Clair, as shown in the example of backward particle tracking for June 16th in Figure 3-15a. The increased flow contribution and reduction in *E. coli* decay (due to lower travel time) resulted in a maximum predicted contribution of about 2 - 8 CFU/100 ml by the Thames River to the *E. coli* concentration at Sandpoint Beach.

We studied two days in more detail using backward particle tracking to find the reason behind the higher relative contribution of the Thames River on June 16th and the high absolute and relative contributions of the St. Clair River on August 23rd (Figure 3-15). On June 16th (Figure 3-15a), between noon and 6 pm, all the particles backtracked to the area near to the Thames River. The simulation time shows that it takes about 3.5-5 days for the water from Sandpoint Beach to backward track to the Thames River mouth. On the same day, water from the St. Clair River has very low concentrations that dilute the Thames River plume. On August 23rd, as can be seen from Figure 3-15b, between 8 am and 2 pm, all the particles backward track to the channels of the St. Clair River. This is why the absolute contribution of the St. Clair River for this day is very high at 7.2% (Figure 3-12c) and its relative contribution is more than 99% (Figure 3-14b).

It is interesting to note that the geographical locations of the St. Clair and Thames rivers with respect to Sandpoint Beach are such that their impact on *E. coli* concentrations are never coincidental. The relative contributions of the various tributaries to *E. coli*

contributions at Sandpoint Beach are shown in Figure 3-14b. When the contribution from the St. Clair River is high (e.g. August 23rd), the contribution of the Thames River is low at <1 CFU/100 ml. This is due to the high water age (> 20 days; Figure 3-9) of the St. Clair River water near the Thames River confluence with Lake St. Clair (eastern/southeastern part of Lake St. Clair) that allows for a greater decay in *E. coli* concentration. For the Sydenham and Clinton rivers, a combination of dilution by the much larger flow from the St. Clair River, and decay resulting from the time taken for the flow to reach Sandpoint Beach, resulted in maximum predicted contributions to *E. coli* concentrations at Sandpoint Beach to be less than 1 CFU/100 ml.

The simulation results presented in Figure 3-14a also show the predicted *E. coli* concentrations to be quite dynamic, with significant variations from one day to the next. This is in agreement with the observed daily monitoring data collected over a month during the simulation period by McPhedran et al. (2013) and presented in Figure 3-14a (inset). The rapid changes in total *E. coli* concentration (Figure 3-14) are caused by fast water movement (mean west-east velocity of about 0.3 m/s, see Figure 3-10) in the area near Sandpoint Beach. Flushing time describes the amount of time needed for the water in a specific area to be replaced by surrounding water (Andutta et al., 2013; Shore et al., 2016). Flushing times for the Sandpoint Beach over the Summer of 2010, calculated from model simulations, were found to be short and varied between a few hours to half a day. Thus the rapid changes in *E. coli* concentrations observed at Sandpoint Beach may be attributed to it's short flushing time by surrounding waters. In addition, water that passed through the vicinity of the Sandpoint Beach has an average age of 19.3 days (see Figure 3-9a). The quick flushing time and high water age suggest that (i) lake-wide contribution of *E. coli* to

the nearshore results in a reduction of *E. coli* concentration (dilution effects), and (ii) considering the decay rate of *E. coli*, with the high water age, the chance of survival for *E. coli* entering the lake through the rivers before reaching Sandpoint Beach is very low. With this high water age, the *E. coli* concentration decline more than a four-log reduction even when using the lowest decay rate of 0.5 d^{-1}. An important difference however is that while there were several exceedances from 100 CFU/100 ml and one exceedance of a beach closure value of 1000 CFU/100 ml in the measurement data from McPhedran et al. (2013) and weekly sampling data collected by WECHU, the predicted *E. coli* concentrations from the combined input of the major tributaries were only 8 CFU/100 ml or less.

The maximum predicted *E. coli* concentrations over a large area surrounding Sandpoint Beach was <40 CFU/100 ml when using the lowest decay rate of 0.5 d^{-1}. This clearly indicates that the microbial pollution coming to Lake St. Clair from the four major tributaries (St. Clair, Clinton, Thames, and Sydenham) have an insignificant impact on the *E. coli* concentrations at Sandpoint Beach, and are not responsible for exceedances of Ontario recreational water use *E. coli* guidelines observed at that beach. The one-month monitoring data also has several *E. coli* measurements at Sandpoint Beach that are similar to or within a factor of 5 of the maximum predicted *E. coli* concentration of about 8 CFU/100 ml (Figure 3-14a inset), which can be attributed to flushing by surrounding waters. This suggests that the waters surrounding Sandpoint Beach are generally of good quality with *E. coli* concentrations of < 40 CFU/100 ml, which is consistent with the model predictions. Further, this implies that the *E. coli* exceedances observed at Sandpoint Beach may be coming from more local sources.

There are several much smaller streams in the Essex Windsor Region that enter Lake St. Clair near Sandpoint Beach, and the contributions of these streams, either in terms of flows or microbial pollution, to the larger Lake are expected to be negligible. However, they could still have some influence in a local region surrounding their confluence with Lake St. Clair. Storm events usually coincide with the high wave and wind speed that subsequently may result in sediment resuspension and contribute to higher *E. coli* concentrations. Ge et al. (2012) showed that deposition-resuspension cycles are responsible for excessive bacterial contamination of beach water. In addition, *E. coli* can naturalize and grow in the environment such as foreshore beach sand (Staley et al., 2018) and sediment, which can add to the complexity of the processes and consequently lead to false positive results in beach water microbial quality (Chan et al., 2015; Halliday and Gast, 2011; Ishii et al., 2006; Whitman et al., 2014).

Higher *E. coli* observed in the nearshore area can be explained to some extent by characteristics of nearshore waters, and needs more investigation. In particular, more turbid water, which can be caused by the lower depth and higher water current resulting in higher shear stress that leads to sediment resuspension, results in lower sunlight penetration in the water column and hence lower deactivation rates. As might be expected, Lake St. Clair showed very low *E. coli* mostly in the central and eastern basins, which contribute to the nearshore region concentration by diluting and reduction of *E. coli* concentration.

In order to define the potential risk area for microbial contamination and also define the safe zone in the lake, areas with a concentration equal to or higher than 1% of the input concentration are presented in Figure 3-16a for all tributaries. Except for the St. Clair River (showed by the blue color that covers most of the lake), other tributaries MPRA are very

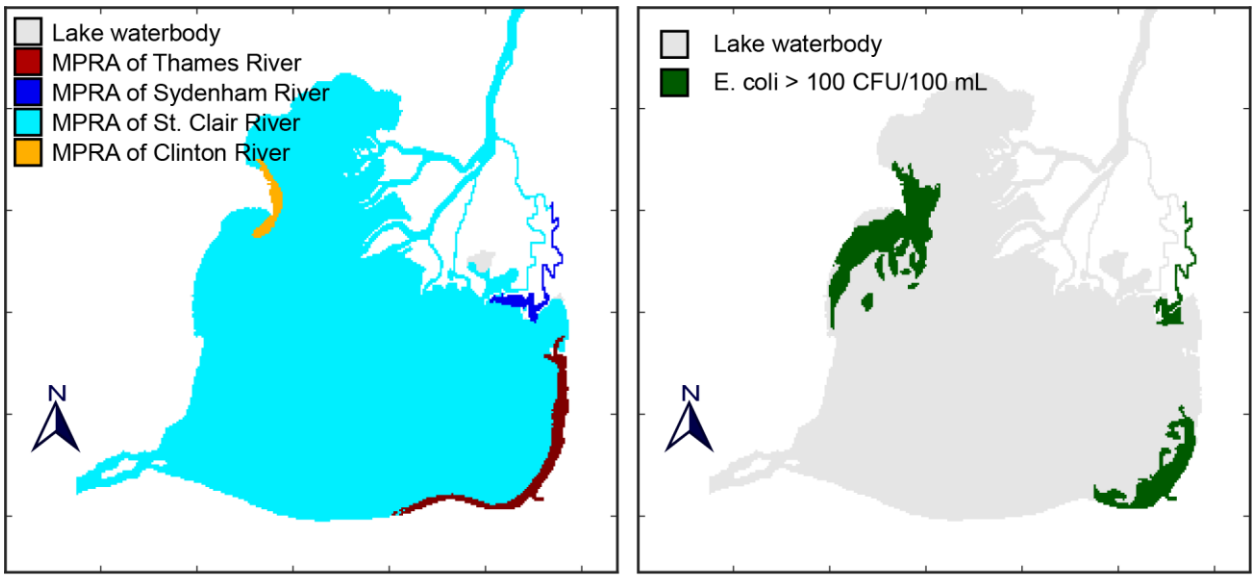

Figure 3-16: (a) Microbiological potential risk area for major tributaries of Lake St. Clair, (b) Potential area that *E. coli* concentration is equal to or greater than 100 CFU/100 ml based on maximum concentration.

small and rarely exceed 5km from the river mouth and shoreline. The MPRA of the St. Clair River covers most of the Lake and its *E. coli* concentration is very low and never exceeds 100 CFU/100 ml (see Figure 3-16b). As can be seen from Figure 3-16b, high *E. coli* concentrations in each river are very quickly diluted by mixing with the very low *E. coli* concentration in the lake water, so theoretically low concentrations will reach the area adjacent to Sandpoint Beach. In addition to the dilution effect, *E. coli* concentration decreases due to biological decay during the time it spends in the water. The longer the period that water stays in the lake before reaching the shoreline of the beach, the lower concentration of *E. coli* will be due to decay.

3.5 Summary and Conclusion

A three-dimensional high-resolution model based on the Aquatic Ecosystem Model (AEM3D) modelling platform was successfully set up for Lake St. Clair, and showed

reasonable agreement with previous studies for its circulation pattern, and its comparison with observed surface water temperature for the summer of 2010 (June to September) and ADCP data (summer 2016). The model can correctly simulate water temperature temporal fluctuations (comparing with mooring data) and spatial fluctuations (comparing with GLSEA data) during the modelling period. In addition to water temperature, flow through the delta channels was also simulated, and found to be in good agreement with previous work. Inputs from the St. Clair, Clinton, Thames and Sydenham rivers account for >99% of all flows into Lake St. Clair. The Thames and Clinton rivers are dominant amongst contributors from urbanized watersheds. Thus the four tributaries are expected to be the dominant contributors to the inputs and budget of all fecal microbial pollution in Lake St. Clair.

The AEM3D Model was used to study the temporospatial distribution of fecal microbial pollution represented by *E. coli* as the fecal indicator bacteria, in Lake St. Clair and Sandpoint Beach on the Canadian side. Estimates of fecal pollution were obtained for each of the tributaries from literature and assumed constant for the entire simulation period. We modelled the fate and transport of *E. coli* assuming it to be freely suspended using a lumped parameter first-order decay function for a range of values obtained from the literature.

Model simulations show that the fate of fecal microbial pollution arriving in Lake St. Clair from the four major tributaries is largely controlled by the flow and microbial concentration in St. Clair River, and the death or decay of that pollution due to the time spent (water age) in the Lake. In addition, *E. coli* simulations with hydrograph-scaled loading based estimation in all the major tributaries (See Figure 3-13) reveals that the influence of the Thames, Sydenham, and Clinton rivers on causing exceedance based on area or volume is

quite limited even at the lowest decay factor of 0.5 d^{-1}. This is due to a combination of dilution by the St. Clair River water as well as decay due to water age.

For coastal regions (including beaches), the contributions of smaller flows and fecal microbial pollution sources can be more significant from time to time than lake-wide contributions if there is an opportunity for the pollution to travel along the coastal edge. The extent of such influences depends on a combination of factors including coastal proximity and weather conditions. Similar to Lake St. Clair, the microbial water quality at Sandpoint Beach was found to be dominantly affected by waters from the St. Clair River that travel along the southern edge. Sandpoint Beach is located in an area of high water age and is least influenced by fecal microbial pollution from the four major tributaries.

The maximum predicted *E. coli* concentration from the combined input of the tributaries is less than 10 CFU/100 ml at the lowest constant decay rate of 0.5 d^{-1}, which is expected to be a conservative estimate. Most of the predictions were dominated by the contribution of the St. Clair River, except on three occasions when the contributions were dominated by the Thames River under weather conditions that drove the flow more quickly from Thames River to Sandpoint Beach along the coastal edge. However, the dilution due to St. Clair River and decay due to travel time still allowed the predicted *E. coli* concentration to be <10 CFU/100 ml. The results clearly indicate that the four major tributaries are not responsible for the *E. coli* exceedances of the Ontario safety guidelines for recreational use at Sandpoint Beach. The source or sources responsible for the exceedances observed need further investigation.

The current study is part of an effort to develop mathematical models that are expected to be useful, complementary and cost-effective additions to monitoring efforts in microbial

source tracking, to help with better beach management, reduce economic loss due to beach closures, and reduce risk to human health of beachgoers from recreational use of beach waters. Despite the assumptions made in the modelling effort, it was useful in establishing that the dominant part of fecal microbial pollution brought into Lake St. Clair, by the four major tributaries, may be expected to have a limited or insignificant effect on the microbial water quality in large parts of Lake St. Clair, including Sandpoint Beach.

To help answer other questions, a different or more detailed modelling effort may be needed. For example, in the current study, *E. coli* is assumed to be freely suspended with its death or decay represented by a lumped first-order decay coefficient, which was assumed to be constant over time. However, *E. coli* decay or inactivation can also be affected by other factors such as UV exposure from sunlight, grazing and die-off due to water temperature and other chemical properties of water (pH, dissolved oxygen etc.) (Brookes et al., 2004), and the effect is expected to be variable over time under the conditions being modelled.

In cases where it may be important, the AEM3D modelling framework has the capabilities that can allow for a time-dependent inactivation rate based on temperature, sedimentation, solar insolation and water chemical properties- provided data is available to parameterize the decay model. Further, it is known that sediment-associated *E. coli* can contribute to increased concentrations in the water column in areas of high shear stress during significant storm events. It has also been shown that a significant part of *E. coli* load in riverine inputs to lakes during storm events could be associated with sediment particles that get mobilized due to large flow velocities during such events (Droppo et al., 2011). However, except for some simplistic settling-resuspension conditions, the AEM3D modelling framework

currently does not have the capability to simulate sediment (cohesive and non-cohesive) associated *E. coli*. Further model development would be required to simulate these more complex fate and transport processes of *E. coli*, which is beyond the scope of the current study.

3.6 Acknowledgments

The authors are grateful to the Ontario Ministry of the Environment and Climate Change (MOECC) for the ADCP data and also to Windsor Essex County Health Unit (WECHU) for providing the weekly *E. coli* data. We thank Environment and Climate Change Canada (ECCC) and Canada Centre for Inland Waters for their support and special thanks to Dr. Tom Edge and his team for providing *E. coli* data for the Thames River. This research received financial support from Canada's Natural Sciences and Engineering Research Council (NSERC) Strategic Project Grant (SPG) and was performed while Mr. Madani (first author) held an Ontario Trillium Scholarship (OTS).

3.7 Supplementary Material

Supplementary data to this article can be found in Appendix 3.

3.8 References

Ackerman, J.D., Loewen, M.R., Hamblin, P.F., 2001. Benthic-Pelagic coupling over a zebra mussel reef in western Lake Erie. Limnology and Oceanography 46(4) 892-904.

Acosta, M., Anguita, M., Fernández-Baldomero, F.J., Ramón, C.L., Schladow, S.G., Rueda, F.J., 2015. Evaluation of a nested-grid implementation for 3D finite-difference semi-implicit hydrodynamic models. Environmental Modelling & Software 64 241-262.

Anderson, E.J., Schwab, D.J., 2011. Relationships between wind-driven and hydraulic flow in Lake St. Clair and the St. Clair River Delta. Journal of Great Lakes Research 37(1) 147-158.

Anderson, E.J., Schwab, D.J., Lang, G.A., 2010. Real-Time Hydraulic and Hydrodynamic Model of the St. Clair River, Lake St. Clair, Detroit River System. Journal of Hydraulic Engineering 136(8) 507-518.

Andutta, F.P., Ridd, P.V., Wolanski, E., 2013. The age and the flushing time of the Great Barrier Reef waters. Continental Shelf Research 53 11-19.

Ansa, E.D.O., Lubberding, H.J., Ampofo, J.A., Gijzen, H.J., 2011. The role of algae in the removal of *Escherichia coli* in a tropical eutrophic lake. Ecological Engineering 37(2) 317-324.

Auer, M.T., Niehaus, S.L., 1993. Modeling fecal coliform bacteria—I. Field and laboratory determination of loss kinetics. Water Research 27(4) 693-701.

Baustian, M.M., Mavrommati, G., Dreelin, E.A., Esselman, P., Schultze, S.R., Qian, L., Aw, T.G., Luo, L., Rose, J.B., 2014. A one hundred year review of the socioeconomic and ecological systems of Lake St. Clair, North America. Journal of Great Lakes Research 40(1) 15-26.

Bocaniov, S.A., Leon, L.F., Rao, Y.R., Schwab, D.J., Scavia, D., 2016. Simulating the effect of nutrient reduction on hypoxia in a large lake (Lake Erie, USA-Canada) with a three-dimensional lake model. Journal of Great Lakes Research 42(6) 1228-1240.

Bocaniov, S.A., Scavia, D., 2018. Nutrient Loss Rates in Relation to Transport Time Scales in a Large Shallow Lake (Lake St. Clair, USA-Canada): Insights From a Three-Dimensional Model. Water Resources Research 54(6) 3825-3840.

Bocaniov, S.A., Ullmann, C., Rinke, K., Lamb, K.G., Boehrer, B., 2014. Internal waves and mixing in a stratified reservoir: Insights from three-dimensional modeling. Limnologica - Ecology and Management of Inland Waters 49 52-67.

Boegman, L., Loewen, M.R., Hamblin, P.F., Culver, D.A., 2008. Vertical mixing and weak stratification over zebra mussel colonies in western Lake Erie. Limnology and Oceanography 53(3) 1093-1110.

Bolsenga, S.J., Herdendorf, C.E., 1993. Lake Erie and Lake St. Clair Handbook. Wayne State University Press.

Bonamano, S., Madonia, A., Borsellino, C., Stefani, C., Caruso, G., De Pasquale, F., Piermattei, V., Zappala, G., Marcelli, M., 2015. Modeling the dispersion of viable and total *Escherichia coli* cells in the artificial semi-enclosed bathing area of Santa Marinella (Latium, Italy). Mar Pollut Bull 95(1) 141-154.

Bonamano, S., Piermattei, V., Madonia, A., Paladini de Mendoza, F., Pierattini, A., Martellucci, R., Stefanì, C., Zappalà, G., Caruso, G., Marcelli, M., 2016. The Civitavecchia Coastal Environment Monitoring System (C-CEMS): a new tool to analyze the conflicts between coastal pressures and sensitivity areas. Ocean Sci. 12(1) 87-100.

Bravo, H.R., McLellan, S.L., Klump, J.V., Hamidi, S.A., Talarczyk, D., 2017. Modeling the fecal coliform footprint in a Lake Michigan urban coastal area. Environmental Modelling & Software 95 401-419.

Brookes, J.D., Antenucci, J., Hipsey, M., Burch, M.D., Ashbolt, N.J., Ferguson, C., 2004. Fate and transport of pathogens in lakes and reservoirs. Environment International 30(5) 741-759.

Chan, S.N., Thoe, W., Lee, J.H., 2013. Real-time forecasting of Hong Kong beach water quality by 3D deterministic model. Water Res 47(4) 1631-1647.

Chan, Y.M., Thoe, W., Lee, J.H.W., 2015. Field and laboratory studies of *Escherichia coli* decay rate in subtropical coastal water. Journal of Hydro-environment Research 9(1) 1-14.

Chen, W.-B., Liu, W.-C., 2017. Investigating the fate and transport of fecal coliform contamination in a tidal estuarine system using a three-dimensional model. Marine Pollution Bulletin 116(1) 365-384.

Cheng, R.T., Casulli, V., Gartner, J.W., 1993. Tidal, Residual, Intertidal Mudflat (TRIM) Model and its Applications to San Francisco Bay, California. Estuarine, Coastal and Shelf Science 36(3) 235-280.

Crane, S.R., Moore, J.A., 1986. Modeling enteric bacterial die-off: A review. Water, Air, and Soil Pollution 27(3) 411-439.

de Brauwere, A., Ouattara, N.K., Servais, P., 2014. Modeling fecal indicator bacteria concentrations in natural surface waters: a review. Critical Reviews in Environmental Science and Technology 3389(March 2015) 140225123512007-140225123512007.

Droppo, I.G., Krishnappan, B.G., Liss, S.N., Marvin, C., Biberhofer, J., 2011. Modelling sediment-microbial dynamics in the South Nation River, Ontario, Canada: Towards the prediction of aquatic and human health risk. Water Research 45(12) 3797-3809.

Emerson, D.G., Vecchia, A.V., Dahl, A.L., 2005. Evaluation of drainage-area ratio method used to estimate streamflow for the Red River of the North Basin, North Dakota and Minnesota, Scientific Investigations Report, Online only ed.

Fewtrell, L., Kay, D., 2015. Recreational Water and Infection: A Review of Recent Findings. Curr Environ Health Rep 2(1) 85-94.

Field, K.G., Samadpour, M., 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Research 41(16) 3517-3538.

Ge, Z., Whitman, R.L., Nevers, M.B., Phanikumar, M.S., Byappanahalli, M.N., 2012. Nearshore hydrodynamics as loading and forcing factors for *Escherichia coli* contamination at an embayed beach. Limnology and Oceanography 57(1) 362-381.

Halliday, E., Gast, R.J., 2011. Bacteria in beach sands: an emerging challenge in protecting coastal water quality and bather health. Environ Sci Technol 45(2) 370-379.

Hamidi, S.A., Bravo, H.R., Val Klump, J., Waples, J.T., 2015. The role of circulation and heat fluxes in the formation of stratification leading to hypoxia in Green Bay, Lake Michigan. Journal of Great Lakes Research 41(4) 1024-1036.

Hannoun, I., List, E.J., Kavanagh, K.B., Chiang, W.-L., Ding, L., Preston, A., Karafa, D., Rackley, I., 2006. Use of Elcom and Caedym for Water Quality Simulation in Boulder Basin. Proceedings of the Water Environment Federation 2006(9) 3943-3970.

He, C., Post, Y., Dony, J., Edge, T., Patel, M., Rochfort, Q., 2016. A physical descriptive model for predicting bacteria level variation at a dynamic beach. Journal of Water and Health 14(4) 617-629.

Health Canada, 2012. Guidelines for Canadian Recreational Water Quality, Third Edition ed. Minister of Health.

Healy, D.F., Chambers, D.B., Rachol, C.M., Jodoin, R.S., 2007. Water quality of the St. Clair River, Lake St. Clair, and their U.S. tributaries, 1946-2005, Scientific Investigations Report: Reston, VA.

Hodges, B.R., 2000. Numerical Techniques in CWR-ELCOM (code release v. 1). CWR Manuscript WP 1422.

Hodges, B.R., Dallimore, C., 2006. Estuary, Lake and Coastal Ocean Model.

Hodges, B.R., Imberger, J., Saggio, A., Winters, K.B., 2000. Modeling basin-scale internal waves in a stratified lake. Limnology and Oceanography 45(7) 1603-1620.

Holtschlag, D.J., Koschik, J.A., 2001. Water-resources Investigations Report: 2001-2004. U.S. Department of the Interior, U.S. Geological Survey.

Holtschlag, D.J., Koschik, J.A., 2002a. Flow Distribution in Selected Branches of St. Clair and Detroit Rivers. Journal of Great Lakes Research 28(3) 379-395.

Holtschlag, D.J., Koschik, J.A., 2002b. A two-dimensional hydrodynamic model of the St. Clair-Detroit River waterway in the Great Lakes basin, Water-Resources Investigations Report: Lansing, MI.

Holtschlag, D.J., Shively, D., Whitman, R.L., Haack, S.K., Fogarty, L.R., 2008. Environmental factors and flow paths related to *Escherichia coli* concentrations at two beaches on Lake St. Clair, Michigan, 2002–2005, Scientific Investigations Report: Reston, VA.

Huziy, O., Sushama, L., 2017. Lake–river and lake–atmosphere interactions in a changing climate over Northeast Canada. Climate Dynamics 48(9) 3227-3246.

Ishii, S., Ksoll, W.B., Hicks, R.E., Sadowsky, M.J., 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. Appl Environ Microbiol 72(1) 612-621.

Leon, L.F., Antenucci, J., Rao, Y., McCrimmon, C., 2012a. Summary performance of the Estuary and Lake Computer Model (ELCOM): application in the Laurentian and other Great Lakes.

Leon, L.F., Smith, R.E.H., Hipsey, M.R., Bocaniov, S.A., Higgins, S.N., Hecky, R.E., Antenucci, J.P., Imberger, J.A., Guildford, S.J., 2011. Application of a 3D hydrodynamic–biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie. Journal of Great Lakes Research 37(1) 41-53.

Leon, L.F., Smith, R.E.H., Malkin, S.Y., Depew, D., Hipsey, M.R., Antenucci, J.P., Higgins, S.N., Hecky, R.E., Rao, R.Y., 2012b. Nested 3D modeling of the spatial dynamics of nutrients and phytoplankton in a Lake Ontario nearshore zone. Journal of Great Lakes Research 38 171-183.

Liang, L., Goh, S.G., Gin, K.Y.H., 2017. Decay kinetics of microbial source tracking (MST) markers and human adenovirus under the effects of sunlight and salinity. Science of The Total Environment 574 165-175.

Liu, L., Phanikumar, M.S., Molloy, S.L., Whitman, R.L., Shively, D.A., Nevers, M.B., Schwab, D.J., Rose, J.B., 2006. Modeling the transport and inactivation of *E. coli* and enterococci in the near-shore region of Lake Michigan. Environ Sci Technol 40(16) 5022-5028.

Loewen, M.R., Ackerman, J.D., Hamblin, P.F., 2007. Environmental implications of stratification and turbulent mixing in a shallow lake basin. Canadian Journal of Fisheries and Aquatic Sciences 64(1) 43-57.

McPhedran, K., Seth, R., Bejankiwar, R., 2013. Occurrence and predictive correlations of *Escherichia coli* and Enterococci at Sandpoint beach (Lake St. Clair), Windsor, Ontario and Holiday beach (Lake Erie), Amherstburg, Ontario. Water Quality Research Journal of Canada 48(1) 99-99.

OMHLTC, O.M.o.H.a.L.-T.C., 2018. Recreational water protocol. Queen's Printer for Ontario: Toronto, ON.

Oveisy, A., Rao, Y.R., Leon, L.F., Bocaniov, S.A., 2014. Three-dimensional winter modeling and the effects of ice cover on hydrodynamics, thermal structure and water quality in Lake Erie. Journal of Great Lakes Research 40 19-28.

Rao, Y.R., Sheng, J., 2008. Application of a Nested-grid hydrodynamic model for circulation and thermal structure in the coastal boundary layer of Lake Huron. Aquatic Ecosystem Health & Management 11(2) 161-166.

Schwab, D.J., Clites, A.H., Murthy, C.R., Sandall, J.E., Meadows, L.A., Meadows, G.A., 1989. The effect of wind on transport and circulation in Lake St. Clair. Journal of Geophysical Research: Oceans 94(C4) 4947-4958.

Schwab, D.J., Leshkevich, G.A., Muhr, G.C., 1999. Automated Mapping of Surface Water Temperature in the Great Lakes. Journal of Great Lakes Research 25(3) 468-481.

Shore, J.A., Valipour, R., Blukcacz-Richards, E.A., 2016. Twenty-eight years of hydrodynamic variability in the Bay of Quinte (ice-free periods of 1979–2006). Journal of Great Lakes Research 42(5) 985-996.

Simpson, J.M., Santo Domingo, J.W., Reasoner, D.J., 2002. Microbial source tracking: state of the science. Environmental Science & Technology 36(24) 5279-5288.

Sokolova, E., 2011. Hydrodynamic modelling of microbial water quality in drinking water sources. 46-46.

Sokolova, E., Astrom, J., Pettersson, T.J., Bergstedt, O., Hermansson, M., 2012a. Decay of Bacteroidales Genetic Markers in Relation to Traditional Fecal Indicators for Water Quality Modeling of Drinking Water Sources. Environmental Science & Technology 46(2) 892-900.

Sokolova, E., Astrom, J., Pettersson, T.J., Bergstedt, O., Hermansson, M., 2012b. Estimation of pathogen concentrations in a drinking water source using hydrodynamic modelling and microbial source tracking. J Water Health 10(3) 358-370.

Spalding, D., 2014. Expert can't identify cause of spike in beach closures, 30 Sep 2014 ed. Windsor Star: Windsor.

Staley, Z.R., He, D.D., Shum, P., Vender, R., Edge, T.A., 2018. Foreshore beach sand as a reservoir and source of total phosphorus in Lake Ontario. Aquatic Ecosystem Health & Management 21(3) 268-275.

Tan, B., Ng, C., Nshimyimana, J.P., Loh, L.L., Gin, K.Y.H., Thompson, J.R., 2015. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. Frontiers in microbiology 6 1027-1027.

Toms, I.P., Owens, M., Hall, J.A., Mindenhall, M.J., 1975. Observations on the performance of polishing lagoons at a large regional works sewage treatment . v. 74.

Trolle, D., Spigel, B., Hamilton, D.P., Norton, N., Sutherland, D., Plew, D., Allan, M.G., 2014. Application of a Three-Dimensional Water Quality Model as a Decision Support Tool for the Management of Land-Use Changes in the Catchment of an Oligotrophic Lake. Environmental Management 54(3) 479-493.

USEPA, 2010. in Water by Membrane Filtration Using membrane-Thermotolerant *Escherichia coli* Agar (mTEC). (March).

USEPA, 2012. Recreational Water Quality Criteria, In: 820-F-12-058, E. (Ed.): Office of Water, Washington, DC.

Valipour, R., León, L.F., Depew, D., Dove, A., Rao, Y.R., 2016. High-resolution modeling for development of nearshore ecosystem objectives in eastern Lake Erie. Journal of Great Lakes Research 42(6) 1241-1251.

Valipour, R., Rao, Y.R., León, L.F., Depew, D., 2019. Nearshore-offshore exchanges in multi-basin coastal waters: Observations and three-dimensional modeling in Lake Erie. Journal of Great Lakes Research.

Whitman, R., Harwood, V.J., Edge, T.A., Nevers, M., Byappanahalli, M., Vijayavel, K., Brandao, J., Sadowsky, M.J., Alm, E.W., Crowe, A., Ferguson, D., Ge, Z., Halliday, E., Kinzelman, J., Kleinheinz, G., Przybyla-Kelly, K., Staley, C., Staley, Z., Solo-

Gabriele, H.M., 2014. Microbes in Beach Sands: Integrating Environment, Ecology and Public Health. Rev Environ Sci Biotechnol 13(3) 329-368.

Whitman, R.L., Nevers, M.B., Korinek, G.C., Byappanahalli, M.N., 2004. Solar and Temporal Effects on *Escherichia coli* Concentration at a Lake Michigan Swimming Beach. Applied and Environmental Microbiology 70(7) 4276-4285.

ZEAS, 2010. Water quality monitoring report for the Thames river, final report. The Corporation of the City of London, Ontario.

Zhao, J., Rao, Y.R., Wassenaar, L.I., 2012. Numerical modeling of hydrodynamics and tracer dispersion during ice-free period in Lake Winnipeg. Journal of Great Lakes Research 38 147-157.

CHAPTER 4 MODELLING THE FATE AND TRANSPORT OF MICROBIAL CONTAMINATION IN LAKE ST. CLAIR: THE EFFECT OF DECAY DYNAMICS

4.1 Introduction

Lake St. Clair is part of the channel connecting Lake Huron and Lake Erie. This channel serves as a recreational waterway, a source of drinking water for Detroit and surrounding cities, as well as a shipping channel to Lakes Huron, Michigan, and Superior. Being relatively shallow and having a large urban area and discharges from many tributaries along the edge on the southern part of the lake, Lake St. Clair is especially susceptible to microbial contamination. Lake St. Clair water quality is a significant concern for thousands of people in Canada and the United States that rely on it for drinking, fishery, and recreational purpose (Hamelin et al., 2007). Drinking water sources are commonly impacted by both human and animal fecal contamination from the point and non-point sources and, as a result, may contain pathogenic microorganisms, including bacteria, viruses, Cryptosporidium oocysts, and Giardia cysts. Detection and enumeration of fecal coliforms, *E. coli*, and Enterococcus are a typical method to measure the existence of fecal contamination and also pathogenic organisms in surface waters. Higher levels of fecal contamination of drinking water sources increase the potential of pathogenic microorganisms surviving the drinking water treatment process and endangering public health. In this regard, raw water is routinely measured for its water quality (e.g., bacteriology, turbidity, color, natural organic matter (NOM), and conductivity).

A comprehensive evaluation of potential sources of fecal contamination and their impact on drinking water sources in Lake St. Clair is currently lacking. The prediction of microbial contamination levels is required to provide data relative to the impact and timing of sources

of fecal pollution which affect the drinking water source. To be able to prevent waterborne disease outbreaks caused by focally contaminated drinking water, an accurate assessment of the contribution from different sources to the total fecal contamination at the raw water intake of a drinking water treatment plant is needed (Sokolova et al., 2013). Many studies tried to use three-dimensional modelling to propose the alternative locations for the current water intakes based on the circulation and spreading patterns of the incoming flows in the lake (Elmoustafa, 2017; Na and Park, 2005). Microbial water quality modelling can allow for better management of fecal contamination by helping to prioritize mitigation measures based on the assessment of the contribution of different contamination sources to the fecal contamination at the raw water intakes Sokolova et al. (2012b). In such condition, the modelling approach can be used to provide more clear pictures of the water quality at the location of water intakes.

Many studies used *E. coli* to model the microbial distribution in the lakes (Bonamano et al., 2015; Chan et al., 2013; Liu et al., 2006; Sokolova et al., 2012b). In a simplified approach, *E. coli* fate and transport is modelled using a constant decay rate based on average conditions (Garcia-Armisen et al., 2006; Lin and Falconer, 2001). Such an approach was applied recently for a preliminary assessment of contributions of the major tributaries on the microbial water quality in Lake St. Clair and at a popular Beach (Sandpoint Beach) in Windsor, Ontario, Canada (Madani et al., 2020). Different formulations based on the different environmental variables such as water temperature, turbidity, pH, dissolved oxygen, salinity, solar radiation, settling and resuspension, grazing, predation and amount of nutrients of growth and mortality rate of enteric organisms were also proposed (Brookes et al., 2004; Fiandrino et al., 2003; Hipsey et al.,

2008; Hipsey et al., 2004; Liu et al., 2006). Many studies have shown that key environmental factors such as changing water temperature over the simulation period and solar insolation can dynamically influence *E. coli* decay rates and resulting concentrations (de Brauwere et al., 2014; Liang et al., 2017; Liu et al., 2006).

Located in the northern hemisphere at a latitude of 42.46^{o} N, Lake St. Clair experiences significant changes in water temperature and solar insolation from one month to another. Lake St. Clair is a source of drinking water to large populations in the United States and Canada. Grosse, Mount Clemens, New Baltimore and Ira Township water treatment plants (WTPs) take water from Lake St. Clair to supply drinking water to residents in the US, while Lakeshore and Stoney Point WTPs supply drinking water to residents in Canada. Thus temporal changes in *E. coli* decay and concentrations due to changes in water temperature and diurnal variation in solar insolation could be important. In the current study, a three-dimensional high-resolution hydrodynamic model based on the Aquatic Ecosystem Model (AEM3D) modelling platform for Lake St. Clair is applied to model the influence of the four major tributaries (St. Clair, Thames, Clinton, and Sydenham rivers) on the microbial water quality (*E. coli* concentration) at the drinking water intakes of all the WTPs. The fate of *E. coli* was modelled using two approaches, i) variable decay rate based on water temperature and solar radiation (Approach 1) and ii) constant decay rate based on average conditions, and the results obtained are compared and discussed.

4.2 Material and Methods

4.2.1 Study Area

Lake St. Clair is part of the Great Lakes region and split between the Canadian province of Ontario and the state of Michigan in the United States (US). Its surface area is about 1114

km^2. About 59% (8988 km^2) of the areal extent of watersheds that drain into Lake St. Clair is on the Canadian side, with the remainder (6317 km^2) on the US side (Baustian et al., 2014). The bathymetry and location of major rivers (inlet and outlet), Lake St. Clair buoys, and location of water treatment intakes are shown in Figure 4-1. The St. Clair River is the main inlet of Lake St. Clair and delivers about 98.2% (5,200 m^3/s) of the total inflow (Holtschlag et al., 2008). Around 1% of the remaining inflow to the lake is contributed by the next three largest tributaries: Thames River and Sydenham River in Ontario, Canada and Clinton River in Michigan, US.

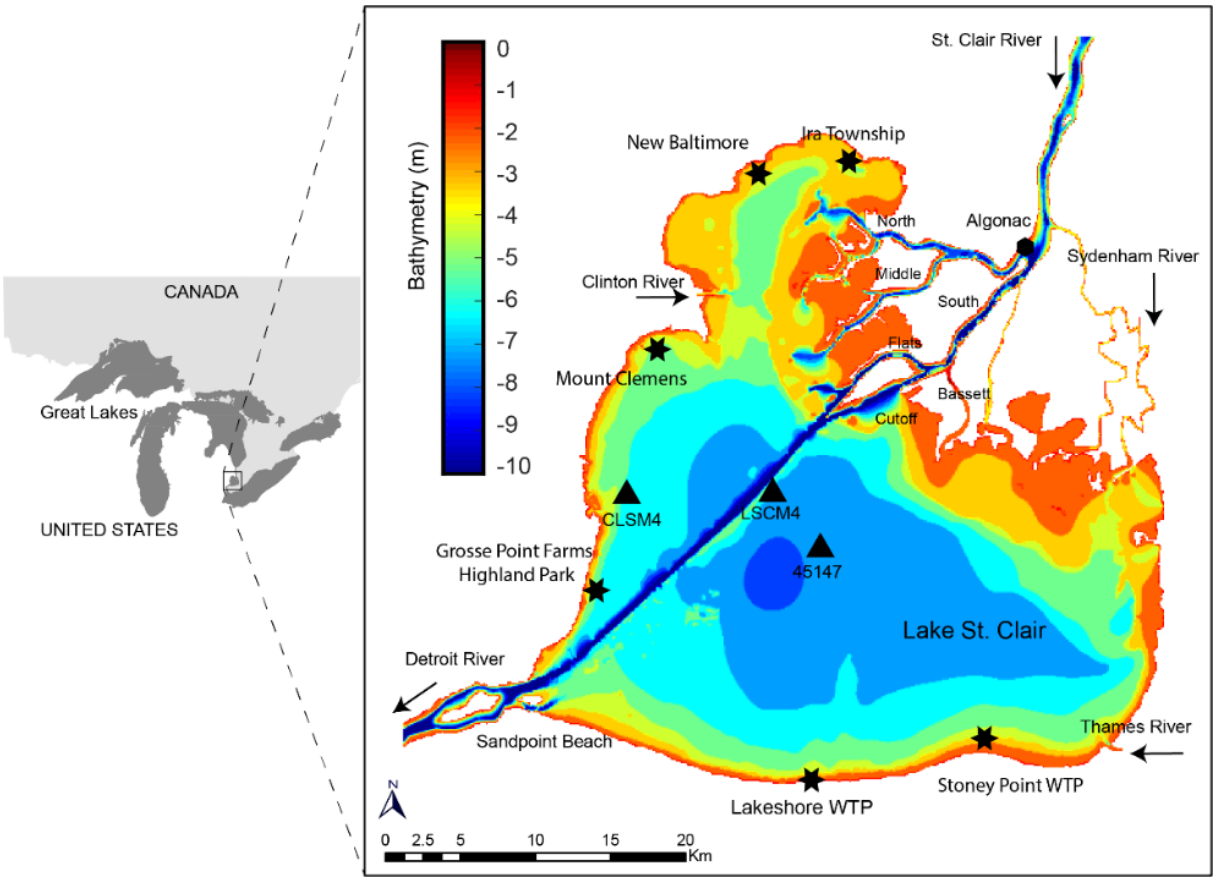

Figure 4-1: Map of Lake St. Clair and its bathymetry. Triangle symbols show the locations of the moorings CLSM4 (42.471 N 82.877 W), LSCM4 (42.465 N 82.755 W) and 45147 (42.430 N 82.680 W)). Locations of water treatment plants intakes are identified by black stars

Lakeshore (capacity = 36400 m^3/day) and Stoney Point (capacity = 4545 m^3/day) are two water treatment plants that supply water to the town of Lakeshore, Ontario on the southern side of Lake St. Clair. Other water treatment plants located on the US side include Grosse Point Farms Highland Park (Grosse), Mount Clemens (Clemens), New Baltimore (Baltimore) and Ira Township (Ira).

4.2.2 Bathymetry and Forcing Data

Lake St. Clair bathymetry, obtained from National Geophysical Data Center (https://www.ngdc.noaa.gov/), has depths referenced to a generic datum of 176.784 meters (580 feet). The original rectangular grid (cell sizes of ~70 m by ~90 m) was processed to generate coarser 400 m uniform grids. Meteorological forcing data sets include wind direction, wind speed, air temperature, humidity, atmospheric pressure, rainfall, and solar radiation. Data is from the nearest stations (Windsor A. Station ID=4716; Windsor Riverside Station ID=4715) and moored surface buoys (Buoy 45147 (42.430N 82.680W); LSCM4 (42.465N 82.755W); CLSM4 (42.471N 82.877W)). Buoy 45147, which is maintained by Environment and Climate Change Canada (ECCC), was used to fill out missing wind direction and wind speed data and for model validation of water temperature. Solar radiation data were obtained from the Canadian Regional Climate Model (CRCM5) provided by the University of Quebec at Montreal (UQAM) (Huziy and Sushama, 2017). It was assumed that the entire study area receives the same amount of solar radiation.

4.2.3 Inflows and Boundary Conditions

Flow and temperature data for the four major tributaries to the lake (Thames, Sydenham, St. Clair, and Clinton rivers) and the outflow (Detroit River) were obtained from the nearest gauged stations. The data on the *E. coli* load from the four major tributaries were estimated

using measured data and a hydrograph-scaled loading approach. Detail information about flow, water temperature and *E. coli* data and methods used for data processing are presented in our previous paper (Madani et al., 2020).

4.2.4 Model Description

In the current study, the existing AEM3D (extended version of ELCOM-CAEDYM) modelling framework for Lake St. Clair (Madani et al., 2020) was used over five months (May-September) of 2010. The first month is considered as warm-up period and results are not shown here. The AEM3D framework has been well documented (Hodges and Dallimore, 2006) and previously applied for hydrodynamic modelling of various small and large lakes and reservoirs (Imberger et al., 2017; Leon et al., 2005; Madani et al., 2020; Mosley et al., 2015; Romero et al., 2004; Tranmer et al., 2018; Trolle et al., 2012; Valipour et al., 2016; Valipour et al., 2018). More information regarding hydrodynamic model preparation, boundary conditions for Lake St. Clair can be found in a recent study by the authors (Madani et al., 2020). The hydrodynamic driver of AEM3D can be coupled with Computational Aquatic Ecosystem Dynamics Model (CAEDYM) for investigations involving biological, microbial and chemical processes. The pathogen module of CAEDYM allows for incorporating the effects of temperature, salinity, pH, dissolved oxygen, sunlight, nutrients and turbidity on the growth and mortality of microorganisms (Hipsey et al., 2008).

4.2.5 Analysis Approach for *E. coli*

Fate and transport of *E. coli* released from major tributaries of Lake St. Clair were simulated using the developed hydrodynamic and water quality model, taking the decay of the *E. coli* into account. *E. coli* concentrations were simulated using the first-order decay

rate based on two approaches: i) variable decay rate using the pathogen module of CAEDYM and ii) constant decay rate. The governing equation that explains the transport of the *E. coli* has the following general form:

$$\frac{dC}{dt} + \frac{\partial}{\partial x_j}(CU_j) = \frac{\partial}{\partial x_j}\left(\kappa_j \frac{\partial C}{\partial x_j}\right) + C_{in} - C_{out} + KC \quad (1)$$

Where C denotes the *E. coli* concentration ($orgs/m^3$), t is time, x_j is the distance in the j^{th} dimension (m), U_j is the velocity in the j^{th} dimension (m/s^1), κ_j is eddy-diffusivity, C_{in} and C_{out} are the inflow and outflow fluxes ($orgs/(m^3 s)$) and K is the overall decay rate.

In the first approach (Approach 1), the dynamic fate of *E. coli* was modelled using the pathogen module of CAEDYM, using the empirical formulation proposed by Hipsey et al. (2008), which is based on experimental results from several studies. Within the formulation, the time-dependent decay rate was modelled as a function of water temperature and sunlight intensity with different bandwidth. It was assumed that the solar radiation incident is encompassing visible-light 45%, UV-A 3.5%, UV_B 0.5% and rest are Near-Infrared bandwidths. Due to a lack of available data, the effect of pH, salinity, dissolved oxygen and nutrients was not considered. The overall decay rate K in Eq. 1 in the first approach was expressed as:

$$K = K_T + K_l + K_p - K_g \quad (2)$$

where K_T is the natural mortality or die-off rate due to water temperature and can be expressed as the Arrhenius expression:

$$K_T = k_{d20}\, \theta^{T-20} \quad (3)$$

where T is the temperature ($°C$), k_{d20} is the dark death rate at 20 $°C$ in freshwater and θ controls the sensitivity of K_T to water temperature change.

In Eq. 2, K_l is total die-off due to exposure to sunlight with different bandwidth. It takes the form:

$$K_l = \sum_{b=1}^{N_B} \varphi \, k_b \, f_b \, I_0 \left(\frac{1- e^{-\eta_b \Delta z}}{e^{-\eta_b \Delta z}}\right) \quad (4)$$

where N_B is the number of discrete solar bandwidths to be modelled which here is 3 (visible, UV_A and UV-B), b is the bandwidth class [1, 2, .. ., N_B], k_b is the freshwater inactivation rate coefficient for exposure to the b^{th} class ($m^2\ MJ^{-1}$), φ is a constant to convert units from seconds to days and J to MJ (=$8.64 \times\ 10^{-2}$). In Eq. 4, Δz is the depth of the computational cell and η_b is the extinction coefficient for each bandwidth region which governs how incident light is attenuated within the water column according to the Beer Law. Details about the other parameters and range of parameters are presented by Hipsey et al. (2008).

K_p is a simple temperature-dependent inactivation of enteric organisms due to predation and grazing:

$$K_p = k_{p20} \, \theta_p^{T-20} \quad (5)$$

where k_{p20} is the minimum rate due to predation at 20 $°C$ and θ_p accounts for the sensitivity of predation to water temperature. Finally K_g is the growth rate:

$$K_g = \mu_{max} \left[c_{T1}(T - T_{min})\left(1 - \exp\left(c_{T2}(T - T_{max})\right)\right)\right]^2 \quad (6)$$

where μ_{max} is the maximum growth rate at 20 $°C$ and c_{T1}, c_{T2} , T_{min} and T_{max} are species-specific constants controlling the exact shape of the growth function.

Using Eq. 2, solar radiation data, and the available simulated water temperature, the decay rate for each computational cell and the whole simulation period was calculated. In the second approach (Approach 2), the net reduction in *E. coli* concentration was simulated as a tracer with a constant decay rate. The constant decay rate (k=0.9 d^{-1}) in Approach 2 was obtained by averaging the decay rate using Approach 1 over all the cells and the simulation period.

4.3 Results and Discussion

4.3.1 Lake-wide Hydrodynamics

The model predicted hydrodynamics in Lake St. Clair was in good agreement with previous studies for lake circulation pattern, as was shown in a recent study by the authors (Madani et al., 2020). The study also showed a high degree of agreement between the observed and model-predicted temporal and spatial distribution of temperature in the lake.

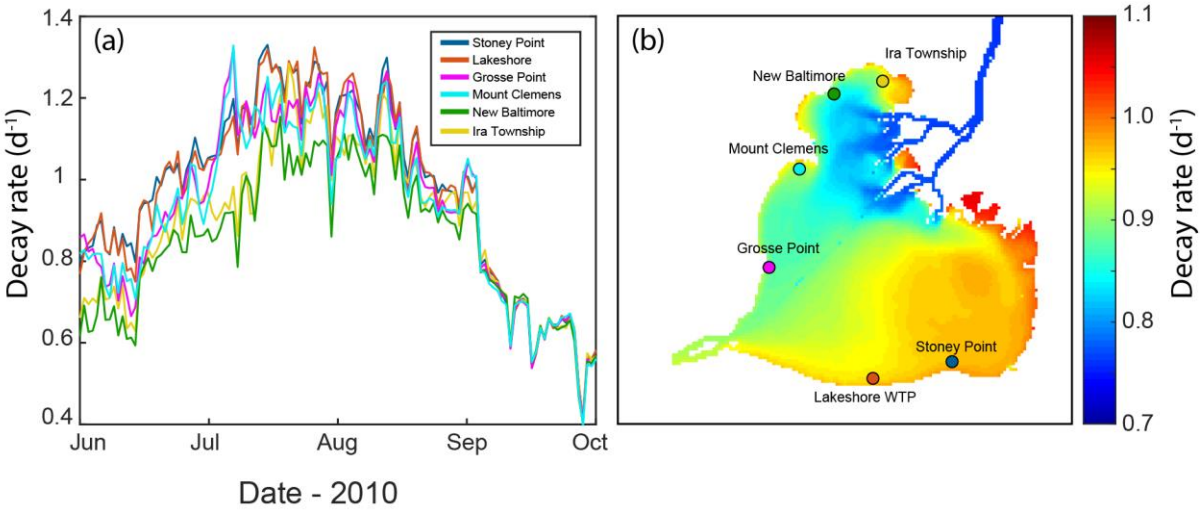

Figure 4-2: a) Comparison of the overall decay rate at six locations of water treatment plants in Lake St. Clair and b) mean overall decay rate at Lake St. Clair during the simulation period. Circles show the location of the water treatment plants.

4.3.2 Temporal and Spatial Variability in *E. coli* Decay

As decay rate components are either a function of water temperature or solar insolation, they are changing during the simulation time and in different locations of the lake. Figure 4-2a shows the temporal variation of the total decay rate at six locations of water treatment intakes indicated in Figure 4-2b. Total decay rates in June and September are lower than the average decay rate ($k=0.9\ day^{-1}$). Although the trend for all the locations is similar, decay rate at Stoney Point water treatment plant (SP_WTP) and Lakeshore water treatment plant (LS_WTP) and most of the southern shoreline is similar to the decay rate in the eastern region of the lake but is statistically different from the decay rate at the northwest part of the lake where New Baltimore and Ira Township water intakes are located. The decay rates at locations of southern and northwest regions have more differences during the July month. Figure 4-2b presents the spatial variation of the mean total decay rate over the simulation period. *E. coli* survival rates are dependent on temperature (Blaustein et al., 2013). As the solar radiation for the entire study domain was assumed to be uniform, the spatial variation and rapid changes in decay rates can be explained by the changes that were

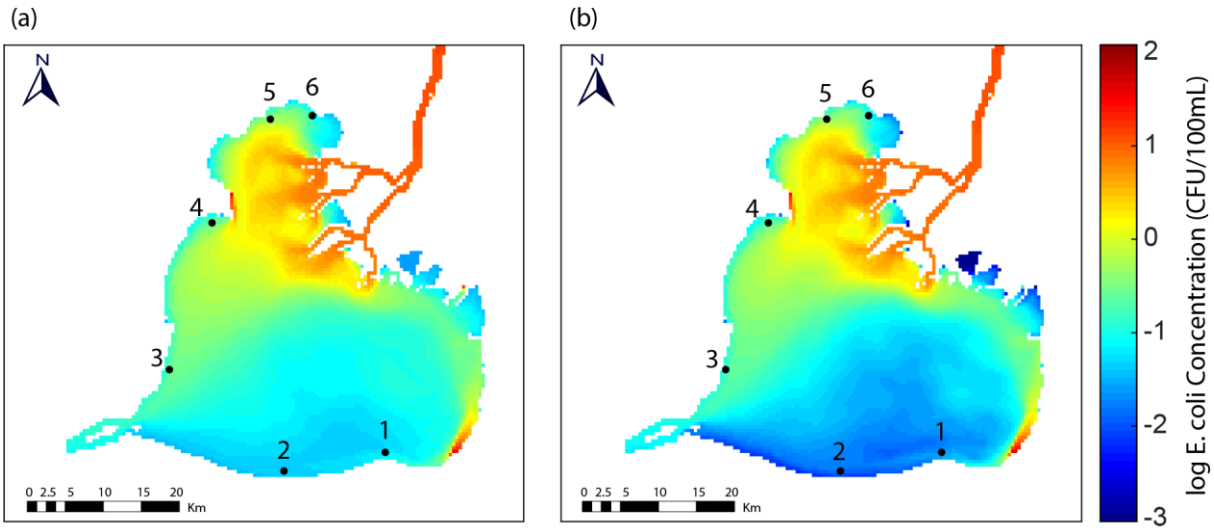

Figure 4-3: Mean *E. coli* concentration of Lake St. Clair for the simulation period (June - September 2010) using a) formulated decay rate b) constant decay rate k= 0.9 d-1. Numbers show the location of the water treatment plants intakes as illustrated in Figure 4-1.

observed in water temperature (Madani et al., 2020). However, water temperature is strongly influenced by temporal variations in incident solar radiation.

4.3.3 Lake-wide Microbial Water Quality

The comparison of mean simulated *E. coli* concentration using the two approaches over the entire period (June- September 2010) is presented in Figure 4-3. Spatially both approaches show a similar pattern in mean *E. coli* concentration in the lake, with higher *E. coli* concentration in the northwest as compared to that in the middle and southern regions of the lake. Figure 4-4 shows the mean contribution of each tributary to *E. coli* concentration in the lake for the simulation period. The contribution from different tributaries to the predicted *E. coli* concentrations at the various raw water intakes is seen to be variable and it is a function of the contaminant load from, and proximity to the tributaries.

Sydenham River, compared to the other tributaries, has a little effect and contribution to *E. coli* concentration in the lake, especially at the location of the water intakes. For the two water intakes at the northwest part of the lake, St. Clair River is the dominant contributor. At Mount Clemens WTP, the contribution is split between St. Clair and Clinton rivers. At

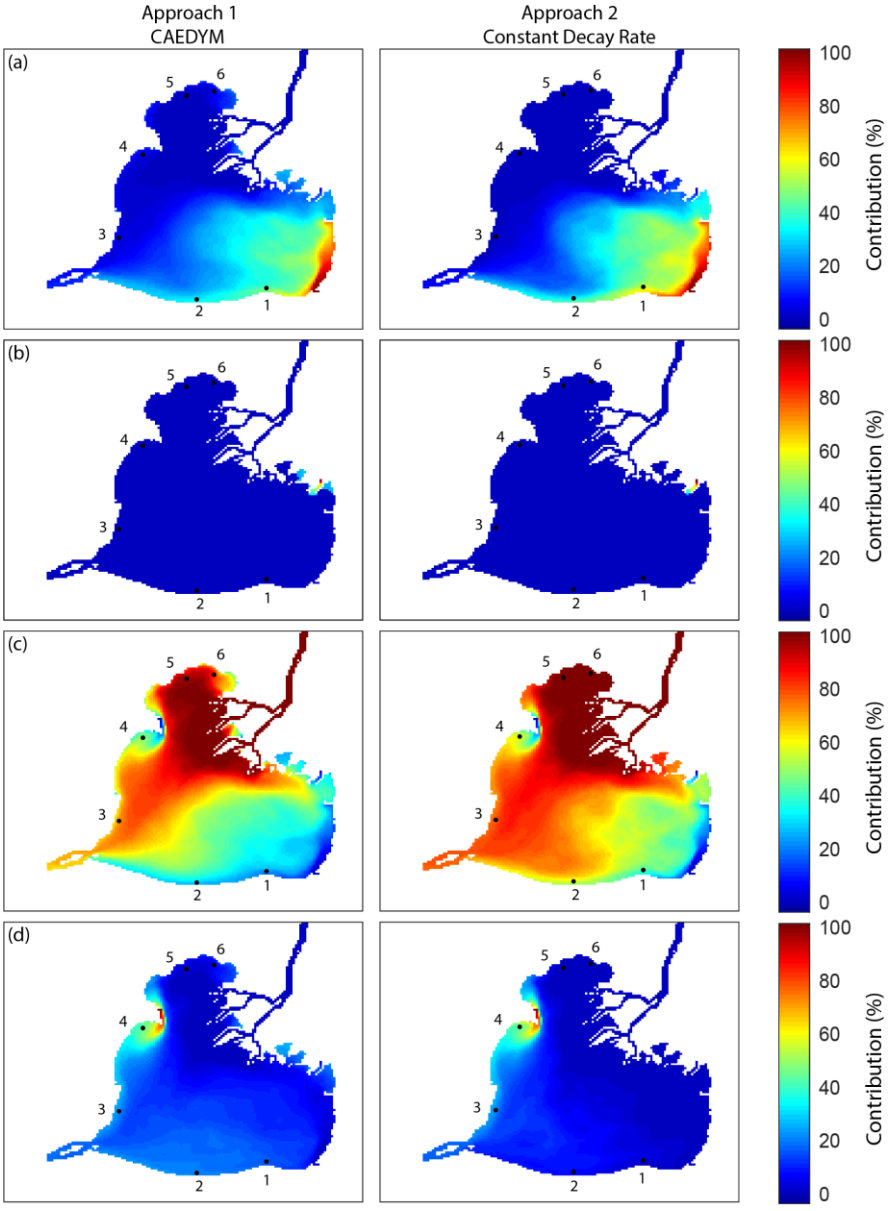

Figure 4-4: Contribution of a) Thames River, b) Sydenham River, c) St. Clair River and d) Clinton River to mean E. coli concentration in Lake St. Clair. Figures in the left column show results with the time-variable decay rate (implemented in CAEDYM). Figures in the right column show results using constant decay rate based on average conditions. Numbers show the location of water treatment plants intakes as were shown in Figure 4-1.

the Stoney Point WTP, the contribution of Thames River is much higher despite St. Clair River being the dominant flow and microbial loading contributor to the lake. Similar to Figure 4-3, while the overall pattern of contributions from the various tributaries to predicted *E. coli* concentrations were similar using the two approaches, there were some differences observed due to variable water temperature and sunlight as discussed in Sections 4.3.4 and 4.3.5.

4.3.4 Effect of Water Temperature

The monthly variation of *E. coli* concentration at six water intakes using two approaches is presented in Figure 4-5a and the relative difference between the two is shown in Figure

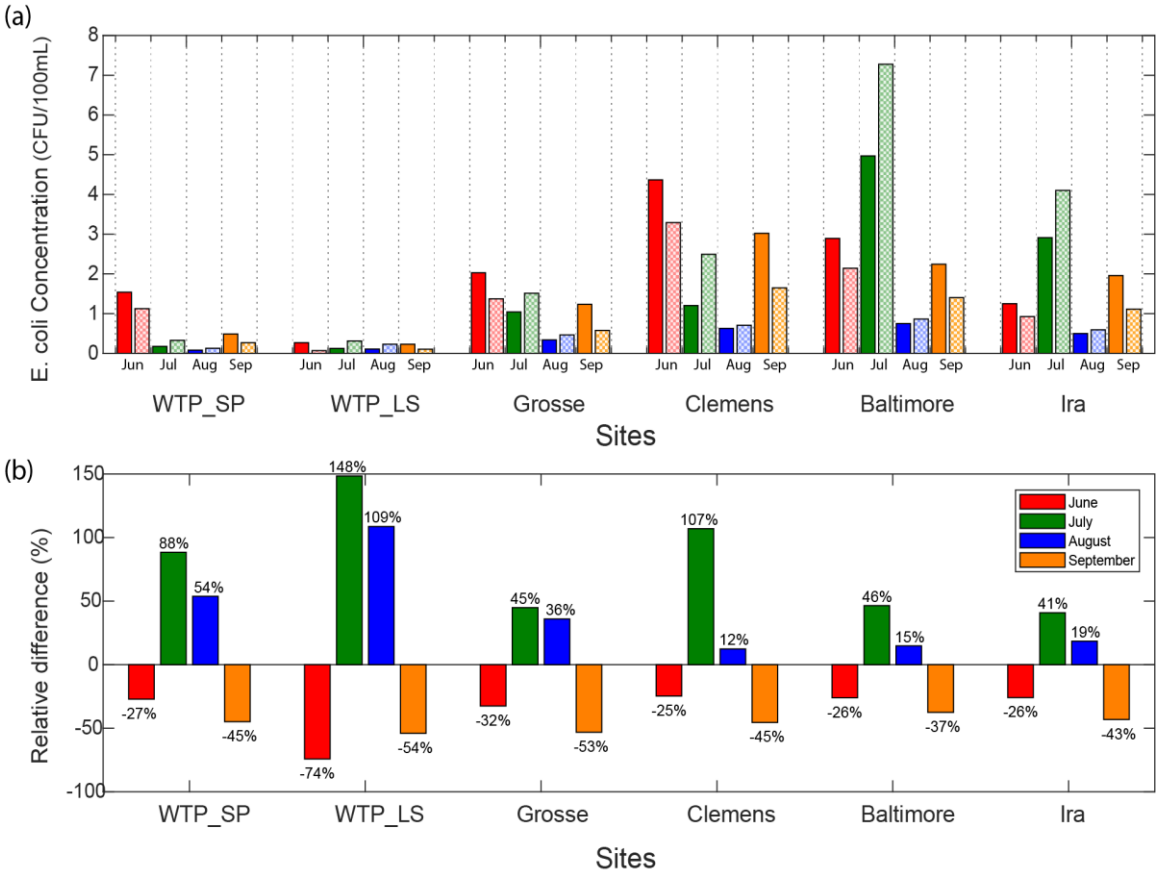

Figure 4-5: Comparison of predicted *E. coli* concentrations using Approach 1 and 2. a) Monthly average concentration of *E. coli* calculated using Approach 1 (solid color) and Approach 2 (pattern filled color) at various WTPs; and b) relative difference between two

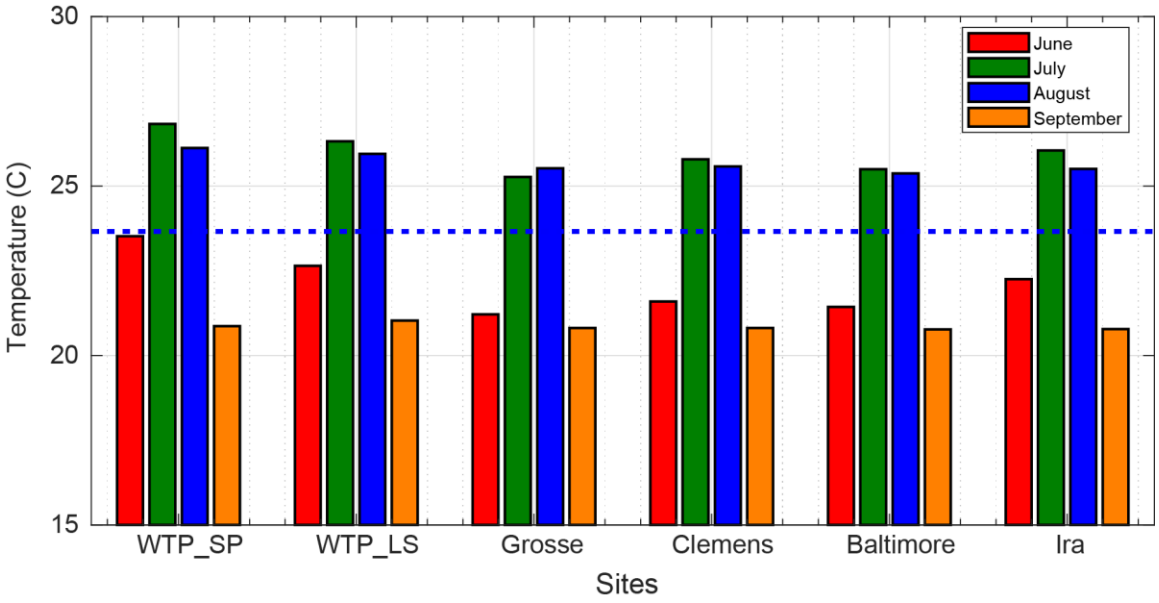

Figure 4-6: Monthly-averaged simulated water temperature at the water treatment plant intakes. The blue dashed line shows the average water temperature of Lake St. Clair for the simulation period.

4-5b. The monthly maximum predicted *E. coli* concentration at any of the WTPs was < 10 CFU/ 100 ml which suggests that the microbial water quality at none of the WTPs is significantly impacted by the microbial loadings from the four major tributaries. Amongst the WTPs, the predicted *E. coli* concentrations were largely determined by their proximity to the St. Clair River. Those sites that are much closer (e.g. Baltimore, Ira, and Clemens WTPs) had much higher predicted *E. coli* concentrations than Lakeshore and Stoney Point WTPs. Except for Baltimore and Ira, where July month has the highest *E. coli* concentration, the predicted *E. coli* concentrations were highest in June for all the WTPs.

While the mean results presented in Figure 4-4 are quite similar, significant month to month variation in predicted *E. coli* concentration is seen using the two approaches. Figure 4-5b shows that the predicted *E. coli* concentrations at various WTPs assuming a constant decay rate (Approach 2) for June and September are 25 – 74% lower as compared to assuming a variable decay rate (Approach 1). At all the sites except for Ira WTP intake, the relative

Table 4-1: Comparison of monthly average E. coli concentration during day and night for different water treatment plant intakes. a) Using time-dependent decay rate and b) the constant decay rate

(a)	WTP_SP		WTP_LS		Grosse		Clemens		Baltimore		Ira	
	Night	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**
June	1.8	1.3	0.4	0.2	2.4	1.6	4.6	4.1	3.5	2.3	1.5	1.0
July	0.2	0.1	0.2	0.0	1.5	0.6	1.6	0.8	5.4	4.6	3.3	2.5
August	0.2	0.0	0.2	0.0	0.7	0.0	1.0	0.3	1.4	0.1	0.8	0.2
September	0.6	0.4	0.3	0.1	1.5	1.0	3.1	2.9	2.8	1.7	2.4	1.5
Avg.	**0.7**	**0.4**	**0.3**	**0.1**	**1.5**	**0.8**	**2.6**	**2.0**	**3.3**	**2.2**	**2.0**	**1.3**
RD (%)	**37**		**68**		**48**		**21**		**33**		**35**	

(b)	WTP_SP		WTP_LS		Grosse		Clemens		Baltimore		Ira	
	Night	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**
June	1.3	1.0	0.1	0.1	1.4	1.3	3.4	3.2	2.1	2.2	0.9	0.9
July	0.3	0.4	0.3	0.3	1.6	1.4	2.3	2.7	7.2	7.4	4.4	3.8
August	0.1	0.2	0.3	0.2	0.4	0.5	0.7	0.7	0.8	0.9	0.6	0.6
September	0.3	0.3	0.1	0.1	0.5	0.6	1.5	1.8	1.4	1.4	1.2	1.0
Avg.	**0.5**	**0.4**	**0.2**	**0.2**	**1.0**	**1.0**	**2.0**	**2.1**	**2.9**	**3.0**	**1.8**	**1.6**
RD (%)	**9**		**-1**		**5**		**-5**		**-2**		**12**	

difference between the two approaches is the highest for July month. These observations can be explained by monthly water temperature variation in the lake. Figure 4-6 shows the monthly averaged simulated water temperature at the water treatment plant intakes. For June and September, the two months that the *E. coli* concentration in the first approach is higher, water temperatures at the various WTPs are lower than the average. When the water temperature is low, according to the Eqs. 3 and 5 the rate of mortality and predation is lower and hence predicted *E. coli* is higher using Approach 1. Although increasing water temperature can slightly increase the growth rate, the rate of increasing mortality and predation is always higher than the growth rate. Water temperatures higher than average are observed during July and August, resulting in 12 – 148% lower predicted *E. coli* concentration using variable decay rate in Approach 1.

4.3.5 Effect of Solar Radiation

As formulated in Eq. 4, inactivation of *E. coli* by sunlight occurs mainly due to visible, UV-A and UV-B wavelengths. The impact of sunlight is included in decay rate calculation under Approach 1 using Eq. 4, with day defined as 8 am - 8 pm, and night as 8 pm - 8 am when there is no contribution due to sunlight inactivation. Monthly averaged day and night *E. coli* concentration for all the sites using the two approaches are presented in Table 4-1. The results (Table 4-1a) show that due to the absence of sunlight inactivation, the average predicted *E. coli* concentrations during nighttime are 21 – 68% higher as compared to the daytime for the various WTPs. As can be seen from Table 4-1b, no particular pattern can be found when the constant decay rate is used to simulate *E. coli* concentration. While the effect of sunlight is averaged out in Approach 2 by using constant decay rate, a small variation of 5 – 12% is still seen, which may be attributed to differences in hydrodynamics during day and nighttime. Sunlight is a major factor influencing the persistence of organisms in environmental waters and its impact on *E. coli* was studied in several studies (Chan et al., 2015; Liang et al., 2016; Liu et al., 2006; Sokolova et al., 2012a).

4.4 Conclusion

In the current study, a three dimensional hydrodynamic model for Lake St. Clair was applied to examine the effect of water temperature and sunlight (Approach 1) on temporal and diurnal microbial water quality at several drinking water intakes during Summer 2010 as contributed by the four major tributaries. The results were compared with those assuming average conditions (Approach 2). The results show that while the overall patterns were similar, a 12 – 148% difference in predicted *E. coli* concentrations at the various WTPs were observed between the two approaches. The differences observed were different

at each of the WTPs and varied temporally. Predicted *E. coli* concentrations using Approach 1 were higher during June and September, and lower during July and August, as compared to Approach 2. Average nighttime *E. coli* predictions were 21 – 68% higher as compared to daytime levels. These results suggest that varying water temperature and sunlight can significantly affect the dynamics of microbial water quality. Decay rate formulations including these effects are thus maybe preferable in microbial water quality models and should be used over assuming constant rates wherever possible.

4.5 Acknowledgments

The authors are grateful to the Ontario Ministry of the Environment and Climate Change (MOECC) for the ADCP data and also to Windsor Essex County Health Unit (WECHU) for providing the weekly *E. coli* data. We thank Environment and Climate Change Canada (ECCC) and Canada Centre for Inland Waters for their support and special thanks to Dr. Tom Edge and his team for providing *E. coli* data for the Thames River. This research received financial support from Canada's Natural Sciences and Engineering Research Council (NSERC) Strategic Project Grant (SPG) and was performed while Mr. Madani (first author) held an Ontario Trillium Scholarship (OTS).

4.6 Supplementary Material

Supplementary data to this article can be found in Appendix 4.

4.7 References

Baustian, M.M., Mavrommati, G., Dreelin, E.A., Esselman, P., Schultze, S.R., Qian, L., Aw, T.G., Luo, L., Rose, J.B., 2014. A one hundred year review of the socioeconomic and ecological systems of Lake St. Clair, North America. Journal of Great Lakes Research 40(1) 15-26.

Blaustein, R.A., Pachepsky, Y., Hill, R.L., Shelton, D.R., Whelan, G., 2013. Escherichia coli survival in waters: Temperature dependence. Water Research 47(2) 569-578.

Bonamano, S., Madonia, A., Borsellino, C., Stefani, C., Caruso, G., De Pasquale, F., Piermattei, V., Zappala, G., Marcelli, M., 2015. Modeling the dispersion of viable and

total Escherichia coli cells in the artificial semi-enclosed bathing area of Santa Marinella (Latium, Italy). Mar Pollut Bull 95(1) 141-154.

Brookes, J.D., Antenucci, J., Hipsey, M., Burch, M.D., Ashbolt, N.J., Ferguson, C., 2004. Fate and transport of pathogens in lakes and reservoirs. Environment International 30(5) 741-759.

Chan, S.N., Thoe, W., Lee, J.H., 2013. Real-time forecasting of Hong Kong beach water quality by 3D deterministic model. Water Research 47(4) 1631-1647.

Chan, Y.M., Thoe, W., Lee, J.H.W., 2015. Field and laboratory studies of Escherichia coli decay rate in subtropical coastal water. Journal of Hydro-Environment Research 9(1) 1-14.

de Brauwere, A., Ouattara, N.K., Servais, P., 2014. Modeling fecal indicator bacteria concentrations in natural surface waters: a review. Critical Reviews in Environmental Science and Technology 3389(March 2015) 140225123512007-140225123512007.

Elmoustafa, A.M., 2017. Evaluation of water intake location suitability using a hydrodynamic approach. Journal of Applied Water Engineering and Research 5(1) 31-39.

Fiandrino, A., Martin, Y., Got, P., Bonnefont, J.L., Troussellier, M., 2003. Bacterial contamination of Mediterranean coastal seawater as affected by riverine inputs: simulation approach applied to a shellfish breeding area (Thau lagoon, France). Water Research 37(8) 1711-1722.

Garcia-Armisen, T., Thouvenin, B., Servais, P., 2006. Modelling faecal coliforms dynamics in the Seine estuary, France. Water Science and Technology 54(3) 177-184.

Hamelin, K., Bruant, G., El-Shaarawi, A., Hill, S., Edge, T.A., Fairbrother, J., Harel, J., Maynard, C., Masson, L., Brousseau, R., 2007. Occurrence of virulence and antimicrobial resistance genes in Escherichia coli isolates from different aquatic ecosystems within the St. Clair River and Detroit River areas. Appl Environ Microbiol 73(2) 477-484.

Hipsey, M.R., Antenucci, J.P., Brookes, J.D., 2008. A generic, process-based model of microbial pollution in aquatic systems. Water Resources Research 44(7).

Hipsey, M.R., Antenucci, J.P., Brookes, J.D., Burch, M.D., Regel, R.H., Linden, L., 2004. A three dimensional model of Cryptosporidium dynamics in lakes and reservoirs: A new tool for risk management. International Journal of River Basin Management 2(3) 181-197.

Hodges, B.R., Dallimore, C., 2006. Estuary, Lake and Coastal Ocean Model.

Holtschlag, D.J., Shively, D., Whitman, R.L., Haack, S.K., Fogarty, L.R., 2008. Environmental factors and flow paths related to Escherichia coli concentrations at two beaches on Lake St. Clair, Michigan, 2002–2005, Scientific Investigations Report: Reston, VA.

Huziy, O., Sushama, L., 2017. Lake–river and lake–atmosphere interactions in a changing climate over Northeast Canada. Climate Dynamics 48(9) 3227-3246.

Imberger, J., Marti, C., Dallimore, C., Hamilton, D., Escriba, J., Valerio, G., 2017. REAL-TIME, ADAPTIVE, SELF-LEARNING MANAGEMENT OF LAKES.

Leon, L.F., Lam, D., Schertzer, W., Swayne, D., 2005. Lake and climate models linkage: a 3-D hydrodynamic contribution. Adv. Geosci. 4 57-62.

Liang, L., Goh, S.G., Gin, K.Y.H., 2016. Decay kinetics of microbial source tracking (MST) markers and human adenovirus under the effects of sunlight and. Science of The Total Environment 574(October) 165-175.

Liang, L., Goh, S.G., Gin, K.Y.H., 2017. Decay kinetics of microbial source tracking (MST) markers and human adenovirus under the effects of sunlight and salinity. Science of The Total Environment 574 165-175.

Lin, B., Falconer, R.A., 2001. Numerical modelling of 3-d tidal currents and water quality indicators in the Bristol Channel. Proceedings of the Institution of Civil Engineers - Water and Maritime Engineering 148(3) 155-166.

Liu, L., Phanikumar, M.S., Molloy, S.L., Whitman, R.L., Shively, D.A., Nevers, M.B., Schwab, D.J., Rose, J.B., 2006. Modeling the transport and inactivation of E. coli and enterococci in the near-shore region of Lake Michigan. Environmental Science & Technology 40(16) 5022-5028.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020. Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach. Journal of Great Lakes Research.

Mosley, L.M., Daly, R., Palmer, D., Yeates, P., Dallimore, C., Biswas, T., Simpson, S.L., 2015. Predictive modelling of pH and dissolved metal concentrations and speciation following mixing of acid drainage with river water. Applied Geochemistry 59 1-10.

Na, E.H., Park, S.S., 2005. A HYDRODYNAMIC MODELING STUDY TO DETERMINE THE OPTIMUM WATER INTAKE LOCATION IN LAKE PALDANG, KOREA1. JAWRA Journal of the American Water Resources Association 41(6) 1315-1332.

Romero, J.R., Antenucci, J.P., Imberger, J., 2004. One- and three-dimensional biogeochemical simulations of two differing reservoirs. Ecological Modelling 174(1) 143-160.

Sokolova, E., Astrom, J., Pettersson, T.J., Bergstedt, O., Hermansson, M., 2012a. Decay of Bacteroidales Genetic Markers in Relation to Traditional Fecal Indicators for Water Quality Modeling of Drinking Water Sources. Environmental Science & Technology 46(2) 892-900.

Sokolova, E., Astrom, J., Pettersson, T.J., Bergstedt, O., Hermansson, M., 2012b. Estimation of pathogen concentrations in a drinking water source using hydrodynamic modelling and microbial source tracking. J Water Health 10(3) 358-370.

Sokolova, E., Pettersson, T.J.R., Bergstedt, O., Hermansson, M., 2013. Hydrodynamic modelling of the microbial water quality in a drinking water source as input for risk reduction management. Journal of Hydrology 497 15-23.

Tranmer, A.W., Marti, C.L., Tonina, D., Benjankar, R., Weigel, D., Vilhena, L., McGrath, C., Goodwin, P., Tiedemann, M., McKean, J., Imberger, J., 2018. A hierarchical modelling framework for assessing physical and biochemical characteristics of a regulated river. Ecological Modelling 368 78-93.

Trolle, D., Hamilton, D.P., Hipsey, M.R., Bolding, K., Bruggeman, J., Mooij, W.M., Janse, J.H., Nielsen, A., Jeppesen, E., Elliott, J.A., Makler-Pick, V., Petzoldt, T., Rinke, K., Flindt, M.R., Arhonditsis, G.B., Gal, G., Bjerring, R., Tominaga, K., Hoen, J.t., Downing, A.S., Marques, D.M., Fragoso, C.R., Søndergaard, M., Hanson, P.C., 2012. A community-based framework for aquatic ecosystem models. Hydrobiologia 683(1) 25-34.

Valipour, R., León, L.F., Depew, D., Dove, A., Rao, Y.R., 2016. High-resolution modeling for development of nearshore ecosystem objectives in eastern Lake Erie. Journal of Great Lakes Research 42(6) 1241-1251.

Valipour, R., Rao, Y.R., León, L.F., Depew, D., 2018. Nearshore-offshore exchanges in multi-basin coastal waters: Observations and three-dimensional modeling in Lake Erie. Journal of Great Lakes Research.

CHAPTER 5 MICROBIAL MODELLING OF LAKE ST. CLAIR: IMPACT OF LOCAL TRIBUTARIES ON THE SHORELINE WATER QUALITY

5.1 Introduction

Fecal pollution of surface water resources is a global issue of concern. The greatest risks are associated with the ingestion of water that is contaminated with water-borne pathogen contamination from human or animal feces (Pandey et al., 2014; WHO, 2003). Detection of waterborne fecal pathogens is very difficult and costly, and thus not recommended as part of a regular monitoring program. Hence, various fecal indicator bacteria (FIB), such as *Escherichia coli* (*E. coli*) are usually used to detect fecal pollution in natural waters (Health-Canada, 2012; USEPA, 2012; WHO, 2017).

In Canada, raw drinking water sources and recreational water are assessed through routine monitoring of *E. coli* (Health-Canada, 2012, 2020). Although monitoring for *E. coli* can provide good information that can be used in assessing microbiological risks and treatment requirements for surface water sources, it has several limitations that need to be addressed. For example, in large water bodies with upstream watersheds impacted by human and animal fecal pollution, significant fecal microbial pollution is mobilized into their tributaries during storm events. This results in a significant and dynamic increase in microbial loadings to the receiving waters. Such changes in the loads produce dynamic and unexpected changes in the microbial water quality of the receiving water body, which are difficult and costly to capture through monitoring studies. Additional monitoring obstacles include requirements for equipment, laboratory and field technicians, and the fact it cannot help identifying potential sources of fecal pollution, which could aid in its practical management.

To overcome the monitoring limitations of monitoring, mechanistic mathematical models can be used that can link pollution sources to receptors as a function of changing environmental conditions. Such models can be calibrated and tested using the monitoring data collected. They then can be very useful tools for water resources managers to predict and control expected water quality changes at water intakes, or to identify and control potential sources of pollution. Models can also provide a platform to compare the effectiveness of source water protection or hazard control measures at a fraction of the cost and time required to build and test them, to select the ones that may give the best value for the money. Many mechanistic models have been developed and applied to simulate fate and transport of microbial pollution in water bodies (Bonamano et al., 2015; Bravo et al., 2017; Garcia-Armisen et al., 2006; Liu et al., 2006; Muirhead and Monaghan, 2012; Safaie et al., 2016). These models have demonstrated better spatial and temporal resolution of FIB distributions at a lower cost when combined with judicious monitoring.

Lake St. Clair is a precious natural resource that provides drinking water for millions of people in the United States and Canada and numerous recreational opportunities. The water quality concerns in Lake St. Clair include pathogens, toxic contaminants and eutrophication. Although there have been some studies that examined hydrodynamic and/or nutrient transport modelling (Anderson et al., 2010; Bocaniov and Scavia, 2018; Healy et al., 2007; Holtschlag and Koschik, 2002), to the best of our knowledge there is no study on 3D microbial water quality modelling for Lake St. Clair. Accounting for >99.5% of flows into Lake St. Clair, the four major tributaries (St. Clair, Sydenham, Thames and Clinton) are also the dominant contributors of microbial pollution to the Lake. In a recent study by (Madani et al., 2020a), a high-resolution 3D hydrodynamic model was developed

using the Aquatic Ecosystem Model (AEM3D) modelling platform and applied to assess the impact of microbial pollution from the four major tributaries on microbial water quality in the lake as well as on a popular beach (Sandpoint Beach, Windsor, Ontario, Canada) located on the southern edge of the lake. The results showed that in a large part of the lake, starting from the middle of the lake to the southern shoreline, the maximum predicted *E. coli* concentration from the combined input of the four major tributaries was very low at < 10 CFU/100 ml.

Eight small but seasonally significant tributaries of widely varying drainage areas and hydrologic features (hereafter referred to as "small tributaries") discharge into Lake St. Clair along the southern shores. The flow and pollution load from these small tributaries are expected to be negligible in the context of the larger lake, and therefore have been ignored or not studied comprehensively in previous hydrodynamic and water quality modelling studies of Lake St. Clair (Anderson et al., 2010; Bocaniov and Scavia, 2018; Healy et al., 2007; Holtschlag and Koschik, 2002; Madani et al., 2020a). However, monitoring by Essex Region Conservation Authority (ERCA) and Provincial (Stream) Water Quality Monitoring Network (PWQMN) provides evidence of frequent occurrence, and at times appreciable high levels of *E. coli* in these tributaries (ERCA, 2015). There are two popular public beaches on, and two drinking water intakes close to, the southern shores of Lake St. Clair. The potential impact of one or more of the small tributaries on the microbial water quality at the two beaches or the drinking water intakes is currently unknown.

In general, many laboratory and field investigations have been conducted to examine environmental and ecological variables affecting *E. coli* decay, and the derived

relationships are being included in microbial water quality models. In another recent study on Lake St. Clair, Madani et al. (2020b) showed that a time-dependent decay rate using such relationships may be expected to give more realistic predictions than using a conservative estimate based on literature values. Such relationships are been included in an ecological modelling framework and then coupled with hydrodynamic models for microbial water quality modelling (Brookes et al., 2004; Chen and Liu, 2017; Cho et al., 2016; Gao et al., 2015; Hipsey et al., 2008).

Overall the fate and transport of microbial pollution in a lake environment are influenced by hydrodynamics, microbial load (from tributaries), and die-off rate (decay) of the microbes which depend on the environmental conditions and physicochemical characteristics of the water body. The primary objective of the current study was to examine the added impact of the eight small tributaries along the southern shores of Lake St. Clair on the microbial water quality in Lake St. Clair during the summer of 2016, particularly in the area surrounding the southern shoreline. The hydrodynamic was simulated using the high-resolution AEM3D hydrodynamic model developed in our previous study (Madani et al., 2020a) and adapted for 2016. The fate of *E. coli* was modelled using CAEDYM (Computational Aquatic Ecosystem DYnamics Model), available within the AEM3D modelling framework. Decay formulation based on water temperature and solar radiation that accounts for dark mortality rate, light inactivation, growth and predation in the environment was implemented. The spatiotemporal features of impacts on the flow and water quality at the shoreline adjacent to the tributary outlets were investigated. Modelling results predict the areas of highly *E. coli* across the southern

shoreline of Lake St. Clair and help identify the contribution extent of each tributary to microbial pollution in the region.

5.2 Data and Methods

5.2.1 Study Site:

The geographic extent of this study is limited to the Lake St. Clair (42°17'33.23"- 42°41'46.02" N- Latitude, 82°25'9.32"- 82°55'45.59"W-Longitude). A subsection of the southern shoreline - an approximate 40 km of shoreline from the mouth of the Thames River to Sandpoint Beach - is of interest for a more detailed analysis of shoreline water quality and environmental conditions (Figure 5-1). This shoreline also encompasses the two Lake St. Clair beaches: Sandpoint Beach (42°20'19.41" N, 82°55'8.61" W) is a 300 m long and relatively shallow (2 m) beach located at the mouth of the Detroit River; and Lakeview Park West Beach (42°17'51.10"N, 82°42'41.61"W) (LP Beach), located in the mouth of the Belle River, is a small (150 m long) but still a popular beach in the region especially because of the adjacent Belle River Marina. According to the Windsor-Essex County Health Unit (WECHU) report and field measurements, both beaches are frequently posted as unsafe for water recreation for the summer seasons due to incidents involving high bacterial counts (McPhedran et al., 2013).

5.2.2 *E. coli* Data:

E. coli data for Lake St Clair and Thames River was obtained over the 2016-2018 sampling season (unpublished results; Tom Edge 2018). Samples were collected by boat at sites a few hundred meters offshore, usually associated with river mouths. Sampling sites are illustrated in Figure 5-1.

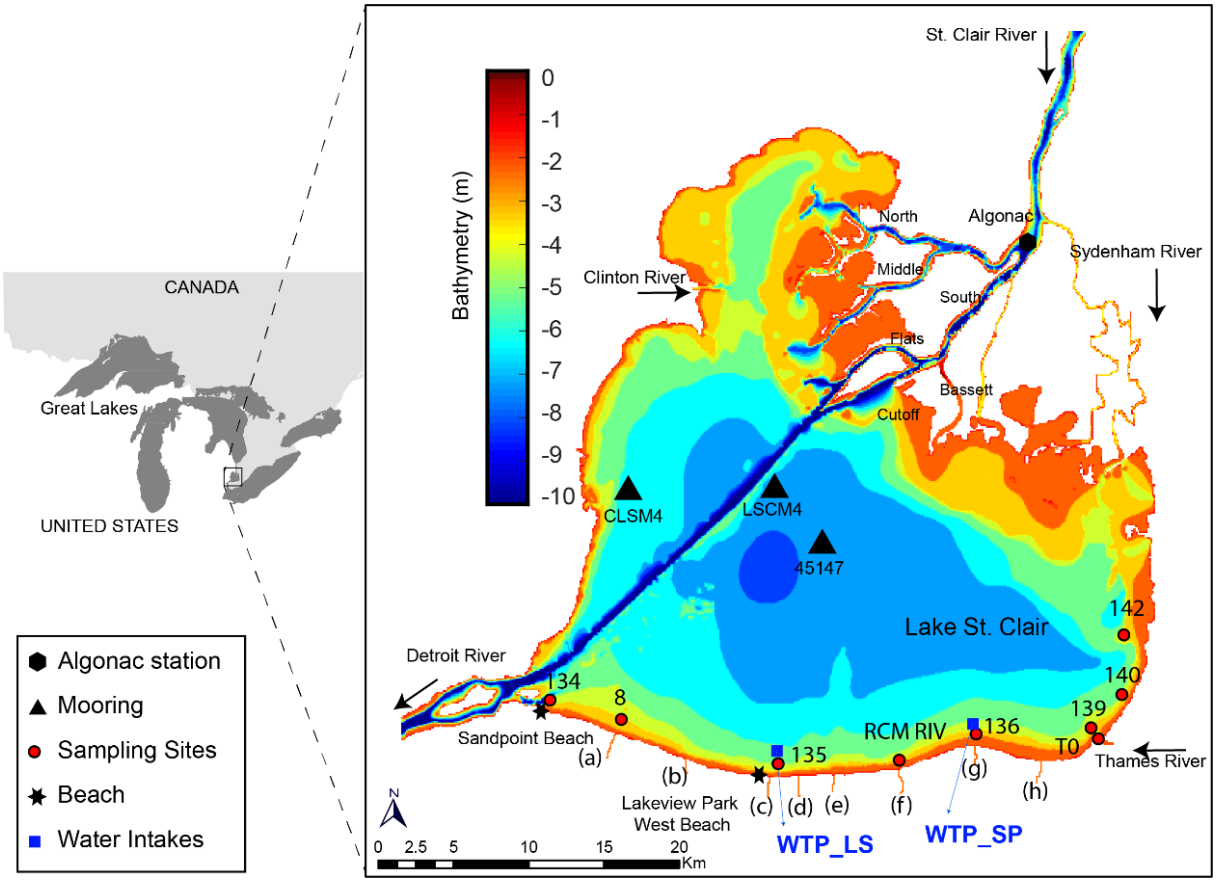

Figure 5-1: Map of Lake St. Clair and its bathymetry. Triangle symbols show the locations of the moorings (CLSM4 (42.471 N 82.877 W), LSCM4 (42.465 N 82.755 W) and 45147 (42.430 N 82.680 W)). Location of Sandpoint Beach and Lakeview Park Beach (LP) are identified by black stars. Bathymetric data are in meters and all major tributaries are labelled on the map. Small tributaries are a) Pike Creek, b) Puce Creek, c) Belle River, d) Duck Creek, e) Moison Creek, f) Ruscom River, g) Stoney Point drainage area and h) Little Creek.

More *E. coli* data was obtained from the water treatment plants in the region. Lakeshore water treatment plant (LS_WTP) with a treatment capacity of 36,400 m3/day, located near the Belle River and Stoney Point water treatment plant (SP_WTP) with a rated treatment capacity of 4,545 m3/day is located near the Stoney Point drainage area (Figure 5-1) draw their source waters from Lake St. Clair. For both treatment plants, *E. coli* sampling from the raw water typically occurred on a weekly basis.

5.2.3 Bathymetry, Forcing Data and Flow and *E. coli* Boundary Conditions:

The bathymetry, sourced from National Geophysical Data Center (https://www.ngdc.noaa .gov/), has depths referenced to a generic datum of 176.784 meters (580 feet) and is a rectangular grid with cell sizes of ~70 m by ~90 m that was processed into coarser grid size of 400 m uniform grids. Meteorological forcing data sets include wind direction, wind speed, air temperature, humidity, atmospheric pressure, rainfall, and solar radiation. Data is from the nearest stations (Windsor A. Station ID=4716; Windsor Riverside Station ID=4715) and moored surface buoys (Buoy (West Erie) 45005 (41.677 N 82.398 W); Buoy 45147 (42.430N 82.680W); LSCM4 (42.465N 82.755W); CLSM4 (42.471N 82.877W)). Buoy 45147, which is maintained by Environment and Climate Change Canada (ECCC), was used to fill out missing wind direction and wind speed data and for model validation of water temperature. Solar radiation data were obtained from the Canadian Regional Climate Model (CRCM5) provided by the University of Quebec at Montreal (UQAM) (Huziy and Sushama, 2017).

Flow and temperature data of the four major tributaries to the lake (Thames, Sydenham, St. Clair, and Clinton Rivers) and the outflow (Detroit River) were obtained from the nearest gauged stations as described in the previous paper (Madani et al., 2020a) for the year 2016. The boundary condition for Thames River is based on the data from site T0 at the river mouth. In the case of small tributaries (Pike Creek, Puce River, Belle River Duck Creek, Moison Creek Ruscom River and Stoney Point drainage area and Little Creek) the observed flow is only available for Ruscom River at one station (02GH002). Flow for other tributaries was obtained using their watershed area ratio to the Ruscom River watershed. *E. coli* data for small tributaries are limited to few grab samples collected by ERCA and

PWQMN. LOADEST was used to correlated flow with *E. coli* to obtain time series of *E. coli* concentration for each tributary. Information of mean, max and 90th percentile *E. coli* concentration from all tributaries are presented in Table 5-S1 in Appendix 5-1.

5.2.4 Modelling Framework:

5.2.4.1 Hydrodynamic Driver and Ecological Model

In the current study, ELCOM (Estuary and Lake COmputer Model), an advanced three-dimensional model, was used as the hydrodynamic driver and coupled with CAEDYM for modelling pathogen transport in Lake St. Clair. The simulation period is from 1 May 2016 until 7 October 2016, encompassing a range of meteorological and *E. coli* occurrence events. The first month is considered as warm-up period and results are not shown here. ELCOM-CAEDYM is presently available as AEM3D from Hydronumerics (http://hydronumerics.com.au/) simulate three-dimensional transport and interactions of flow physics, thermodynamics and ecology in the reservoir. It has been successfully applied to many large lakes around the world (Caramatti et al., 2020; Romero et al., 2004; Silva et al., 2014) and Great Lakes including Lake St. Clair (Bocaniov and Scavia, 2018; Madani et al., 2020a), Lake Erie (Leon et al., 2005; Leon et al., 2011; Oveisy et al., 2014; Valipour et al., 2016) and Lake Ontario (Leon et al., 2012; Paturi et al., 2015) and a review of the coupled model system has recently been performed by Trolle et al. (2012).

5.2.4.2 Decay Formulation

A detailed description of the configuration of AEM3D used here and its validation against previous studies and observation data such as water temperature for 2010 is presented in detail in (Madani et al., 2020a).

In the current study, we assumed the survival of *E. coli* is dependent on several factors including physio-chemical (abiotic) (e.g. temperature, and sunlight) and biological (biotic) factors (e.g. growth and also predation in the presence of other competing predation, and grazing by larger eukaryotic organisms such as protozoa ((Byappanahalli et al., 2012; Hipsey et al., 2008; Ishii and Sadowsky, 2008).

A detailed description of the formula we used for each of the parameters of decay and growth rate are presented by Hipsey et al. (2008), as summarized in Section 4.2.5.

5.2.4.3 Relative Contribution of Tributaries to Flow and *E. coli* Concentration

The flow contribution of each tributary is defined as the relative quantity of water received from the inflow of that tributary to the location of concern. To calculate this a conservative soluble tracer is used to simulate the distribution of the effluent plume. Different tracers are defined for each tributary and their concentration set to unity for all the tributaries, so it represents the inflow. The conservative tracer with zero decay rate is employed so that the impacts of receiving water hydrodynamics isolated from those associated with decay processes.

To calculate the relative contribution of tributaries to the *E. coli* concentration at different sites, a similar process is repeated but the tracer concentration for each tributary are set to the estimated *E. coli* concentration and decay rate considered to be equal to 0.9 d^{-1} which is obtained from the average condition during the simulation period according to the Eq. 2 in Section 4.2.5. Flow and *E. coli* contribution of each tributary at a point in the lake can be expressed mathematically as:

$$\text{Relative flow or } E.\ coli \text{ contribution } (\%) = \frac{C_{trib_i}}{\sum_{i=1}^{N} C_{trib_i}} \times 100 \quad (1)$$

where C_{trib_i} is the tracer concentration receiving from tributary i at the location of interest and N is the number of tributaries.

5.3 Results and Discussion

5.3.1 Lake Hydrodynamic:

The model predicted circulation pattern and currents in Lake St. Clair were in good agreement with previous studies for the lake, as was shown in a recent study by the authors (Madani et al., 2020a). We also showed a high degree of agreement between the observed and model-predicted temporal and spatial distribution of water temperature in the lake.

5.3.2 *E. coli* Distribution in Lake St. Clair and Effect of Small Tributaries

Model simulations were carried out with and without several small tributaries along the southern edge of Lake St. Clair shown in Figure 5-1, and the results for maximum, 90th percentiles and mean are predicted *E. coli* concentration presented in Figure 5-2. The

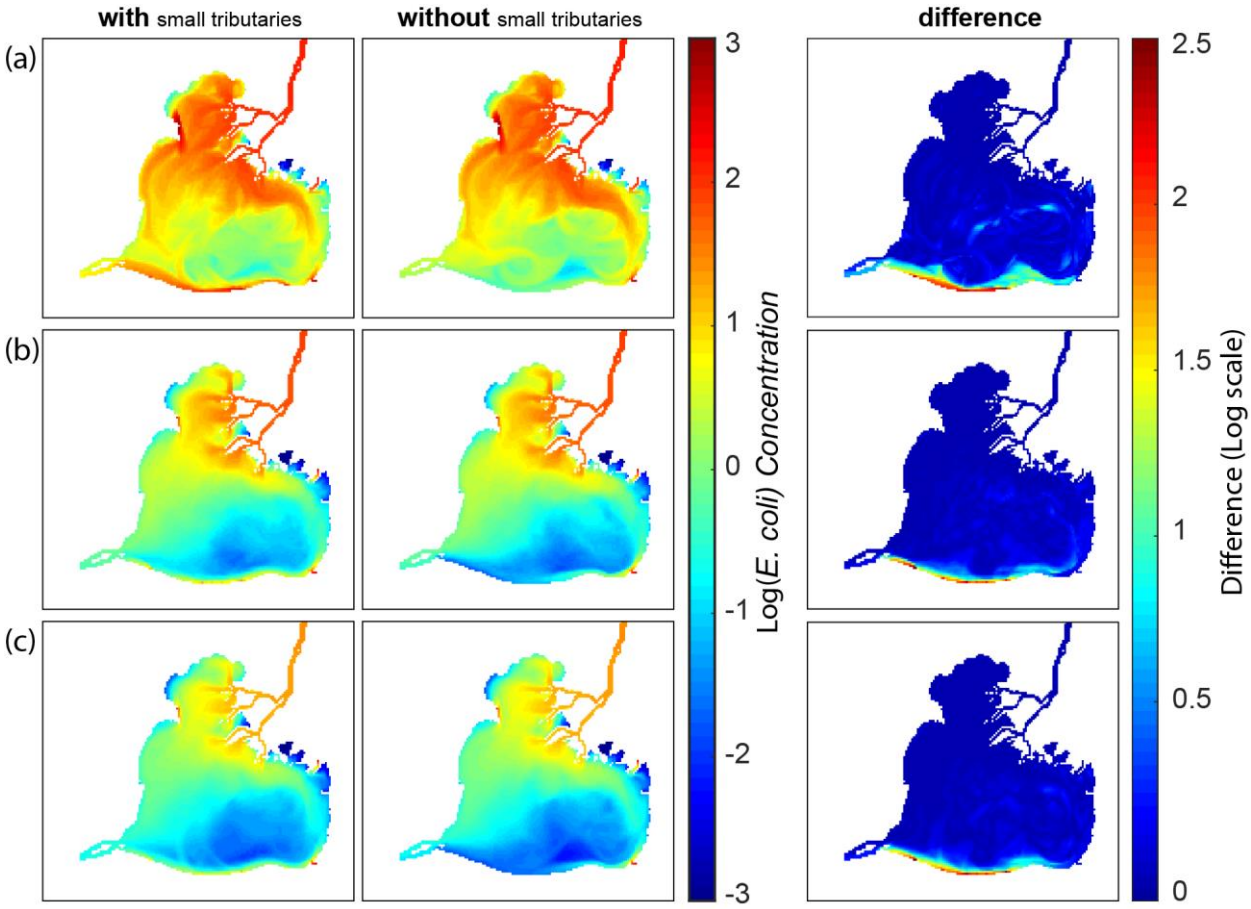

Figure 5-2: (a) Max, (b) 90th percentile, and (c) mean predicted *E. coli* concentrations in Lake St. Clair over June 1st to October 7th, 2016 with (left) and without (middle) considering small tributaries. The graphs on the right show the difference between the two.

results show that while the predicted *E. coli* concentrations for much of Lake St. Clair are not significantly influenced by the small tributaries, their impact on a narrow nearshore region along the southern edge is quite significant. The similarity in regions of influence seen for the 90th percentile and mean differences in *E. coli* concentrations (Figure 5-2b, c) shows that in this region, the influence of the small tributaries is significant most of the time. Occasionally the influence can extend beyond this narrow nearshore region, as seen for the maximum (Figure 5-2a), but are relatively rare. High water age (Madani et al.,

2020a) and high decay rate in the middle, south and southeastern part of the lake are responsible for lower *E. coli* count in those regions.

Limited *E. coli* monitoring data in the nearshore region is available from the field measurement (unpublished paper; Tom Edge, 2018) and the two Town of Lakeshore water treatment plants. The locations of the sites are shown in Figure 5-1. A comparison of the time series of model simulated results with and without the inclusion of the small tributaries, and these limited observations is shown in Figure 5-3. For Sites 134, 135, 136 and the two drinking water intakes, observed *E. coli* concentration are <10 CFU/100 ml most of the time, which is in good agreement with model results. The *E. coli* concentrations simulated without the inclusion of the smaller tributaries were close to zero most of the times, with occasional spikes of <4 CFU/100 ml. More spikes in simulated *E. coli* concentrations were seen with the inclusion of the smaller tributaries, occasionally ranging between 10 – 60 CFU/100 ml, which are in better agreement with the observations. These sites are thus significantly influenced by microbial pollution from the smaller tributaries. The higher *E. coli* concentrations at these sites as a result of this influence, ranging between 20 – 60 CFU/100 ml, are predicted during a significant rainfall event (See Appendix 5: Figure 5-S2 for rainfall data) near the end of September 2016. Results of simulated *E. coli* concentration with and without the inclusion of small tributaries at Site S140 are very similar, except for occasional differences, and also are in good agreement with observations. This site is on the northeast of and close to Thames River, which is one of the major tributaries and included in both models. It is also farther away from the smaller tributaries as compared to most of the other sampling sites. The results suggest that the site

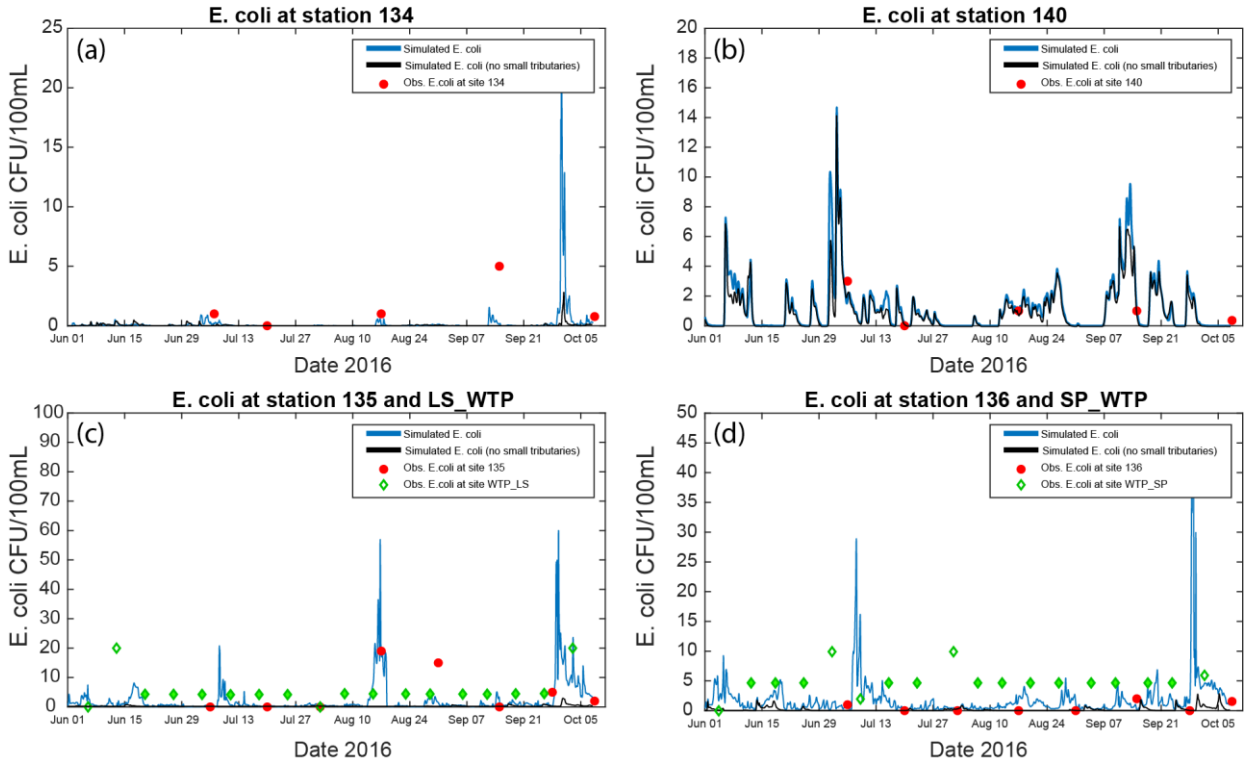

Figure 5-3: Time series of simulated *E. coli* concentration with (blue solid line) and without (black solid line) considering small tributaries and its comparison with measured data at different locations. In c) and d) observations that are reported as <10 CFU/100 ml (filled green diamonds) are shown as 5 CFU/100 ml for illustration.

is dominantly influenced by Thames River, with little to no influence of the smaller tributaries, most of the time which explains the lack of difference observed in the model results. The smaller tributaries can have some influence on the *E. coli* concentrations during unique weather conditions.

The range and mean observed concentrations at the various sites are compared with various model simulations in Figure 5-4. The simulations with small tributaries (ST) include those with a time-variable decay rate (predicted using CAEDYM; orange bars) and constant decay rate (average of decay rates over the domain and simulation period; blue bars). Model simulations were also obtained with time-variable decay but without including the

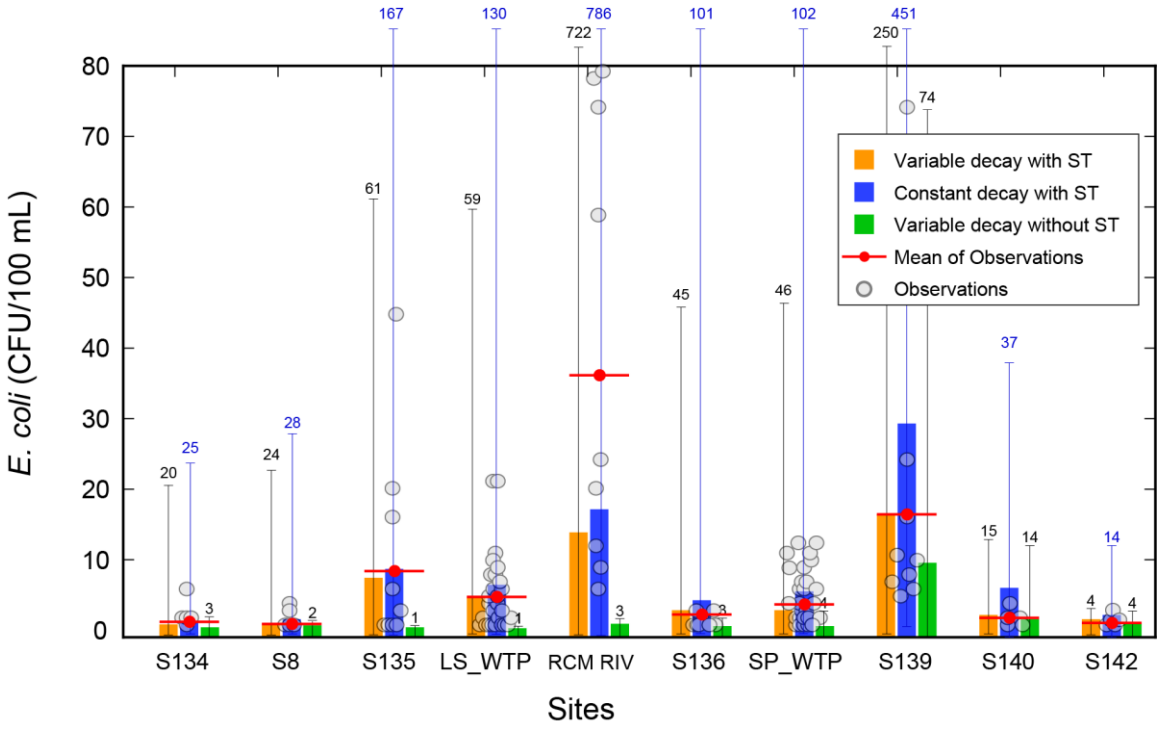

Figure 5-4: Comparison of range and mean *E. coli* concentration at different sites with model simulations using i) time-variable decay with small tributaries (Variable decay with ST); ii) constant decay with small tributaries (Constant decay with ST), and iii) time-variable decay without small tributaries (Variable w/o ST). Whiskers and the value above them show the maximum simulated values. Whiskers passed the y-axis limit provide the number but are not plotted to scale.

small tributaries (green bars). For Sites S134, S8, S140, and S142 which are farther away from the smaller tributaries, the observed *E. coli* concentrations are <5 CFU/100 ml, and the model simulations are similar. For Sites S135, S136, Town of Lakeshore drinking water intakes (LS-WTP and SP-WTP), and the site near the confluence of Ruscom River (RCM RIV), the mean and range of *E. coli* concentrations predicted using model simulations including small tributaries ("Variable decay with ST" and "Constant decay with ST"; Figure 5-4) are higher than those without inclusion small tributaries, and also in better agreement with the observations. These results confirm that the microbial water quality in the nearshore region close to the confluence of the eight smaller tributaries along the

Table 5-1: Flow contribution (%) of the tributaries at the site locations along the southern shoreline of Lake St. Clair shown in Figure 5-1

Tributaries\Sites	**S134**	**S8**	**LP**	**S135**	**LS_WTP**	**RCM RIV**	**S136**	**SP_WTP**	**S139**	**S140**	**S142**
Pike Creek	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Puce River	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.0	0.0	0.0
Belle River	0.8	0.8	16.8	12.6	8.5	2.9	2.2	2.2	0.3	0.9	1.1
Duck Creek	0.3	0.3	1.1	1.5	1.7	0.8	0.6	0.6	0.1	0.3	0.3
Moison Creek	0.4	0.4	1.0	1.4	1.6	1.1	0.8	0.8	0.1	0.4	0.4
Ruscom River	2.1	2.3	3.1	5.2	5.6	29.0	7.4	7.4	0.8	2.9	2.8
Little Creek	0.4	0.5	0.5	0.8	0.8	0.7	1.8	1.8	0.2	0.7	0.7
Stoney Point	0.3	0.4	0.4	0.6	0.6	0.6	5.0	5.0	0.1	0.5	0.5
Major tributaries	95.2	95.3	76.9	77.6	80.9	64.7	82.2	82.2	98.4	94.2	94.1

southern edge of Lake St. Clair is significantly affected by one or more of these tributaries, and the effect is captured by both models using a time-variable or constant decay rates. The range of predicted *E. coli* concentrations using the time-variable decay rate is in better agreement with the range of observed values than that using the constant decay rate, suggesting that the use of time-variable decay may be preferable for dynamic simulations over time. A similar conclusion has been drawn in previous studies (Madani et al., 2020b).

5.3.3 Area of *E. coli* Influence and Relative Contribution of Each Tributary

The proximity of a study site to the tributary locations, hydrodynamic conditions, and inflows from the other major tributaries determine the contribution of the tributary to the water quality at that site. The combined flow from the eight small tributaries is very small and estimated to be <0.2% of the total inflows arriving into Lake St. Claire. The relative contribution of the eight small tributaries to the water at different sites near the southern shore of Lake St. Clair, as compared to the major tributaries is presented in Table 5-1. Given the proximity of the small tributaries to sites, their contribution to the water at these sites is much higher than the flow whole lake contribution of <0.2%. The resulting

Table 5-2: Contribution of each tributary to the *E. coli* concentration (%) at different site locations along the southern shoreline of Lake St. Clair showed in Figure 5-1.

Tributaries\Sites	**S134**	**S8**	**LP**	**S135**	**LS_WTP**	**RCM RIV**	**S136**	**SP_WTP**	**S139**	**S140**	**S142**
Pike Creek	15.9	0.5	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Puce River	2.9	2.9	0.7	0.8	0.8	0.0	0.0	0.0	0.0	0.0	0.0
Belle River	7.7	10.6	77.3	71.6	55.3	2.5	2.5	2.5	0.6	0.2	0.8
Duck Creek	3.5	4.5	6.4	7.7	14.2	1.1	0.9	0.9	0.1	0.2	0.5
Moison Creek	2.9	3.7	4.2	5.5	9.1	1.7	1.3	1.3	0.2	0.3	0.6
Ruscom River	8.3	9.8	8.3	10.6	15.0	92.4	19.9	19.9	2.9	3.4	5.6
Little Creek	1.1	1.6	0.2	0.4	0.7	0.3	9.3	9.3	3.1	2.2	3.6
Stoney Point	1.1	1.5	0.3	0.5	0.9	0.5	50.1	50.1	1.5	1.6	2.4
Major tributaries	56.6	64.8	2.5	2.7	3.7	1.5	16.0	16.0	91.6	92.1	86.7

contribution of the major tributaries is still high, ranging between 64.7% in site RCM RIV to 98.4% in site S139.

For the same sites, the contribution to the *E. coli* concentration was quite different. For six of the 11 sites located in the nearshore region of ~40 km stretch between Pike Creek and Little Creek, *E. coli* concentrations are dominated by one or more of the eight small tributaries with the combined total exceeding 80% (see Figure 5-5). For sites S135 and LS_WTP for example, while their source water is dominated by the major tributaries (77.6% and 80.9% respectively), the contribution to *E. coli* concentration is dominated by the input from Belle River (71.6% and 55.3% respectively) (see Table 5-2). Site S134 which is the nearest site to the location of Sandpoint Beach is highly influenced by *E. coli* input from major tributaries, Pike Creek, and Ruscom River with a contribution of 56.6%, 15.9% and 8.3% respectively. Belle River is much closer to the site S134, but its *E. coli* contribution is less than the Ruscom River because of the higher flow contribution of the Ruscom River (2.1%) compared to the Belle River (0.8%).

Figure 5-5 shows the map and contour plot of the mean and 90th percentile of the relative *E. coli* contribution from small tributaries combined. The entire southern shoreline is

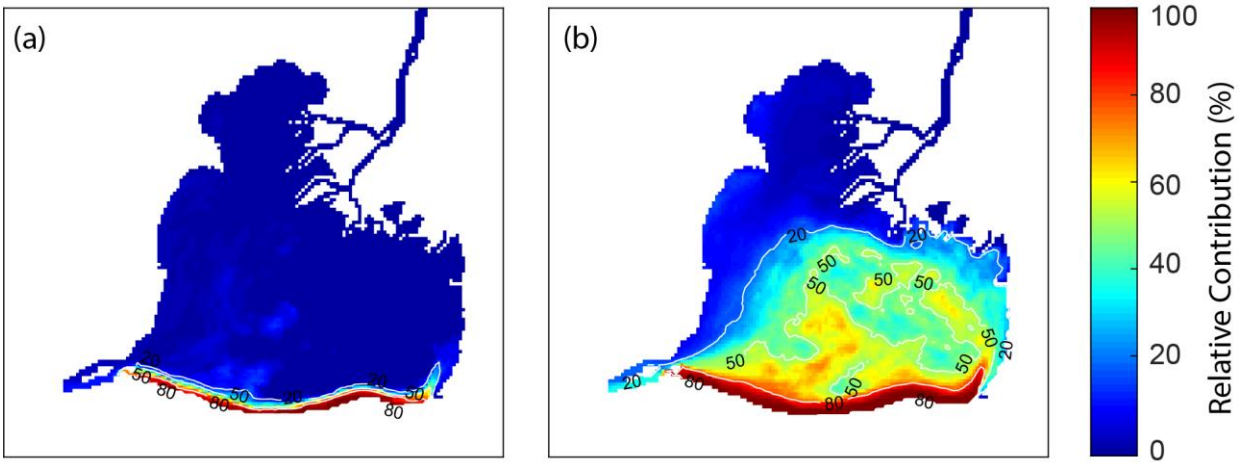

Figure 5-5: Relative contribution of all small tributaries to *E. coli* concentration (a) mean and (b) 90th percentiles

affected by more than 80% contribution from small tributaries. Areas in the middle and eastern parts of the lake can occasionally have up to 50% relative contribution from small tributaries to *E. coli* concentration (Figure 5-5b). However, the total *E. coli* concentration in these regions is very low (see Figure 5-2). To show the areal extent of influence for each tributary, areas in which the *E. coli* concentrations are reduced by 90% (1-Log reduction) and by 99% (2-Log reduction) from those in the tributary just before its confluence with the lake were identified, as illustrated in Figure 5-6. For example, sites S135 and LS_WTP are located in the area of 1-Log reduction from Belle River, and 2-Log reduction from Duck Creek, Moison Creek, Ruscom River, Stony Point, and Little Creek. That means the mean *E. coli* concentration of Belle River of 414 CFU/100 ml (Table 5-S1) is reduced to ~40 CFU/100 ml at those two sites. In a case when maximum *E. coli* concentration is input from Belle River (~ 5600 CFU/100 ml) and coincides with an east to west current in the

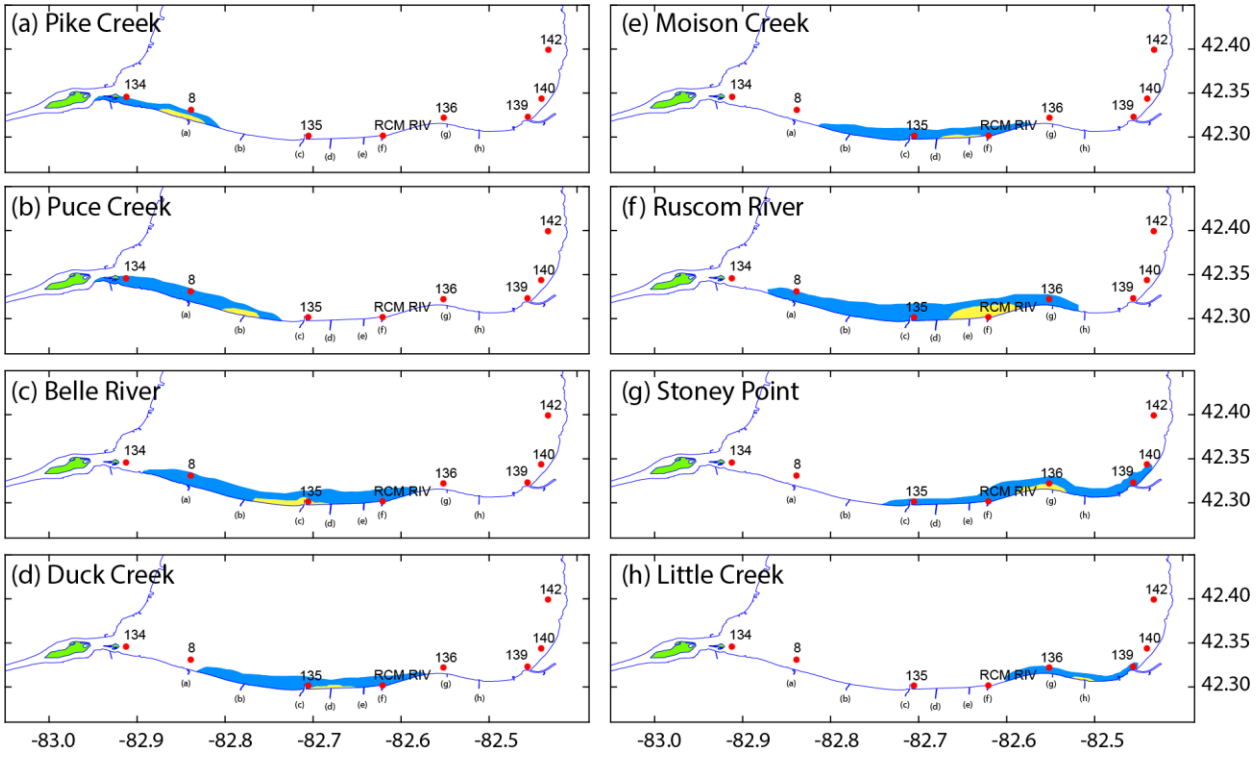

Figure 5-6: Area of *E. coli* influence: yellow are shows where the *E. coli* is reduced from the input value by or less than 1-log reduction and blue area shows the reduction less than 2-log reduction. a) Pike Creek, b) Puce Creek, c) Belle River, d) Duck Creek, e) Moison Creek, f) Ruscom River, g) Stoney Point drainage area and h) Little Creek.

southern shoreline, the tributary concentrations is still reduced to less than 56 CFU/100 ml before reaching Sandpoint Beach. Besides, for each site, whether it is affected by the other tributaries can be obtained from Figure 5-6. For instance, sites S135 and LS_WTP are affected by the discharges from Duck Creek, Moison Creek, Ruscom River, Stony Point, and Little Creek. In the case of site S134, it is only affected by inputs from Pike Creek and Puce River. Any *E. coli* entered from other tributaries will reduce by more than 2-Log reduction (more than 99% reduction) before arriving at site S134.

5.4 Conclusion

A 3D high-resolution AEM3D model was used to assess the impact on microbial water quality in Lake St. Clair from eight smaller tributaries located on its southern edge. The

combined flow contribution of these tributaries to Lake St. Clair is less than 0.2%. The impact of these tributaries on the water quality in the larger areas of the lake is expected to be minimal and the results did show that except for a small nearshore region near the southern edge of Lake St. Clair, the eight small tributaries do not impact the microbial water quality.

For sampling sites located within the study area, model simulations that included the smaller tributaries inflows were in better agreement with observations that those without. This suggests that the modelling framework used in the study can allow for the identification of areal zones where the contribution of a tributary may be significant. The framework was then used to then identify areal zones within which there is a 1- or 2-log reduction expected from each small tributary loading concentration. Depending on the location of interest, such zones can then be used to identify whether or not the location is likely to be impacted by the tributary.

For sampling sites in the nearshore region of a ~ 40 km stretch along the southern edge of the lake, while the flow contributions from the eight small tributaries were estimated in the range between 18-35%, the contribution of *E. coli* microbial pollution was estimated to be >80%. These sites include a beach (Lakeview Park West Beach) and two drinking water treatment plant intakes where the contributions from one tributary alone (Belle River) is estimated to be >50%.

The hydrodynamics in the nearshore region is expected to be more complex than the larger lake. Water quality in these regions can be further influenced by the confluence of tributaries, or the presence of structures or changes in shoreline characteristics. Water

quality changes in nearshore regions affected by such features are unlikely to be accurately predicted using the uniform grid size of 400 m used in the current modelling framework.

5.5 Acknowledgement

The authors are grateful to the Ontario Ministry of the Environment and Climate Change (MOECC) for the ADCP data and also to Windsor Essex County Health Unit (WECHU) for providing the weekly *E. coli* data. We thank Environment and Climate Change Canada (ECCC) and Canada Centre for Inland Waters for their support and special thanks to Dr. Tom Edge and his team for providing *E. coli* data for the Thames River. A special thanks to Katie Stammler from Essex Region Conservation Authority (ERCA) for all her supports during the project and providing E. coli data for small tributaries. We thank Nicole Drumm, Erin Carroll and Kelli Smith from St. Clair Region Conservation Authority and Lilly Snobelen from Chatham-Kent PUC to provide data that helped us in *E. coli* estimation for the Sydenham and St. Clair River. This research received financial support from Canada's Natural Sciences and Engineering Research Council (NSERC) Strategic Project Grant (SPG) and was performed while Mr. Madani (first author) held an Ontario Trillium Scholarship (OTS).

5.6 Supplementary Material

Supplementary data to this article can be found in Appendix 5.

5.7 References

Anderson, E.J., Schwab, D.J., Lang, G.A., 2010. Real-Time Hydraulic and Hydrodynamic Model of the St. Clair River, Lake St. Clair, Detroit River System. Journal of Hydraulic Engineering 136(8) 507-518.

Bocaniov, S.A., Scavia, D., 2018. Nutrient Loss Rates in Relation to Transport Time Scales in a Large Shallow Lake (Lake St. Clair, USA—Canada): Insights From a Three-Dimensional Model. Water Resources Research 54(6) 3825-3840.

Bonamano, S., Madonia, A., Borsellino, C., Stefani, C., Caruso, G., De Pasquale, F., Piermattei, V., Zappala, G., Marcelli, M., 2015. Modeling the dispersion of viable and

total Escherichia coli cells in the artificial semi-enclosed bathing area of Santa Marinella (Latium, Italy). Mar Pollut Bull 95(1) 141-154.

Bravo, H.R., McLellan, S.L., Klump, J.V., Hamidi, S.A., Talarczyk, D., 2017. Modeling the fecal coliform footprint in a Lake Michigan urban coastal area. Environmental Modelling & Software 95 401-419.

Brookes, J.D., Antenucci, J., Hipsey, M., Burch, M.D., Ashbolt, N.J., Ferguson, C., 2004. Fate and transport of pathogens in lakes and reservoirs. Environment International 30(5) 741-759.

Byappanahalli, M.N., Nevers, M.B., Korajkic, A., Staley, Z.R., Harwood, V.J., 2012. Enterococci in the Environment. Microbiology and Molecular Biology Reviews 76(4) 685.

Caramatti, I., Peeters, F., Hamilton, D., Hofmann, H., 2020. Modelling inter-annual and spatial variability of ice cover in a temperate lake with complex morphology. Hydrological Processes 34(3) 691-704.

Chen, W.-B., Liu, W.-C., 2017. Investigating the fate and transport of fecal coliform contamination in a tidal estuarine system using a three-dimensional model. Marine Pollution Bulletin 116(1) 365-384.

Cho, K.H., Pachepsky, Y.A., Oliver, D.M., Muirhead, R.W., Park, Y., Quilliam, R.S., Shelton, D.R., 2016. Modeling fate and transport of fecally-derived microorganisms at the watershed scale: State of the science and future opportunities. Water Research 100 38-56.

ERCA, 2015. Updated Assessment Report – Essex Region Source Protection Area, Watershed Characterization. Retrived from: https://essexregionconservation.ca/wp-content/uploads/2018/04/chapter-2-watershed-characterization.pdf.

Gao, G., Falconer, R.A., Lin, B., 2015. Modelling the fate and transport of faecal bacteria in estuarine and coastal waters. Marine Pollution Bulletin 100(1) 162-168.

Garcia-Armisen, T., Thouvenin, B., Servais, P., 2006. Modelling faecal coliforms dynamics in the Seine estuary, France. Water Science and Technology 54(3) 177-184.

Health-Canada, 2012. Guidelines for Canadian Recreational Water Quality, Third Ed., Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, (Catalogue No H129-15/2012E).

Health-Canada, 2020. Guidelines for Canadian Drinking Water Quality: Guideline Technical Document — Escherichia coli. Water and Air Quality Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario. (Catalogue No. H129-27/2020E-PDF).

Healy, D.F., Chambers, D.B., Rachol, C.M., Jodoin, R.S., 2007. Water quality of the St. Clair River, Lake St. Clair, and their U.S. tributaries, 1946-2005, Scientific Investigations Report: Reston, VA.

Hipsey, M.R., Antenucci, J.P., Brookes, J.D., 2008. A generic, process-based model of microbial pollution in aquatic systems. Water Resources Research 44(7).

Holtschlag, D.J., Koschik, J.A., 2002. A two-dimensional hydrodynamic model of the St. Clair-Detroit River waterway in the Great Lakes basin, Water-Resources Investigations Report: Lansing, MI.

Huziy, O., Sushama, L., 2017. Lake–river and lake–atmosphere interactions in a changing climate over Northeast Canada. Climate Dynamics 48(9) 3227-3246.

Ishii, S., Sadowsky, M.J., 2008. Escherichia coli in the Environment: Implications for Water Quality and Human Health. Microbes and Environments 23(2) 101-108.

Leon, L.F., Lam, D., Schertzer, W., Swayne, D., 2005. Lake and climate models linkage: a 3-D hydrodynamic contribution. Adv. Geosci. 4 57-62.

Leon, L.F., Smith, R.E.H., Hipsey, M.R., Bocaniov, S.A., Higgins, S.N., Hecky, R.E., Antenucci, J.P., Imberger, J.A., Guildford, S.J., 2011. Application of a 3D hydrodynamic–biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie. Journal of Great Lakes Research 37(1) 41-53.

Leon, L.F., Smith, R.E.H., Malkin, S.Y., Depew, D., Hipsey, M.R., Antenucci, J.P., Higgins, S.N., Hecky, R.E., Rao, R.Y., 2012. Nested 3D modeling of the spatial dynamics of nutrients and phytoplankton in a Lake Ontario nearshore zone. Journal of Great Lakes Research 38 171-183.

Liu, L., Phanikumar, M.S., Molloy, S.L., Whitman, R.L., Shively, D.A., Nevers, M.B., Schwab, D.J., Rose, J.B., 2006. Modeling the transport and inactivation of E. coli and enterococci in the near-shore region of Lake Michigan. Environmental Science & Technology 40(16) 5022-5028.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020a. Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach. Journal of Great Lakes Research.

Madani, M., Seth, R., Leon, L.F., valipour, R., McCrimmon, C., 2020b. Modelling The Fate and Transport of Microbial Contamination in Lake St. Clair: The Effect of Decay Dynamics. CSCE.

McPhedran, K., Seth, R., Bejankiwar, R., 2013. Occurrence and predictive correlations of Escherichia coli and Enterococci at Sandpoint beach (Lake St Clair), Windsor, Ontario and Holiday beach (Lake Erie), Amherstburg, Ontario. Water Quality Research Journal of Canada 48(1) 99-99.

Muirhead, R.W., Monaghan, R.M., 2012. A two reservoir model to predict Escherichia coli losses to water from pastures grazed by dairy cows. Environment International 40 8-14.

Oveisy, A., Rao, Y.R., Leon, L.F., Bocaniov, S.A., 2014. Three-dimensional winter modeling and the effects of ice cover on hydrodynamics, thermal structure and water quality in Lake Erie. Journal of Great Lakes Research 40 19-28.

Pandey, P.K., Kass, P.H., Soupir, M.L., Biswas, S., Singh, V.P., 2014. Contamination of water resources by pathogenic bacteria. AMB Express 4 51-51.

Paturi, S., Boegman, L., Bouffard, D., Rao, Y.R., 2015. Three-Dimensional Simulation of Lake Ontario North-Shore Hydrodynamics and Contaminant Transport. Journal of Hydraulic Engineering 141(3) 04014082.

Romero, J.R., Antenucci, J.P., Imberger, J., 2004. One- and three-dimensional biogeochemical simulations of two differing reservoirs. Ecological Modelling 174(1) 143-160.

Safaie, A., Wendzel, A., Ge, Z., Nevers, M.B., Whitman, R.L., Corsi, S.R., Phanikumar, M.S., 2016. Comparative Evaluation of Statistical and Mechanistic Models of Escherichia coli at Beaches in Southern Lake Michigan. Environmental Science & Technology 50(5) 2442-2449.

Silva, C.P., Marti, C.L., Imberger, J., 2014. Physical and biological controls of algal blooms in the Río de la Plata. Environmental Fluid Mechanics 14(5) 1199-1228.

Trolle, D., Hamilton, D.P., Hipsey, M.R., Bolding, K., Bruggeman, J., Mooij, W.M., Janse, J.H., Nielsen, A., Jeppesen, E., Elliott, J.A., Makler-Pick, V., Petzoldt, T., Rinke, K., Flindt, M.R., Arhonditsis, G.B., Gal, G., Bjerring, R., Tominaga, K., Hoen, J.t., Downing, A.S., Marques, D.M., Fragoso, C.R., Søndergaard, M., Hanson, P.C., 2012. A community-based framework for aquatic ecosystem models. Hydrobiologia 683(1) 25-34.

USEPA, 2012. Recreational Water Quality Criteria, In: 820-F-12-058, E. (Ed.): Office of Water, Washington, DC.

Valipour, R., León, L.F., Depew, D., Dove, A., Rao, Y.R., 2016. High-resolution modeling for development of nearshore ecosystem objectives in eastern Lake Erie. Journal of Great Lakes Research 42(6) 1241-1251.

WHO, 2003. Emerging issues in water and infectious disease. World Health Organization: Geneva.

WHO, 2017. Guidelines for drinking-water quality: fourth edition incorporating the first addendum. World Health Organization: Geneva.

CHAPTER 6 NESTED 3D HIGH RESOLUTION MODELLING OF THE MICROBIAL WATER QUALITY IN NEARSHORE REGION OF LAKE ST. CLAIR

6.1 Introduction

Understanding the fate and transport of microbial contamination within the zone of influence of the beaches and drinking water intakes is critical for managers to effectively reduce health risks. Microbial water quality in nearshore waters, being the interface between land and coastal waters and the primary zone of contact during recreational activities, has direct implications for human health (Health-Canada, 2012; Nevers and Whitman, 2005). The nearshore zone has been defined in various ways in the literature, which can be as a simple descriptive feature such as depth or distance from the shoreline, or depending on specific processes under consideration (Huang et al., 2019; Leon et al., 2012; Makarewicz and Howell, 2012; Valipour et al., 2018; Warren et al., 2018). As far as microbial pollution is concerned, the nearshore region can be defined as a zone extending offshore from the shoreline that is exposed to microbial loading from the local sources.

Microbial water quality in the nearshore region is affected by the open lake waters, which is impacted by the inflows of the major rivers and the hydrodynamics in the lake. It can also be disproportionately affected by relatively small discharges and tributaries in closer proximity of the region in question. The concentrations of microbial pollutions can be higher near tributary mouths and discharge sites until they are diluted by mixing with the nearshore waters. They remain in the nearshore for extended periods before ultimately diluted with larger volumes of less impacted offshore waters (Edsall and Charlton, 1997). Many such small tributaries are usually ignored in the lake-wide impact studies. To mitigate and manage issues posed by elevated levels of microbial contamination in the

nearshore regions, it is critical to identify the primary water sources of contamination entering the lake environment.

Lake St. Clair is a freshwater waterbody and part of the international boundary between the United States and Canada. It is an important navigational, recreational and drinking water resource in the Great Lakes basin. Microbial water quality of Lake St. Clair is important for more than four million people in the region that rely on the lake for drinking and recreational purposes. Modelling studies have been conducted to better simulate and understand hydrodynamics and water quality in the lake (Anderson and Schwab, 2011; Anderson et al., 2010; Bocaniov and Scavia, 2018; Madani et al., 2020a). In the recent study by Madani et al. (2020a), a structured grid AEM3D model was developed and applied to investigate the impact of four major tributaries (Thames, Sydenham, St. Clair and Clinton River) on microbial water quality in Lake St. Clair and in particular at Sandpoint Beach. Lake-wide microbial water quality of Lake St. Clair was shown to be dominated by major tributaries inputs especially from St. Clair River but it had a lower impact on the southern shoreline which includes at least two popular beaches and two drinking water intakes for the town of Lakeshore.

The study also identified unique hydrodynamics that resulted in the disproportionate impact from the Thames River on the southern edge of the lake, near the Windsor Essex Region in Canada. Microbial pollution from non-point sources and urban development in the Windsor Essex County find their way into the southern edge of Lake St. Clair through eight smaller tributaries. The impact of these tributaries has been included in previous Lake-wide St. Clair water quality studies (Bocaniov and Scavia, 2018; Madani et al., 2020b). Using structured grid AEM3D model, Madani et al. (2020b) estimated that the

combination of small tributaries contributed to more than 80% of *E. coli* concentration at the location of Lakeshore water treatment plants (WTP) and Lakeview Park West Beach (LP Beach) located within the nearshore region of the southern edge. Belle River alone was estimated to contribute to >50% of the *E. coli* concentration at LP Beach and Lakeshore WTP. Complex bathymetry and geomorphic layout have a noticeable effect on the hydrodynamic (flow circulation) and contamination transport, thus should be represented with an appropriate computational mesh (Ganju et al., 2016; Ge et al., 2012). The grid sizes of 400 m used in the study by (Madani et al., 2020b) is insufficient to adequately model the hydrodynamics and water quality impacts in the nearshore region from small tributaries such as Belle River with mouth widths <50 m (a grid size of 5–10 m may be desirable). Further reducing the grid size in this structured grid model is not practical due to computational expense. Also, it is difficult for the structured grid to correctly resolve the complex geometry in the nearshore region.

A model with an unstructured grid can be used to accurately resolve the bathymetry in complex shoreline with high resolution while using coarser cells elsewhere. Using unstructured grids is preferred when it is needed to resolve the domain with complex geometries such as irregular coastlines, islands, barriers or inlets that need to be correctly represented. In this way, increasing the overall resolution is avoided and hence has a computational advantage over regular grids (Chen et al., 2003; Martyr-Koller et al., 2017). Still, that level of refinement (5-10 m) will be computationally very demanding in larger lakes such as Lake St. Clair. Powerful computational resources and the use of parallel computation techniques are required. A possible way of overcoming this deficiency is to

developing a nested model for the region of interest whilst obtaining forcing conditions from a lower-resolution larger-scale lake-wide model.

Lakeview Park West Beach in the Town of Lakeshore, Ontario, Canada is located right next to Belle River on the southern edge of Lake St. Clair and separated from it by a 25 m steel breakwall jetty. This breakwall was replaced by a new 150 m breakwall in 2018. Other features nearby include another 600 m breakwall at the Lakeview Park Marina, and a drinking water crib intake for the Lakeshore Water Treatment Plant (WTP) that extends about 1.2 km into Lake St. Clair. The first main objective of this study is to develop a lake-wide unstructured grid hydrodynamic and microbial (*E. coli*) water quality model using TUFLOW-FV coupled with AED2+ framework. This lake-wide model is then tested against ADCP and water temperature data for Lake St. Clair during the summer of 2016, and also compared with hydrodynamic simulations obtained with the previously developed structured grid hydrodynamic model using AEM3D (Madani et al., 2020a). The second objective is to develop a higher resolution nested model in the area around the Belle River, including Lakeview Beach and Lakeshore WTP, that uses the lake-wide model output (current and scalar concentration) as the boundary condition. Microbial water quality simulations at LP Beach and Lakeshore WTP obtained using the nested model are compared with those using the lake-wide model and available monitoring data. Microbial simulations were also carried out using the nested model including the new 150 m replacement to the old 25 m jetty, built in 2018, to simulate the expected impact on the microbial water quality at Lakeview Beach and Lakeshore WTP.

6.2 Data and Methods

6.2.1 Study Site:

The geographic extent of this study is Lake St. Clair and a high-resolution area with a radius of 2 km with center at the mouth of Belle River. Lakeview Park, located at the west side of the Belle River mouth, includes a beach with pavilion, splash pad, playground, walking trails, marina and restaurant and it is very popular in the region. A schematic drawing showing the beach and relevant features are presented in Figure 6-1. Lakeview Park West Beach (42°17'51.10"N, 82°42'41.61"W) (hereafter LP Beach) is located in the mouth of the Belle River is a small (150 m long) but still a popular beach in the region especially because of the Belle River Marina at the adjacent. The marina is protected by 600 m breakwaters on the east and west side. Recently a jetty was built that extends into Lake St. Clair from the LP Beach to replace the failing 25-metre steel break wall. According to the Windsor-Essex County Health Unit (WECHU) report and field measurements both beaches are frequently posted as unsafe for water recreation for the summer seasons due to incidents involving high bacterial counts. John George Water Treatment Plant that is located in Belle River supplies water to the north-western portion of the Town of Lakeshore. The intake is located at 1050 m from the shoreline (See Figure 6-1).

6.2.2 Bathymetry, Forcing Data and Flow and *E. coli* Boundary Conditions:

The bathymetry, from the National Geophysical Data Center (https://www.ngdc.noaa.gov/), has depths referenced to a generic datum of 176.784 meters (580 feet). It is a gridded scatter data with nodes interval of ~70 m by ~90 m in easting and northing direction. The lake flexible mesh consists of 21,019 combinations of triangular and quadrilateral cells and

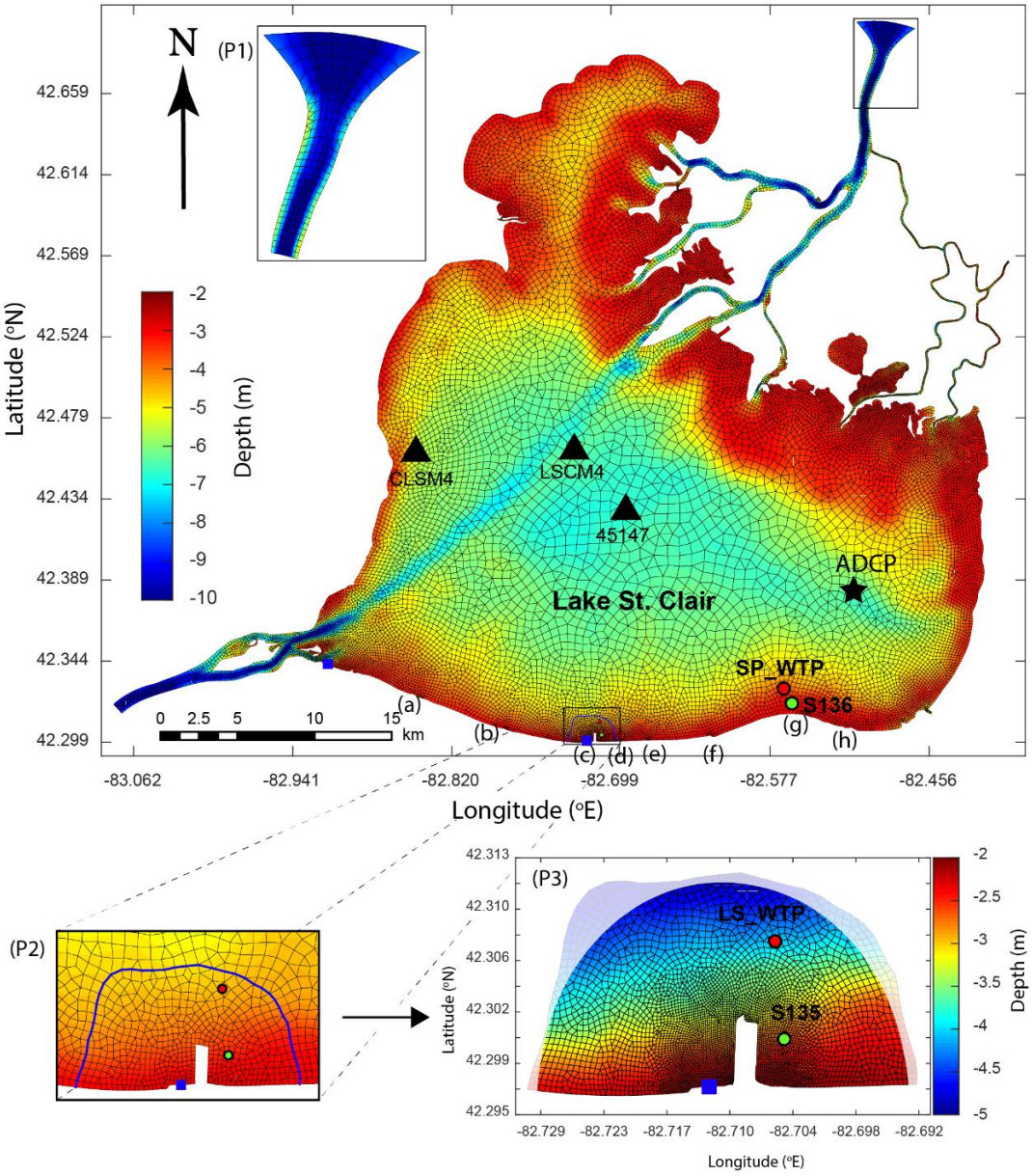

Figure 6-1: Map of Lake St. Clair and its bathymetry. Triangle symbols show the locations of the moorings (CLSM4 (*42.471 N 82.877 W*), LSCM4 (*42.465 N 82.755 W*) and 45147 (*42.430 N 82.680 W*)). Location of Sandpoint Beach and Lakeview Park Beach are identified by blue squares, intakes by the red circle. Bathymetric data are in meters and all major tributaries are labelled on the map. Small tributaries are a) Pike Creek, b) Pike Creek, c) Belle River, d) Duck Creek, e) Moison Creek, f) Ruscom River, g) Stoney Point drainage area and h) Little Creek. P1) shows the extension of the model inlet at the St. Clair River, P2 and P3 shows the bathymetry of an area around Belle River and the location of Beach and water intakes.

was generated in the Surface Modeling System (SMS) mesh generation software package

(SMS community version 13.0, AQUAVEO) developed by the Engineering Graphics Laboratory at Brigham Young University. The model was configured with a variable mesh ranging from <60m to ~600m, thereby capturing high spatial resolution in the littoral regions and around complex coastlines, while using a coarser resolution lake wide. To resolve the geometry without the great computational expense, an unstructured mesh is employed to simulate the fluid dynamics in the Belle River mouth. Nested mesh with a resolution between 5-10 meters at a zone stretching 4 km alongshore and roughly 2 km onshore-offshore around Belle River mouth was established as the computational domain to perform the simulations with the higher spatial resolution model.

Similar to the work done by Anderson et al. (2010) artificial cells and bathymetry were created at the mouth of St. Clair River. This extension is implemented to stabilize the high flow condition at the boundary that eliminates flow anomalies that can occur near the boundary of unstructured grids. This is not expected to affect the results in the realistic grid geometry. To simulate events in the high-resolution model, the lake-wide setup is first used to generate output on the flow, water temperature, scalar and ecological values which are then captured for input at the boundaries of the nested domain. A similar approach was successfully employed elsewhere (Acosta et al., 2015; Leon et al., 2012; Rao and Sheng, 2008; Spillman et al., 2008).

Meteorological forcing data sets include wind direction, wind speed, air temperature, humidity, atmospheric pressure, rainfall, and solar radiation. Data is from the nearest stations (Windsor A. Station ID=4716; Windsor Riverside Station ID=4715) and moored surface buoys (Buoy (West Erie) 45005 (41.677 N 82.398 W); Buoy 45147 (42.430N 82.680W); LSCM4 (42.465N 82.755W); CLSM4 (42.471N 82.877W)). Buoy 45147,

which is maintained by Environment and Climate Change Canada (ECCC), was used to fill out missing wind direction and wind speed data and for model validation of water temperature. Solar radiation data were obtained from the Canadian Regional Climate Model (CRCM5) provided by the University of Quebec at Montreal (UQAM) (Huziy and Sushama, 2017). (See Table S1 for more information).

Flow and temperature data of the four major tributaries to the lake (Thames, Sydenham, St. Clair, and Clinton Rivers) and the outflow (Detroit River) were obtained from the nearest gauged stations as described in the previous paper (Madani et al., 2020a) for the year 2016. *E. coli* concentration for major tributaries was estimated using the hydrological scaled method described in our previous work (Madani et al., 2020a). Data for Thames River at the mouth of the river (Site T0) is used to get the boundary condition for Thames River. In the case of small tributaries (Pike Creek, Puce River, Belle River Duck Creek, Moison Creek Ruscom River and Stoney Point drainage area and Little Creek) the observed flow is only available for Ruscom River at the station (02GH002). Flow for other tributaries was obtained from watershed modelling using SWAT and results was calibrated using observed flow of Ruscom River. *E. coli* data for small tributaries are limited to few grab samples collected by ERCA and PWQMN. A MATLAB code was written to correlate the flow with the *E. coli* to obtain time series of *E. coli* concentration for each tributaries.

6.2.3 Modelling Framework:

6.2.3.1 Hydrodynamics and Ecological Model

In the current study, the 2-Dimensional Unsteady Flow Finite Volume (TUFLOW-FV) is used. The model solves the Non-Linear Shallow Water Equations (NLSWE) numerically using the finite volume method integrating primitive equations over the unstructured

triangular and quadrilateral grid mesh with 10 vertical sigma layers for a 3D simulation (BMT_Pty._Ltd., 2019). The NLSWE is a system of equations describing the conservation of fluid mass/volume and momentum in an incompressible fluid, under the hydrostatic pressure and Boussinesq assumptions. This approach combines the advantage of an unstructured grid for shoreline fitting and the flexibility of local mesh refinements (similar to finite element methods), as well as numerical efficiency and code simplicity (similar to finite difference methods). Although this study focuses on the Belle River nearshore (Figure 6-1), TUFLOW-FV is configured to simulate the lake-wide physical dynamics of Lake St. Clair, thus providing a reliable representation of large-scale background circulation and the role of remote (onshore) forcing in driving nearshore water movement. The nested approach was adopted by extracting information from the coarser grid to be used as boundary conditions in finer mesh resolution. This configuration avoids the impact of setting an artificial numerical boundary condition for our target region. The advantage of an unstructured grid is that model resolution varies around 800-1000 m (coarse) in the open lake to 50–100 m (finer) in the near vicinity of Belle River (the targeted nearshore), affording a high degree of resolution across the 4 km study site and adequately resolving the geographic complexity and coastal hydrodynamic conditions of that system (Figure 6-1).

Aquatic Eco-Dynamics (AED2+ V. 1.3), an open-source community-driven library of model components. AED2 processes include the cycling of carbon, nitrogen and phosphorus, and other relevant components such as oxygen, and can simulate organisms including different functional groups of phytoplankton and zooplankton, and organic matter (Hipsey et al., 2019). The coupled framework of TUFLOW-FV and AED2 is

recently applied successfully to simulate hydrodynamics and water quality in several studies (Liu et al., 2020; Zhang et al., 2017). AED2+ is the advanced version of AED2 that contains 12 modules including the pathogen module and was used in the current study. Pathogen module implemented in AED2+ can be used to simulate organisms such as protozoa, bacteria and viruses. For more details, readers are referred to (Hipsey et al., 2008). The AED2+ code is freely available for possible users at https://aed.see.uwa.edu.au/research/models/AED/download.html

6.2.3.2 Decay Formulation

The governing equation that explains the transport of the *E. coli* has the following general form:

$$[1] \qquad \frac{dC}{dt} + \frac{\partial}{\partial x_j}\left(CU_j\right) = \frac{\partial}{\partial x_j}\left(\kappa_j \frac{\partial C}{\partial x_j}\right) + C_{in} - C_{out} + K\,C$$

Where C denotes the *E. coli* concentration ($orgs/m^{-3}$), t is time, x_j is the distance in the j^{th} dimension (m), U_j is the velocity in the j^{th} dimension (m/s^{-1}), κ_j is eddy-diffusivity, C_{in} and C_{out} are the inflow and outflow fluxes ($orgs/(m^{-3}.s)$) and K is the overall decay rate. Within the TUFLOW-FV coupled with AED2+ modelling framework, the unsteady, advection and diffusion terms on the right hand side of Eq. 1 are solved via the finite volume scalar transport routines with TUFLOW-FV and the fourth and fifth terms are simulated by the aed2_pathogens module of the AED2+ aquatic ecological modelling library.

In the AED2+ pathogen module, the growth and predation terms suggested by (Hipsey et al., 2008) are not considered directly since the likelihood of growth is small (Toze et al.,

2012) and predation/grazing is factored into the die-off rate. Hence, the overall decay rate can be formulated using the simplest approach assuming additive effects:

[2] $K = K_T + K_l$

where K_T is the natural mortality or die-off rate due to water temperature and can be expressed as the Arrhenius expression:

[3] $K_T = k_{d20}\, \theta^{T-20}$

where T is the temperature ($°C$), k_{d20} is the dark death rate at 20 $°C$ in freshwater and θ controls the sensitivity of K_T to water temperature change.

In Eq. 2, K_l is total die-off due to exposure to sunlight with different bandwidth. It takes the form:

[4] $K_l = \sum_{b=1}^{N_B} \varphi\, k_b\, f_b\, I_0 \left(\frac{1- e^{-\eta_b \Delta z}}{e^{-\eta_b \Delta z}}\right)$

where N_B is the number of discrete solar bandwidths to be modelled which here is 3 (visible, UV_A and UV-B), b is the bandwidth class [1, 2, ..., NB], k_b is the freshwater inactivation rate coefficient for exposure to the b^{th} class ($m^2\, MJ^{-1}$), φ is a constant to convert units from seconds to days and J to MJ (=8.64×10^{-2}). In Eq. 4, Δz is the depth of the computational cell and η_b is the extinction coefficient for each bandwidth region which governs how incident light is attenuated within the water column according to the Beer Law.

A detailed description of the formula for each of the parameters of decay and growth rate is presented by Hipsey et al. (2008).

6.2.3.3 Metrics to Measure Model Performance

The model output and ADCP data were averaged in a window of 12 hours and compared using r -correlation of determination, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) and Normalized RMSE (NRMSE) metrics. Besides, to quantitatively compare the observed and modelled currents, Fourier norms (F_n) was calculated according to (Beletsky and Schwab, 2001; Huang et al., 2010). F_n represents the uncertainty in the modelled currents relative to the variance in the observed currents from ADCP. The smaller the F_n better the model results fit the observations. The calculation involves the average difference between the V component (north-south) and the U component (west-east) of velocity as follows:

$$F_n = \frac{\|v_o, v_m\|}{\|v_o, 0\|} \quad and \quad \|v_o, v_m\| = \left(\frac{1}{n} \sum_{t=1}^{n} |v_o - v_m|^2 \right)^{0.5}$$

Typical F_n values of 0.4-1.2 are reported in the Great Lakes (Beletsky et al., 2006) and for Lake Michigan yielded between 0.75-1.01 (Beletsky and Schwab, 2001). F_n scores below 0.85 are considered to be very good. The other measure of model accuracy used in this exercise was the Root Mean Square Error (RMSE) for each vector component.

6.3 Results and Discussion

6.3.1 Lake Hydrodynamics

The hydrodynamic model was tested using measured data, which included: (1) water temperature and (2) depth-averaged flow velocities measured using one Acoustic Doppler Current Profiler (ADCPs) devices deployed at (42.3771 N, 82.5025 W). Here, both results are taken at the ADCP location and further comparison of water temperature at the location

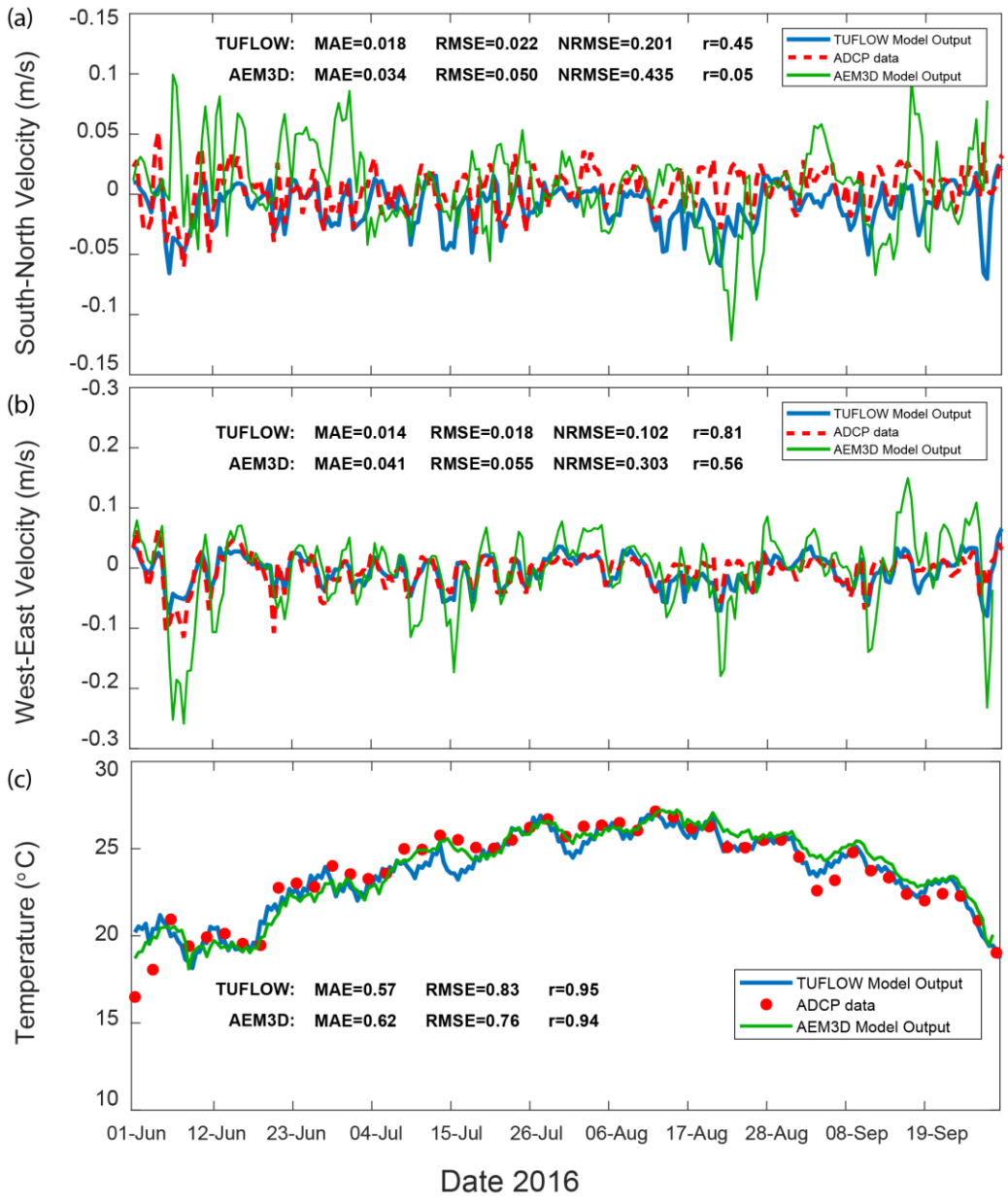

Figure 6-2: Comparison of simulated a) South-North and b) West-East velocity components and c) water temperature from TUFLOW-FV and AEM3D model with the ADCP data.

of buoy 45147 is presented in Appendix 3-3. The model predicted circulation pattern and currents in Lake St. Clair was compared with previous studies for the lake, and with the results obtained from AEM3D hydrodynamic presented in a recent study by the authors (Madani et al., 2020a). The time series and metrics are presented in Figure 6-2. Model-

simulated current conditions using TUFLOW-FV agree well with ADCP measurements for both south-north and west-east flow components with NRMSE 0.2 and 0.1 respectively. It showed better agreement compared with the results obtained by AEM3D for the same simulation period. From the ADCP data, the south-north component varied between −5 $cm.s^{-1}$ and 5 $cm.s^{-1}$ and was weaker than the west-east flow (−10 $cm.s^{-1}$ and 8 $cm.s^{-1}$). TUFLOW-FV simulates both components in the same range as ADCP data however, results from the AEM3D show a higher range. Both models can simulate low and high velocity and its rapid changes in the west-east direction better than the south-north component. Simulated water temperature shows excellent agreement with the observed water temperature for both models with RMSE 0.83 and 0.76 ^{o}C and a correlation coefficient of 0.95 and 0.94 for TUFLOW-FV and AEM3D respectively. F_n value of 0.8 was obtained which is in the acceptable range of 0.75 -1.01 that is reported in the literature (Beletsky and Schwab, 2001).

6.3.2 Microbial Water Quality Simulations

Using the lake-wide and nested models, *E. coli* concentration was simulated and time series of the results at LP Beach and Lakeshore WTP are compared against observations in Figure 6-3. Figure 6-3a shows the predicted *E. coli* concentration at the beach using the lake-wide model is significantly higher (up to a factor of 4) than results from the high resolution nested model. That may be because the ~50 m grid size in the lake-wide model at LB beach is too large to simulate the effect of the 25 m small breakwall that separates the Belle River from the LP Beach. (See Appendix 6-1 for high-resolution images from the beach). However, using the nested model the grid size in the same area is about 5 m, which is fine

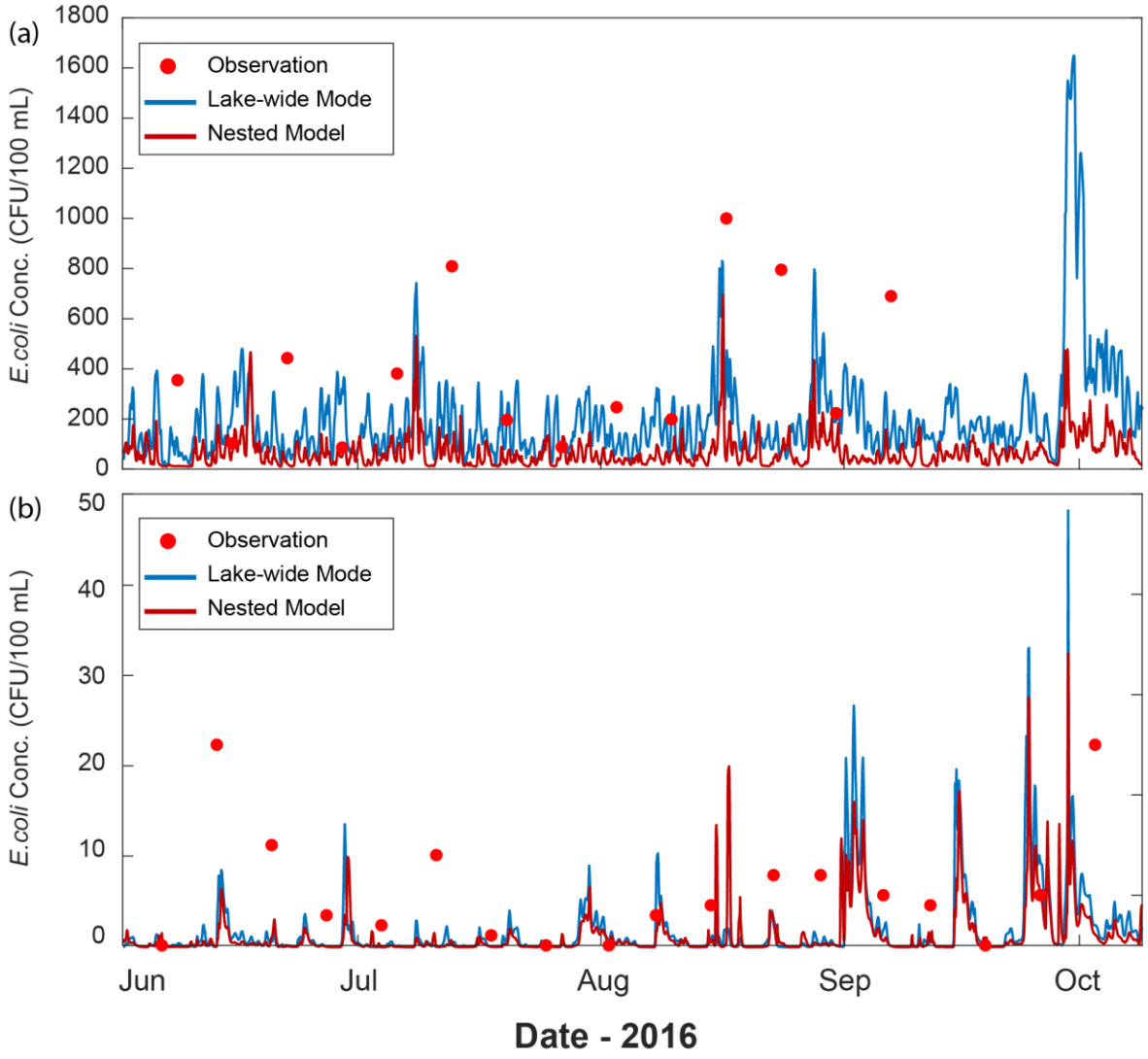

Figure 6-3: Time series comparison of the lake-wide model with the nested model at a) LP Beach b) Lakeshore water treatment plant intake location.

enough to resolve the complex feature of the beach and the breakwall. Figure 6-3a also shows that at lower *E. coli* values (<300 CFU/100 ml), the nested model results seem to be in better agreement with the monitored data, while lake-wide model results are often too high. For observed values of >400 CFU/100 ml, the predicted values are much lower than observed for both the models.

Examining the weather data shows that these high *E. coli* events occur during a rainfall event or shortly thereafter. As explained earlier in Section 6.2.2, very limited monitoring data is available to define the flow and *E. coli* loading input from Belle River. Improper representation of the flow and *E. coli* loading may be responsible for the difference between the observed and predicted values for the high values. Model comparison at the location of Lakeshore water treatment plant intake is presented in Figure 6-3b. Results show that the trend from the lake-wide model and nested model is similar, with some differences observed between the two. Most of the time the simulated *E. coli* concentration from the lake-wide model is higher than the nested model but still within a factor of two. Occasionally and for short durations, a unique combination of *E. coli* inputs and wind events (strong 30-45 km/h winds from the westerly direction) resulted in spikes in *E. coli* concentrations observed in the nested model but not in the lake-wide model. Such events were observed in mid-August and late September. Monitoring data obtained from the Lakeshore WTP show that most of the time the observed *E. coli* concentrations are <10 CFU/100 ml, which are simulated well by the nested model. On two occasions the observed *E. coli* concentration was 20 CFU/100 ml. While spikes in predicted concentrations by the nested model were seen on both occasions but were <10 CFU/100 ml. On a few occasions, the model predicted spikes in *E. coli* concentrations ranging between 10 – 30 CFU/100 ml. However, at those times, no monitoring data is available.

6.3.3 Contribution of Belle River to *E. coli* Concentration

To estimate the extent of the impact of the Belle River, its contribution to *E. coli* concentration was calculated. Figure 6-4a shows the contour plot of the mean contribution of the Belle River to *E. coli* concentration in the second scenario. Considering *E. coli*

coming from all the small and major tributaries, on an average about 90% of the microbial pollution at LP Beach from external sources is predicted to be from Belle River. The relative contribution of the Belle River on average at Lakeshore WTP is much lower at around 20%. Still however, the time series presented in Figure 6-4b shows that there are many instances in which the relative contribution of Belle River relative is estimated to be

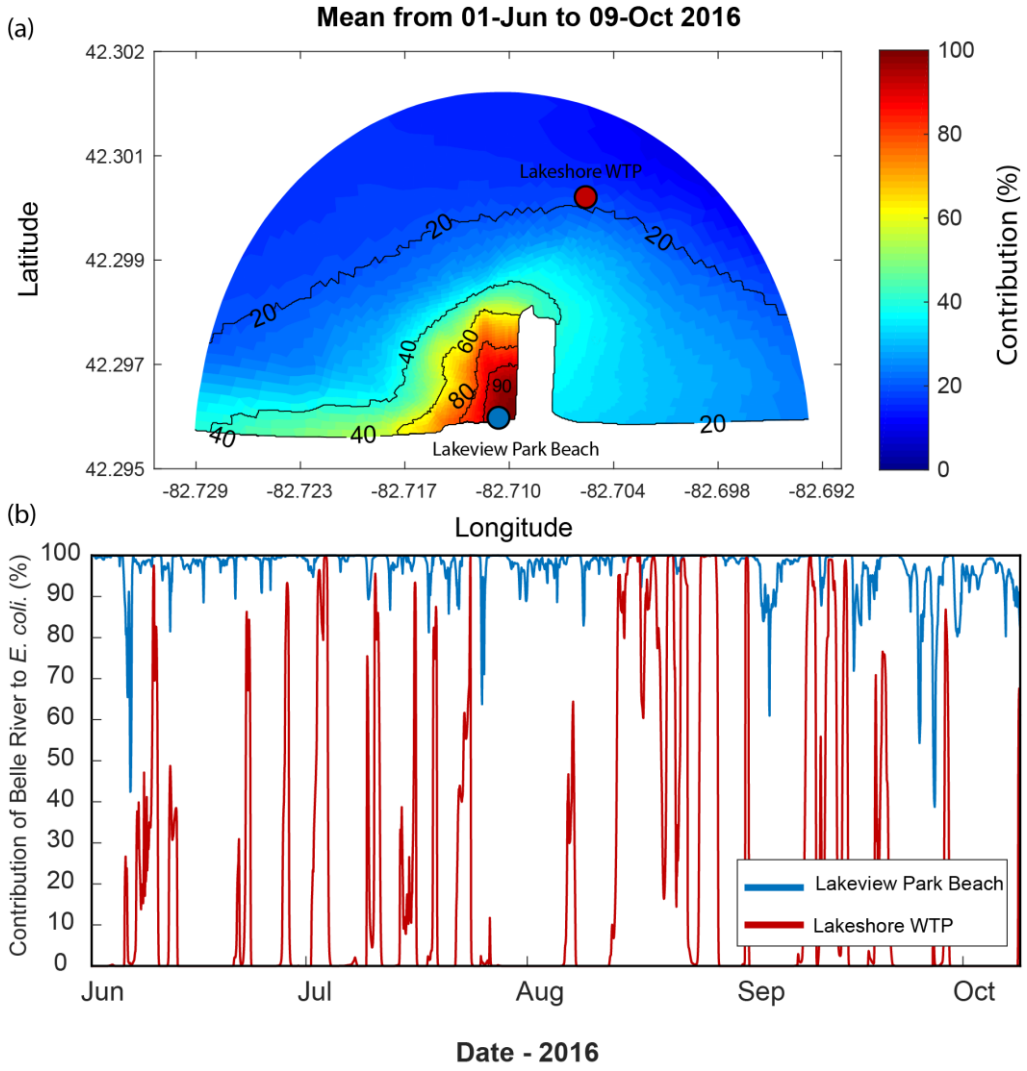

Figure 6-4: Contribution of the Belle River to E. coli concentration. a) shows the contours of the mean contribution of Belle River during the simulation period and b) illustrates the time series of the Belle River contribution at LP beach and Lakeshore WTP intake.

between 80-100%, responsible for the spikes seen in Figure 6-3b. However even in those instances, the maximum predicted *E. coli* concentration was <30 CFU/100 ml.

6.3.4 Effect of Nearshore Features

As mentioned previously, nearshore structures close to LP Beach and Lakeshore WTP in 2016 included a 600 m marina breakwall, and a 25 m jetty breakwall separating Belle River from the adjacent LP Beach. The smaller jetty was replaced by a new 150 m jetty in 2018. Breakwaters are one of the features on the beach that promoting the non-uniform distribution of *E. coli* concentration (USEPA, 2010). The impacts of these features on the temporal variation in microbial water quality at LP Beach and Lakeshore WTP were investigated using the nested model. The effects of conditions in 2016 are compared with no structures (Scenario 1), and replacement of old jetty by a new one in 2018 (Scenario 2). Figure 6-5a shows that at LP Beach, the structures present in 2016 may have at times caused a marginal increase in the maximum *E. coli* concentrations, as compared to those expected without these structures (Scenario 1). However, more instances of *E. coli* concentrations exceeding 200 CFU/100 ml could have resulted under Scenario 1. The construction of the new jetty is expected to result in a dramatic reduction in the contributions of Belle River and the resulting *E. coli* concentrations at LP Beach from external sources. At Lakeshore WTP, the results presented in Figure 6-5b show that most of the time the predicted *E. coli* concentrations were <10 CFU/100 ml and there was not much difference between the predicted microbial water quality under 2016 conditions as compared to the two scenarios. At times, the structures in 2016 and the new jetty in 2018 could result in a marginal increase in *E. coli* concentrations at Lakeshore WTP. The

predicted maximum *E. coli* concentration without any structures (Scenario 1) was 15 CFU/100 ml, as compared to 35 and 40 CFU/100 ml.

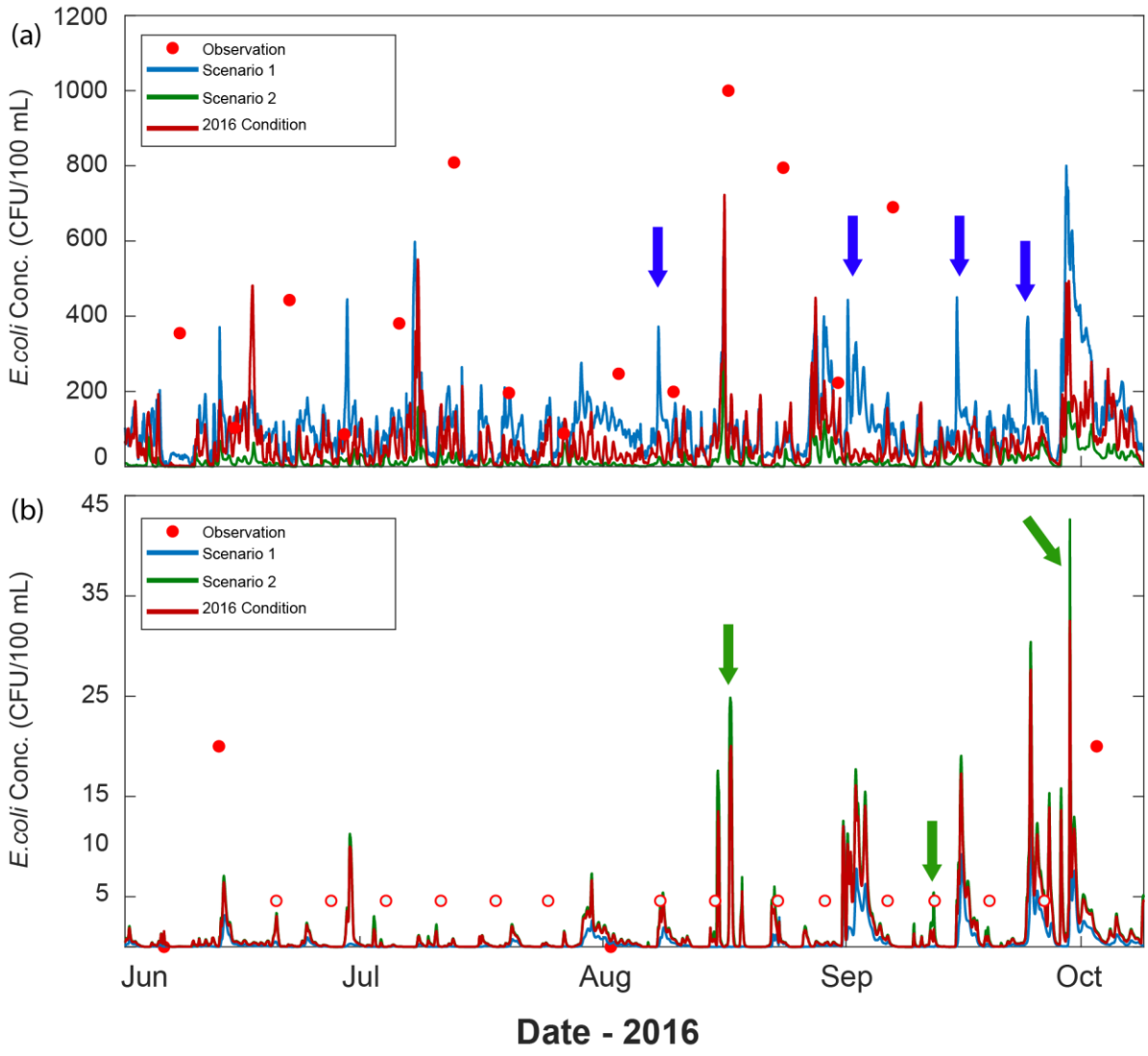

Figure 6-5: Time series of the simulate E. coli concentration compared with the observed values at a) LP Beach b) Lakeshore water treatment plant intake location. Blue arrows indicate some of the instances that peaks are observed in Scenario 1 while low values are predicted in other cases. Green arrows in b) show some instances that the installation of new jetty results in increasing the E. coli concentration. Hollow circles in b) show the measured E. coli <10 CFU/100 ml which plotted as 5 CFU/100 ml for illustration.

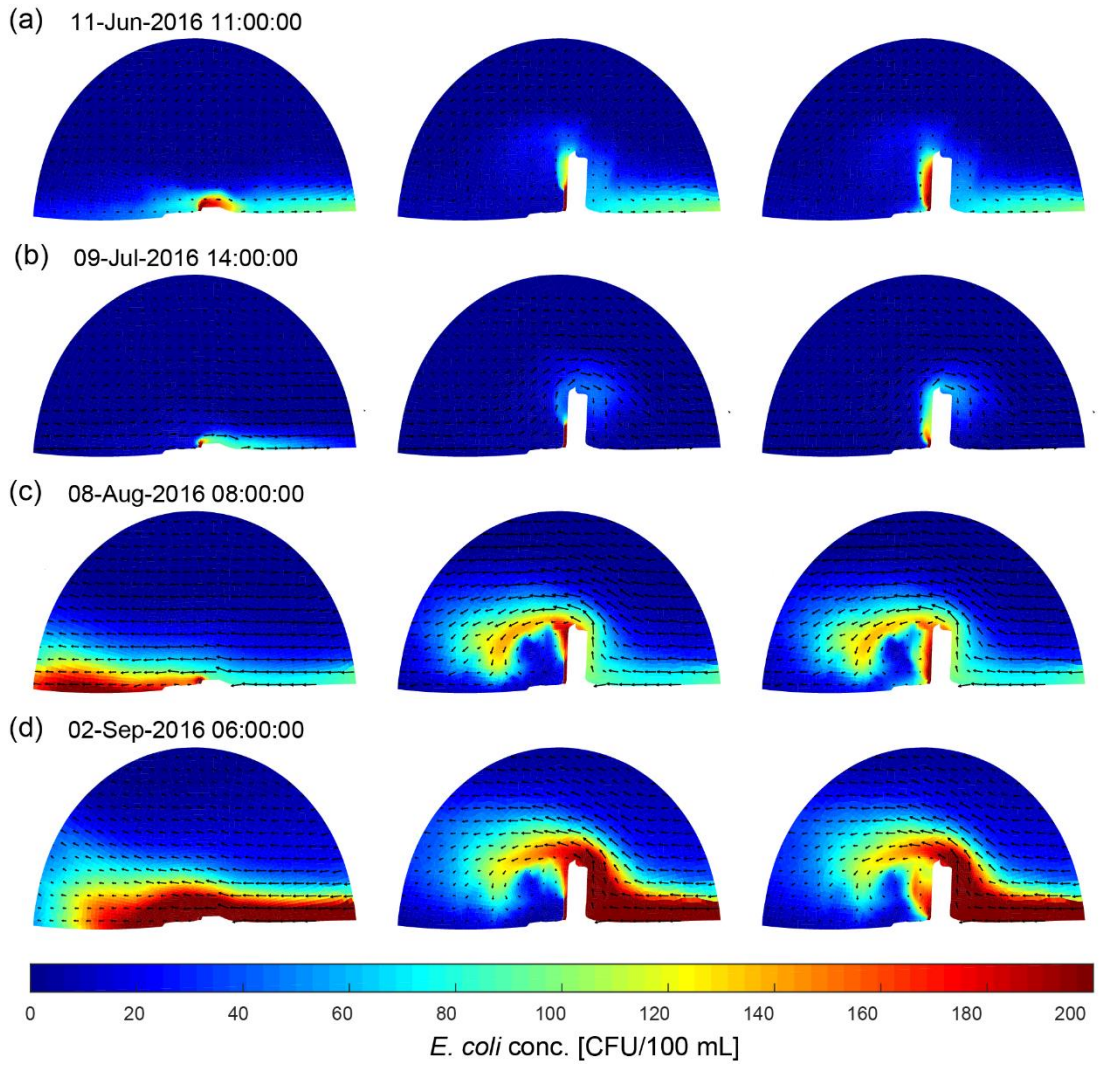

Figure 6-6: comparison of flow and circulation pattern and E. coli concentration of different scenarios at select times during the simulation period. left) no structures (Scenario 1); middle) with new jetty built in 2018 (Scenario 2), and right) 2016 Conditions (Marina and 25 m jetty)

To show how the structures present in 2016 and Scenarios 1 and 2 could have resulted in the differences seen in temporal variations in Figure 6-5, the nested model results from some days with particular features as having strong east-west or west-east current are presented in Figure 6-6. The main forces causing circulation in Lake St. Clair's southern region are currents driven by wind stress and spatial gradients in density and temperature rather than hydraulic flows, which is more dominant in the northwest part of the lake

affected by the St. Clair River. Over a timeframe, e.g. couple of days or a month, tributary plumes distribute their constituents across the nearshore region in a manner that may be likened to a stationary source.

On the shores near the Belle River, the spatial extent of the mixing gradients in the lake will be limited except possibly during high flow conditions due to the dilution with the vast volume of Lake St. Clair. The occurrence of elevated levels of *E. coli* attributable to inputs forms the Belle River to the lake. Figure 6-6 shows that this impact is highly localized and sporadic in time. In the case where there is a strong east-west current, the beach areas are likely impacted by mixing gradients from Belle River input and diluted water reached to the beach as a result of gyre formation (See Figure 6-6 c and d). The occurrence of the high *E. coli* concentration likely coincides with the occasions that the inflow is high, water quality in the tributary is poor, and when wind conditions push the mixing gradient against the shores of the beach.

To show the variability in predicted *E. coli* concentrations going into the lake, mean *E. coli* concentration profile during the simulation period along transects that contains the location of Lakeshore water treatment intakes and at the LP Beach were calculated, and the results are presented in Figure 6-7. At the transect of Lakeshore WTP, Figure 6-7a shows that *E. coli* concentration declines exponentially in the direction away from the shoreline and the mouth of the Belle River. While there is some impact of the structures up to about 800 m from shore, hardly any impact is seen at around 1000 m where the Lakeshore WTP intake is located.

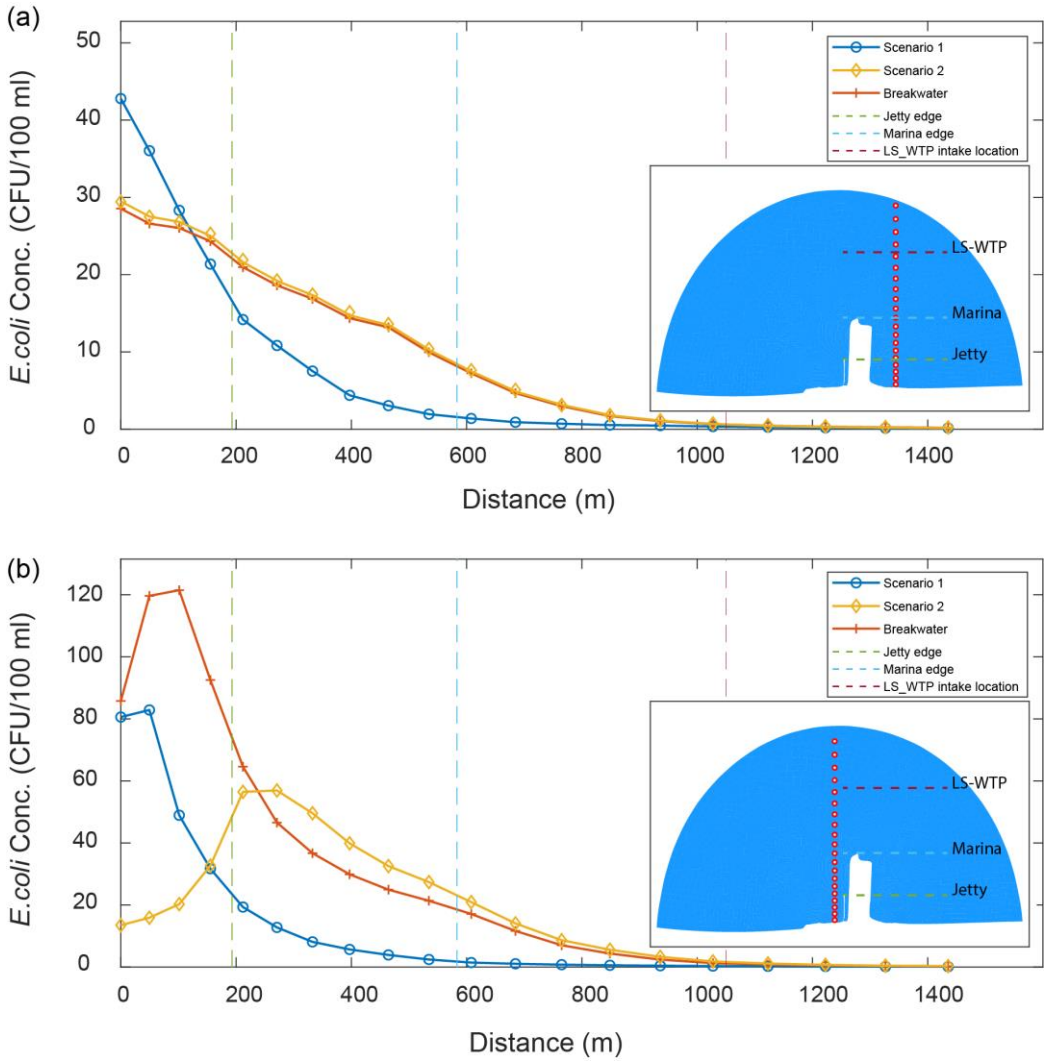

Figure 6-7: E. coli concentration profile vs. distance from the shoreline a) at the location of Lakeshore water treatment intakes and b) at LP Beach.

Results presented in Figure 6-7b at the transect of LP Beach show that the structures present in 2016 may have resulted in some increase in *E. coli* concentrations at the beach as compared to no structures, due to some "dam" effect of Belle River microbial pollution by the marina. However, the construction of the new jetty in 2018 is expected to significantly reduce microbial pollution at LP Beach.

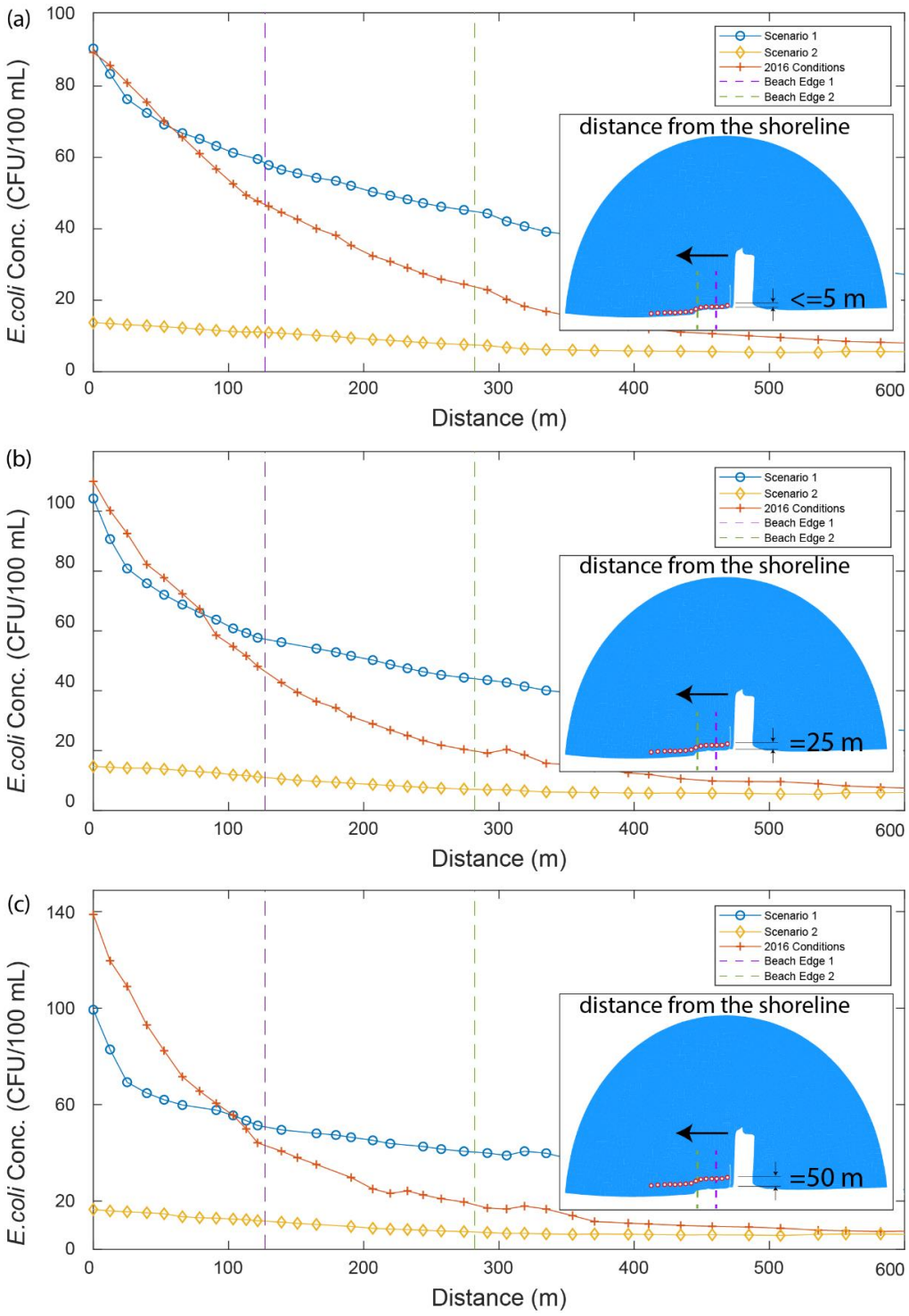

Figure 6-8: Comparison of the simulated alongshore profile of E. coli concentration at LP Beach at a) less than 5 m from shore b) 25 m from shoreline c) 50 m from the shoreline. The x-axis shows the distances from the mouth of the Belle River in the direction of the arrow.

The alongshore profile is illustrated in Figure 6-8. In the area immediately adjacent to the west edge of the Belle River the mean *E. coli* concentration with the presence of the breakwater is higher compared to the case when there is no breakwater and jetty. That has happened because the plum from the Belle River can transfer both east and west directions when there is no obstacle such as breakwater to hold it in one place. On the west side of the breakwater, contamination moves from river mouths to the beach no matter what is the direction of prevailing currents. Once it reaches to the edge of the breakwater, it then moves along the shoreline in any direction that the dominant current is. In the case of having jetty installed (Scenario 2), the *E. coli* concentration along the shoreline is essentially uniform and lower than the other two cases. These findings indicate that without jetty, the loading and mixing can result in highly non-uniform distributions of *E. coli* along the LP Beach with high *E. coli* persist to happen in the area immediately adjacent to the Belle River.

6.4 Conclusion

A 3D high-resolution hydrodynamic modelling was coupled with microbial modelling to identify possible sources and contributors to microbial water quality (*E. coli*) at a beach and drinking water intake in the nearshore region of the southern edge of Lake St. Clair. Both lake-wide 3D hydrodynamic model with structured grid AEM3D and unstructured grid TUFLOW-FV platforms were in good agreement with observed water temperatures and ADCP data. However with comparable computational times, the TUFLOW-FV model with a resolution of 50 m in the nearshore region provided better performance statistics than the AEM3D model with a uniform resolution of 200 m in the lake including nearshore areas.

A finer grid ranging from 5-10 m nested model over a 2 km nearshore region surrounding Belle River was successfully applied to simulate *E. coli* concentrations at LP Beach and Lakeshore WTP intake using forcing conditions from the lake-wide TUFLOW-FV model. Model results showed a significant difference between lake-wide and nested model at LP Beach and near the mouth of the Belle River while the differences reduced to a factor of about 1.3 or less at the Lakeshore WTP water intake, which is located about 1 km from the shore. The differences are attributed to nearshore conditions at Belle River and its vicinity that include its narrow width of about 30 m at its confluence, a 25 m jetty separating the Belle River from the adjacent LP Beach, and a 600 m marina breakwater nearby. The results show that source and shoreline features can significantly influence water quality in the nearshore region. If features likely to impact water quality exist in the nearshore region of interest, it is recommended that an appropriate finer grid nested model be considered to investigate whether the nearshore features can significantly affect the results obtained using the model with a coarser grid.

Results obtained using the nested model are in the same range and in reasonable agreement with observations both at LP Beach and Lakeshore WTP intake location. More than 90% of the predicted *E. coli* concentration at LP Beach is linked to the Belle River at all times. At Lakeshore WTP intake, while the contribution of the Belle River is <20% when the predicted concentrations are low (<10 CFU/ 100 ml), the contribution increased to >80% when higher concentrations (10-35 CFU/ 100 ml) are predicted. The results also indicate that the construction of the marina may have contributed to some increase in *E. coli* concentrations at LP Beach. However, construction of a new 150 m jetty in 2018, in place of the 25 m jetty separating Belle River from LP Beach, is expected to reduce the *E. coli*

concentrations at LP Beach to be even lower than those in the absence of both. The impact of these structures on *E. coli* concentrations at Lakeshore WTP intake is predicted to be marginal.

Further testing of the models developed and their use to identify possible sources of human/ animal fecal contamination is constrained by the very limited monitoring data available at the locations of interest, as well as the that related to the flow and microbial loadings from the eight smaller tributaries affecting the nearshore water quality on the southern edge of Lake Clair adjoining Windsor Essex Region. Besides, *E. coli* data is at times confounded by *E. coli* from non-fecal sources. Also, it is still possible that contributions from some sources such as septic system, storm drainage etc. are missing in the model. For example, there is no field-based study of the effects of shoreline dwelling septic systems on water quality in the nearshore of Lake St. Clair. Although there is no direct evidence to indicate that failing septic systems are the source of *E. coli* affecting an area of the shoreline in the region, but it is worth investigating. Many markers have now been identified for monitoring that are much better indicators for human or animal fecal contamination than *E. coli*. If data on such markers are collected, the same modelling framework as used in the current study can be more useful in identifying potential sources of fecal contamination or to help identify those currently unidentified.

6.5 Acknowledgement

The authors are grateful to the Ontario Ministry of the Environment and Climate Change (MOECC) for the ADCP data and also to Windsor Essex County Health Unit (WECHU) for providing the weekly *E. coli* data. We thank Environment and Climate Change Canada (ECCC) and Canada Centre for Inland Waters for their support and special thanks to Dr.

Tom Edge and his team for providing *E. coli* data for the Thames River. A special thanks to Katie Stammler from Essex Region Conservation Authority (ERCA) for all her supports during the project and providing E. coli data for small tributaries. We thank Nicole Drumm, Erin Carroll and Kelli Smith from St. Clair Region Conservation Authority and Lilly Snobelen from Chatham-Kent PUC to provide data that helped us in *E. coli* estimation for the Sydenham and St. Clair River. A special thanks to Mr. Mitchell Smith from BMT, developer of TUFLOW-FV software, and Dr. Matthew Hipsey and his team to help us in setting up the AED2+ pathogen module and all their supports during the project. This research received financial support from Canada's Natural Sciences and Engineering Research Council (NSERC) Strategic Project Grant (SPG) and was performed while Mr. Madani (first author) held an Ontario Trillium Scholarship (OTS).

6.6 References

Acosta, M., Anguita, M., Fernández-Baldomero, F.J., Ramón, C.L., Schladow, S.G., Rueda, F.J., 2015. Evaluation of a nested-grid implementation for 3D finite-difference semi-implicit hydrodynamic models. Environmental Modelling & Software 64 241-262.

Anderson, E.J., Schwab, D.J., 2011. Relationships between wind-driven and hydraulic flow in Lake St. Clair and the St. Clair River Delta. Journal of Great Lakes Research 37(1) 147-158.

Anderson, E.J., Schwab, D.J., Lang, G.A., 2010. Real-Time Hydraulic and Hydrodynamic Model of the St. Clair River, Lake St. Clair, Detroit River System. Journal of Hydraulic Engineering 136(8) 507-518.

Beletsky, D., Schwab, D., McCormick, M., 2006. Modeling the 1998–2003 summer circulation and thermal structure in Lake Michigan. Journal of Geophysical Research: Oceans 111(C10).

Beletsky, D., Schwab, D.J., 2001. Modeling circulation and thermal structure in Lake Michigan: Annual cycle and interannual variability. Journal of Geophysical Research: Oceans 106(C9) 19745-19771.

BMT_Pty._Ltd., 2019. TUFLOW FV user manual. Flexible Mesh Modelling.

Bocaniov, S.A., Scavia, D., 2018. Nutrient Loss Rates in Relation to Transport Time Scales in a Large Shallow Lake (Lake St. Clair, USA—Canada): Insights From a Three-Dimensional Model. Water Resources Research 54(6) 3825-3840.

Chen, C., Liu, H., Beardsley, R.C., 2003. An Unstructured Grid, Finite-Volume, Three-Dimensional, Primitive Equations Ocean Model: Application to Coastal Ocean and Estuaries. Journal of Atmospheric and Oceanic Technology 20(1) 159-186.

Edsall, T.A., Charlton, M.N., 1997. Nearshore waters of the Great Lakes. Environment Canada.

Ganju, N.K., Brush, M.J., Rashleigh, B., Aretxabaleta, A.L., del Barrio, P., Grear, J.S., Harris, L.A., Lake, S.J., McCardell, G., O'Donnell, J., Ralston, D.K., Signell, R.P., Testa, J.M., Vaudrey, J.M.P., 2016. Progress and Challenges in Coupled Hydrodynamic-Ecological Estuarine Modeling. Estuaries and Coasts 39(2) 311-332.

Ge, Z., Whitman, R.L., Nevers, M.B., Phanikumar, M.S., Byappanahalli, M.N., 2012. Nearshore hydrodynamics as loading and forcing factors for Escherichia coli contamination at an embayed beach. Limnology and Oceanography 57(1) 362-381.

Health-Canada, 2012. Guidelines for Canadian Recreational Water Quality, Third Ed., Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, (Catalogue No H129-15/2012E).

Hipsey, M., Boon, C., Paraska, D., Bruce, L., Huang, P., 2019. Aquatic EcoDynamics (AED) Model Library Science Manual. The University of Western Australia Technical Manual, Perth, Australia 34.

Hipsey, M.R., Antenucci, J.P., Brookes, J.D., 2008. A generic, process-based model of microbial pollution in aquatic systems. Water Resources Research 44(7).

Huang, A., Rao, Y.R., Lu, Y., Zhao, J., 2010. Hydrodynamic modeling of Lake Ontario: An intercomparison of three models. Journal of Geophysical Research: Oceans 115(C12).

Huang, C., Kuczynski, A., Auer, M.T., O'Donnell, D.M., Xue, P., 2019. Management Transition to the Great Lakes Nearshore: Insights from Hydrodynamic Modeling. Journal of Marine Science and Engineering 7(5) 129.

Huziy, O., Sushama, L., 2017. Lake–river and lake–atmosphere interactions in a changing climate over Northeast Canada. Climate Dynamics 48(9) 3227-3246.

Leon, L.F., Smith, R.E.H., Malkin, S.Y., Depew, D., Hipsey, M.R., Antenucci, J.P., Higgins, S.N., Hecky, R.E., Rao, R.Y., 2012. Nested 3D modeling of the spatial dynamics of nutrients and phytoplankton in a Lake Ontario nearshore zone. Journal of Great Lakes Research 38 171-183.

Liu, J., Wang, B., Oldham, C.E., Hipsey, M.R., 2020. Unravelling the metabolism black-box in a dynamic wetland environment using a hybrid model framework: Storm driven changes in oxygen budgets. Science of The Total Environment 723 138020.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020a. Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach. Journal of Great Lakes Research.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020b. Z Microbial Modelling of Lake St. Clair: Impact of Local Tributaries On the Shoreline Water Quality. Unpublished- Chapter 5.

Makarewicz, J.C., Howell, E.T., 2012. The Lake Ontario Nearshore Study: Introduction and summary. Journal of Great Lakes Research 38 2-9.

Martyr-Koller, R.C., Kernkamp, H.W.J., van Dam, A., van der Wegen, M., Lucas, L.V., Knowles, N., Jaffe, B., Fregoso, T.A., 2017. Application of an unstructured 3D finite

volume numerical model to flows and salinity dynamics in the San Francisco Bay-Delta. Estuarine, Coastal and Shelf Science 192 86-107.

Nevers, M.B., Whitman, R.L., 2005. Nowcast modeling of Escherichia coli concentrations at multiple urban beaches of southern Lake Michigan. Water Research 39(20) 5250-5260.

Rao, Y.R., Sheng, J., 2008. Application of a Nested-grid hydrodynamic model for circulation and thermal structure in the coastal boundary layer of Lake Huron. Aquatic Ecosystem Health & Management 11(2) 161-166.

Spillman, C.M., Hamilton, D.P., Hipsey, M.R., Imberger, J., 2008. A spatially resolved model of seasonal variations in phytoplankton and clam (Tapes philippinarum) biomass in Barbamarco Lagoon, Italy. Estuarine, Coastal and Shelf Science 79(2) 187-203.

Toze, S., Hodgers, L., Palmer, A., Sidhu, J., Page, D., Williams, M., Kookana, R.S., Bartkow, M., Sedlak, D.L., Stratton, H., 2012. Natural attenuation of pathogens and trace contaminants in South East Queensland waterways.

USEPA, 2010. Sampling and Consideration of Variability (Temporal and Spatial) For Monitoring of Recreational Waters, report number: EPA-823-R-10-005.

Valipour, R., Rao, Y.R., León, L.F., Depew, D., 2018. Nearshore-offshore exchanges in multi-basin coastal waters: Observations and three-dimensional modeling in Lake Erie. Journal of Great Lakes Research.

Warren, G.J., Lesht, B.M., Barbiero, R.P., 2018. Estimation of the width of the nearshore zone in Lake Michigan using eleven years of MODIS satellite imagery. Journal of Great Lakes Research 44(4) 563-572.

Zhang, L., Hipsey, M.R., Zhang, G.X., Busch, B., Li, H.Y., 2017. Simulation of multiple water source ecological replenishment for Chagan Lake based on coupled hydrodynamic and water quality models. Water Supply 17(6) 1774-1784.

CHAPTER 7 GENERAL CONCLUSION

7.1 Summary

Fecal microbial pollution is currently monitored using *Escherichia coli* (*E. coli*) as fecal indicator bacteria (FIB). Monitoring methods have several limitations including their inability to predict water quality in real-time or identify potential sources of contamination for more effective management. Mathematical models are tools that can be very effective and complementary to monitoring in overcoming its limitations. Model predictions can be real-time and can also help to identify or eliminate potential sources of microbial pollution. Two main types of models are commonly used: statistical and mechanistic models. Currently, no microbial water quality model has been developed for Lake St. Clair or the nearshore region along its southern edge.

A simpler approach that has shown promising results is to use FIB monitoring data and complement them with common weather data and easily measurable water quality data to develop a predictive NowCast statistical model to predict FIB levels for a particular beach, using mathematical techniques such as multiple linear regression (MLR). Such an approach was applied to develop several MLR models for Sandpoint Beach located in Windsor Essex County (WEC). All the MLR models developed performed better than the current assumptions on microbial water quality based on the monitoring data. The inclusion of a commonly available qualitative weather data, that is currently not used in such models, significantly improved model performance. A strategy was presented to conclude that two years of FIB monitoring data could be sufficient for MLR model development at Sandpoint Beach. A new tool based on Principal Coordinate Analysis (PCoA) for MLR model comparison and selection based on multiple performance criteria.

Microbial water quality in a lake environment is influenced by the complex interaction of hydrodynamics and ecological conditions and affected by several factors including microbial load (from tributaries) and location, die-off (decay) of the microbes with time, weather conditions, and lake bathymetry. A 3D unsteady-state mechanistic model is expected to be required to better predict the spatial and temporal variability in microbial water quality within the lake. Several 3D mechanistic models of increasing complexity were developed and applied to better understand and simulate microbial water quality, and presented in Chapters 3 – 6. Microbial loadings from watershed are expected to be flow driven events. Thus it is expected that the major sources of microbial loading to Lake St. Clair are the tributaries that bring most of the flow into the lake. Four major tributaries of Lake St. Clair contribute > 99.5% of the flow. Chapter 3 was the first attempt to assess the contribution of the major tributaries to microbial water quality in the lake and at Sandpoint Beach using AEM3D model. Using a constant but conservative estimate of *E. coli* decay rate of 0.5 d^{-1}, the results showed that while St. Clair River was the most significant contributor to the predicted *E. coli* concentration in most of the lake, including Sandpoint Beach, the maximum predicted *E. coli* concentrations were <100 CFU/100 mL in most of the central to the southern edge of Lake St. Clair. In the nearshore region along the southern edge, including Sandpoint Beach, the predicted concentrations were even lower at <10 CFU/100 mL. The results indicate that the four major tributaries are not responsible for the *E. coli* exceedances of the Ontario safety guidelines for recreational use at Sandpoint Beach.

The fate and transport of *E. coli* in environmental waters are affected by hydrodynamics and a complex array of physical, chemical, and biological factors. Of these factors, water

temperature and incoming solar radiation (insolation) are arguably the most potent in the inactivation and die-off rate of *E. coli* in water. Two main approaches can be used to simulate *E. coli* fate in the water environment, i) using constant decay rate and ii) using the time-variable decay rate. When the data is limited, using a constant decay rate still can provide useful information about the fate of the microbial contamination however, using the time-variable decay rate gives more realistic predictions. The impact of decay rate on the modelling of microbial contamination investigated in Chapter 4. Results showed that using the time-variable decay rate is preferable for dynamic simulations over time. The impact of water temperature was also quite sensible when the trend of *E. coli* concentration was compared with the water temperature data.

Results from Chapter 3 showed that *E.coli* concentrations in the region adjacent to the southern edge of Lake St. Clair were not significantly impacted by the four major tributaries. Eight small but seasonally significant tributaries of widely varying drainage areas and hydrologic features discharge into Lake St. Clair along the southern shores. The flow and pollution load from these small tributaries are expected to be negligible in the context of the larger lake. However, their effect on the nearshore region close to their confluence with Lake St. Clair is investigated in Chapter 5. In the nearshore region of about 40 km stretch along the southern edge of the lake, while the flow contributions from the eight tributaries were estimated to range between 18-35%, the contribution of *E. coli* microbial pollution was estimated to be greater than 80%. This region includes two beaches (Lakeview Park West Beach and Sandpoint Beach) and two drinking water treatment plant intakes for the Town of lakeshore.

The grid sizes of 200 and 400 m used in the study by (Madani et al., 2020a, b) is not deemed sufficient to adequately model the hydrodynamics and water quality impacts in the nearshore region of small tributaries such as Belle River with widths of <50 m, where a grid size of 5 – 10 m may be desirable. Further reducing the grid size in this structured grid model is not practical due to computational expense. A lake-wide unstructured grid TUFLOW-FV coupled with AED2+ ecological model was developed and results compared with a finer grid (up to 5 m) nested model over a 2 km nearshore region surrounding Belle River. The results show up to a 4 fold difference in simulated *E. coli* concentrations between the lake-wide and nested models at Lakeview Park West Beach, which is right next to Belle River, confirming the need for the finer resolution model. The differences reduced to a factor of up to 1.3 at the Lakeshore WTP water intake, which is about 1 km from the shore. Besides results showed that the shoreline features such as marina that is protected by 600 m breakwaters and newly constructed 150 m jetty can significantly influence water quality in the nearshore region.

7.2 Engineering Significance

Minimizing human health risk from microbial water quality at recreational beaches or source waters for drinking water supplies requires accurate and timely information on microbiological conditions and potential sources, to issue advisories or undertake remedial action. For example, in the United States, about 50% of the beach closing/advisory days were attributed to unknown sources of pollution (Feng et al., 2015). Mathematical models are tools that can help in this regard. The choice of a model will depend on many factors such as hydrological features, local dynamic process, area of interest and availability of the data and computational power. For beach managers or public health agencies, mechanistic

models are usually challenging to build and less accessible due to their intrinsic complexity and high computational demands. Statistical models in such cases are preferred because of their simplicity. To use the statistical models, it is needed to collect information on environmental parameters that can determine microbial contamination at a specific site.

The current study for the first time for Sandpoint Beach, which is a popular beach in the region, developed a predictive statistical model based on Multiple Linear Regression that showed a much higher degree of accuracy in prediction *E. coli* concentration, as compared to the status quo. The predictive capability of the model was significantly improved by including qualitative weather data that is widely available but not previously used in such models. The study also provided guidance on time period necessary for the training period, as well as presented a new tool based on Principal Coordinate Analysis (PCoA) for MLR model comparison and selection based on multiple performance criteria that are expected to be useful in future studies.

Mechanistic models on the other hand are implemented to improve the knowledge-base on the fundamentals of fate and transport of microbial pollution to not only predict microbial water quality but to also help with the identification of sources of microbial contamination and their relative impact at any given location. Lake St. Clair's watershed is heavily impacted by human activity, which can result in contamination of its waters by fecal matter of human or animal origin, currently monitored using fecal indicator bacteria (FIB) such as *E. coli*, but the source of the contamination is currently not known.

The current study developed and used a series of mechanistic models of increasing complexity to develop a better understanding of the microbial contamination and contributing sources in Lake St. Clair and the nearshore region in the southern edge of the

lake. Modelling results were compared with observations at locations of interest to identify sources and their contributions or identify the need for model modifications or missing sources or data. The impact of major tributaries, local inputs from the small tributaries and also the existence of the structures or barriers in the near vicinity of the beach were studied for the selected locations in Lake St. Clair. The models developed can be used for further assessment and management of microbial water quality in Lake St. Clair, or used for further model development and applications for the management of other water quality parameters. The modelling framework and strategies used can be useful for microbial water quality assessments in other lake environments.

7.3 Recommendations for Future Work

During the current study, some aspects were identified that have the potential for further investigation in future studies. For example, in the case of the statistical model, it is recommended to:

- Develop other types of models such as artificial neural networks (ANN), logistic regression, classification tree etc. and compare the results with the MLR models.
- Develop a real-time toolkit that continuously collects data from the available weather stations, water quality probs, and captured photos of beach conditions using systems such as WebCAT or COSMOS which are the lightweight video monitoring system (Dusek et al., 2019; Taborda and Silva, 2012) and analyze them to predict the *E. coli* level in the beach area.

In the case of mechanistic modelling frameworks, more sensitive, reproducible and standardized methods have to be developed to evaluate microbial concentrations in

tributaries, waters and other compartments involved such as sewage, catchment systems and sediment. In this regards it is recommended to:

- Simulate for a longer period such as two years and study the ice cover, lake turnover etc.
- Investigate the impact of sediment loading on the *E. coli* concentration. In natural environments, sediments are generally mixtures of clay, silt and sand which *E. coli* can attach with different fractions. Also, bed composition and consolidation will have effects on the fate and transport of *E. coli*.
- Develop a near-future hydrodynamic and *E. coli* model by collecting the forcing and loading data from the weather forecasting models and stream water quality models respectively.
- Integrate the current model with the watershed models with the capability of predicting the *E. coli* concentration in the streams.
- Increase the understanding of nearshore processes by study the impact of the urbanization and threats of future climate change.
- Collect high frequency data of *E. coli* from the point and non-point sources and find more accurate forcing data such as stream flow rates because models are never better than the data feeding them.

7.4 References

Dusek, G., Hernandez, D., Willis, M., Brown, J.A., Long, J.W., Porter, D.E., Vance, T.C., 2019. WebCAT: Piloting the Development of a Web Camera Coastal Observing Network for Diverse Applications. Frontiers in Marine Science 6(353).

Feng, Z.X., Reniers, A., Haus, B.K., Solo-Gabriele, H.M., Wang, J.D., Fleming, L.E., 2015. A predictive model for microbial counts on beaches where intertidal sand is the primary source. Marine Pollution Bulletin 94(1-2) 37-47.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020a. Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach. Journal of Great Lakes Research.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020b. Z Microbial Modelling of Lake St. Clair: Impact of Local Tributaries On the Shoreline Water Quality. Unpublished- Chapter 5.

Taborda, R., Silva, A., 2012. COSMOS: A lightweight coastal video monitoring system. Computers & Geosciences 49 248-255.

APPENDICES

Appendix – 1 Nomenclature

Table 1-S1: Nomenclature of the often-used terms (abbreviations).

Term	Definition	Units
C	*E. coli* concentration	$orgs/m^{-3}$
α	Intercept	$CFU/100\ ml$
β	Coefficient	$-$
γ^2	Gamma squared	$-$
ϵ	Error	$-$
Y	Predicted E. coli Concentration	$CFU/100\ ml$
C_{in}	Inflow flux of E. coli	$orgs/(m^{-3}.s)$
C_{out}	Inflow flux of E. coli	$orgs/(m^{-3}.s)$
x_j	x_j is the distance in the j^{th} dimension	m
U_j	the velocity in the j^{th} dimension	m/s^{-1}
j	dimension	x, y, z
$Ecoli$	E. coli Concentration	$CFU/100\ ml$
K	Overall decay rate	d^{-1}
k_g	Growth rate	d^{-1}
l	Light bandwidth	$-$
k_p	Predation rate	d^{-1}
k_T	Dark mortality rate	d^{-1}
k_l	Light inactivation rate	d^{-1}
κ_j	eddy-diffusivity	$m^2.s^{-1}$
t	time	s

Appendix – 2 Supplementary Information of Chapter 2

2-1: Metrics:

Model performance was examined by determination of metrics such as RMSE, accuracy, sensitivity (ability to predict true positive), and specificity (ability to predict true negative) which are defined as follows:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\log_{10} P_i - \log_{10} O_i)^2}{N}} \quad (1)$$

$$Model\ Sensitivity = \frac{TP}{(TP+FN)} \times 100 \quad (2)$$

$$Model\ Specificity = \frac{TN}{(TN+FP)} \times 100 \quad (3)$$

$$Model\ Accuracy = \frac{(TP+TN)}{N} \times 100 \quad (4)$$

Where N is number of observations, $\log_{10} P_i$ is the log10-transformed predicted model value, $\log_{10} O_i$ is the log10-transformed observation, and TP, TN, FP, and FN - as described below- are the numbers of true positives, true negatives, false positives (Type I error), and false negatives (Type II error) respectively. All the last three metrics range from 0 to 100, with 100 being perfect.

TP: True positive; when the modelled and observed *E. coli* levels are both above the bathing water standard (BWS) of 200 CFU/100 ml

TN: True negative; when the modelled and observed *E. coli* levels are both below the bathing water standard (BWS) of 200 CFU/100 ml

FP: False positive; when the modelled *E. coli* levels are above BWS but the observed *E. coli* levels are below BWS. (Type I error)

FN: False negative; when the modelled *E. coli* levels are below BWS but the observed *E. coli* levels are above BWS. (Type II error)

2-2: Detailed description of explanatory variables:

In the initial development phase, the key to developing a good model is to select the proper set of explanatory variables and ensuring that they are easily available in a timely manner. Table 2-S1 shows the list of all the explanatory variables used in developing M and W series MLR models. These variables are selected based on the literature review and their availability in the study area. Description of each explanatory variable is as follow:

WeatherRank: Please see section "Qualitative data preparation for model series W"

RainCombine: Cumulative rainfall for the past 8, 12, 24, 48 and 72 hours before the sampling time is used to create RainCombine variable. This variable captures the rainfall condition over a longer period. Rainfall is undoubtedly the main contributor to bring contaminations and *E. coli* into the lake. A strong correlation between log-transformed *E. coli* concentration and RainCombine is observed for all the models.

RainCombine calculated based on the following formula.

$$\text{RainCombine} = \beta1*R8 + \beta2*R12 + \beta3*R24 + \beta4*R48 + \beta5*R72$$

β1 - β5 are the coefficients that are calculated to make the best fit between Log *E. coli* and RainCombine.

Atemp: Air Temperature. Survival of *E. coli* in recreational water has been linked to water temperature which directly related to air temperature. An increase in air temperature increases water temperature and consequently a progressive increase in the biological death rate.

WDRanked: Ranked Wind Direction. The direction from which the wind blows. It reported in degrees clockwise from the geographic north pole. For example, 90 degrees means East wind and 180 means South wind. Antecedent hourly wind direction data was compiled for the instantaneous value at sampling time and for the 10-h period ending at sampling time. Wind directions were placed in eight categories as North, Northwest, West, Southwest, South, Southeast, East, and Northeast, and examined to correlate with *E. coli* concentration. One-hot encoding along with a genetic algorithm is used to assign a weight to each category to get the best correlation with *E. coli* concentration.

Turb: Water turbidity. All samples were analyzed for water turbidity on-site at the time of sample collection.

Rain24: Twenty-four-hour antecedent rainfall data from the sampling time.

WSRanked: Ranked wind speed. Antecedent hourly wind speed data was compiled for the instantaneous value at sampling time and for the 10-h period ending at sampling time. Wind speed were placed in four categories (0<ws<2 m/s, 2 m/s<ws<4 m/s, 4 m/s<ws<6 m/s and ws>6 m/s). Similar to WDRanked, one-hot encoding along with a genetic algorithm is applied.

WaveH: Wave height is an independent variable associated with the movement of bacteria through the receiving water. High waves can bring the *E. coli* in the wave-washed zone of the beach. Beach sand can be the potential source of fecal contamination that supports large densities of *E. coli* for a prolonged period.

WSPAIns: The instantaneous value of A-component (alongshore) of the wind moving parallel to the shoreline at sampling time. (See figure 2-S1). A positive value indicates wind moving from

right to left as an observer looks out onto the water. Detail description of beach orientation can be found in the Virtual Beach user manual (Cyterski et al., 2013).

WSPOIns: The instantaneous value of O-component (offshore/onshore) of the wind moving perpendicular to the shoreline at sampling time. Sandpoint Beach shoreline orientation is 306^{o} degrees from the North (See figure 2-S1). A positive value indicates wind direction from the water to the shore.

WSPAVec: Antecedent 10-hour average from the sampling time of the A-component of the wind moving parallel to the shoreline.

WSPOVec: Antecedent 10-hour average from the sampling time of the O-component of the wind moving perpendicular to the shoreline.

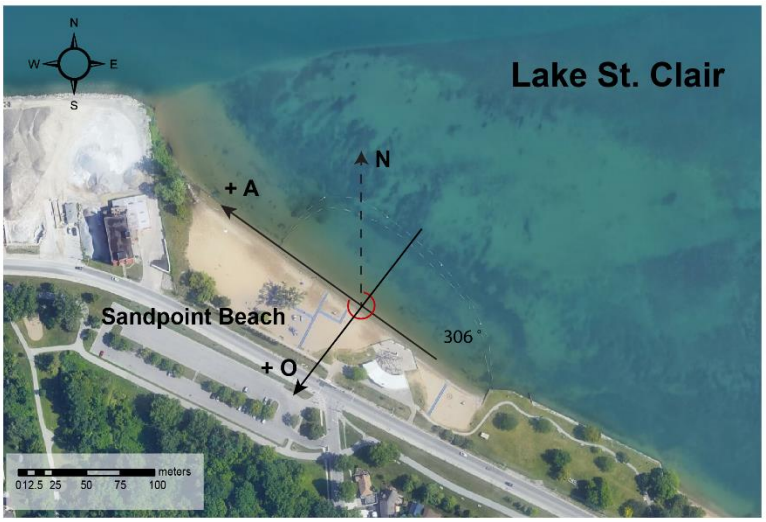

Figure 2-S1: Satellite image of Sandpoint orientation (Google earth V 7.3.0.3832. (Retrieved January 15, 2020)).

Wtemp: Water temperature (^{o}C) at the time of sampling.

WindS: Wind speed (m/s) at the time of sampling.

WSpdVec: Antecedent 10-hour average of wind speed from the sampling time.

WDirVec: Antecedent 10-hour average of wind direction from the sampling time.

NBirds: Number of birds (e.g. gulls, ducks, pelicans, sparrows, geese, and terns) on the beach at the time of sampling. The impact of a single bird on the beach is insignificant. The level of *E. coli* contamination can increase rapidly when dozens of birds settle in the beach shoreline or beach water.

DOY: Day of year

Table 2-S1: List of the explanatory variables that used to develop M and W series models and their Pearson R and p-value in correlation with Log (*E. coli*) concentration. Variables are sorted in ascending order based on their p-value.

W1 and M1 2014-2015	R	P	W2 and M2 2014-2016	R	P	W3 and M3 2014-2017	R	P	W4 and M4 2015-2017	R	P	W5 and M5 2016-2017	R	P
RainCombine	0.387	**0.000**	RainCombine	0.413	**0.000**	RainCombine	0.443	**0.000**	RainCombine	0.426	**0.000**	WeatherRank	-0.497	**0.000**
Atemp	-0.343	**0.000**	WDRanked	0.334	**0.000**	WeatherRank	-0.376	**0.000**	WeatherRank	-0.411	**0.000**	RainCombine	0.491	**0.000**
WeatherRank	0.336	**0.000**	Rain24	0.322	**0.000**	Rain24	0.365	**0.000**	Rain24	0.344	**0.000**	Rain24	0.449	**0.000**
WDRanked	0.318	**0.000**	WeatherRank	-0.321	**0.000**	WDRanked	0.314	**0.000**	Turb	0.317	**0.000**	WDRanked	-0.368	**0.000**
Turb	0.314	**0.000**	Turb	0.302	**0.000**	Turb	0.308	**0.000**	WDRanked	0.277	**0.000**	WSRanked	-0.239	**0.010**
Rain24	0.306	**0.001**	WaveH	0.232	**0.001**	WaveH	0.217	**0.001**	WSRanked	0.213	**0.005**	DOY	0.228	**0.014**
WSRanked	0.305	**0.001**	WSRanked	0.217	**0.003**	Atemp	-0.190	**0.003**	WindS	0.173	**0.022**	WSPAVec	-0.169	0.069
WaveH	0.302	**0.001**	Atemp	-0.215	**0.003**	WSPAVec	-0.188	**0.003**	WSPOVec	0.144	0.058	Turb	0.163	0.080
WSPOVec	0.232	**0.010**	WSPAVec	-0.212	**0.004**	WSPOVec	0.184	**0.004**	WaveH	0.141	0.064	Wtemp	0.133	0.154
Wtemp	-0.231	**0.010**	WindS	0.192	**0.009**	WSRanked	0.163	**0.011**	WSpdVec	0.138	0.068	WSPAIns	-0.129	0.165
WindS	0.231	**0.010**	WSPOVec	0.173	**0.018**	DOY	0.125	0.053	WSPAVec	-0.119	0.115	WSPOVec	0.118	0.204
WSPAVec	-0.211	**0.019**	WSpdVec	0.126	0.085	WSPAIns	-0.112	0.084	WSPOIns	0.084	0.268	NBirds	0.094	0.313
WSpdVec	0.153	0.092	WSPAIns	-0.090	0.222	WSpdVec	0.106	0.100	Atemp	-0.069	0.367	WSPOIns	0.087	0.348
WDirVec	-0.119	0.188	DOY	0.090	0.222	WindS	0.090	0.163	DOY	0.068	0.373	WSpdVec	0.071	0.449
WSPAIns	-0.096	0.292	NBirds	0.090	0.223	WSPOIns	0.089	0.168	WDirVec	-0.049	0.517	WaveH	0.061	0.516
WSPOIns	0.091	0.315	WSPOIns	0.071	0.335	WDirVec	-0.081	0.211	NBirds	0.048	0.524	Atemp	0.034	0.715
NBirds	0.018	0.841	WDirVec	-0.065	0.381	NBirds	0.061	0.350	WSPAIns	-0.043	0.571	WindS	-0.018	0.846
DOY	0.013	0.882	Wtemp	-0.048	0.515	Wtemp	-0.060	0.352	Wtemp	0.008	0.919	WDirVec	0.006	0.946

2-3: Qualitative data preparation for model series W:

One of the problems with MLR models is the fact that categorical data cannot be used directly to develop the model. To include a categorical variable with more than two levels in an MLR prediction model, the researcher needs to do some additional steps to ensure that the results are interpretable. If the categorical variable has two levels, it can be used directly as explanatory variables by coding as 0 and 1. In such a condition, the weighted explanatory variable is added to the response depending on whether it is present or absent. If the categorical coded as -1 and 1, the weighted explanatory variable is subtracted from the response when it is -1 and added to response when it is coded as 1. When the categorical variable has more than two levels, some other methods to deal with the data is needed. In general, a categorical variable with n levels will be transformed into n variables each with two levels. This process is called one-hot encoding.

To generate meaningful data for MLR, one-hot-encoder returns a vector for each unique entries of the categorical data. Each vector contains only one '1' while all other values in the vector are '0'. In the case of weather sky condition data, the original data was put in 10 unique categories. 24-hour antecedent data from the sampling time is used to encode for each sample. That generates a matrix of 24*10 for each sample. Supposed we have 100 samples the final matrix size is 100*24*10. The genetic algorithm was used to find the best weight function that gives the best correlation with the *E. coli* concentration. To do so, a MATLAB code was written that solve the below equation to find w vector to get the best linear fit between the final explanatory variable created from weather data and *E. coli* concentration. Population selection is based on stochastic uniform function and crossover

and mutation rate was set to 0.8 and 0.01 respectively. Optimization was terminated when the average change in the fitness value is less than 10^{-6}.

$$X_i = \sum_{k=1}^{24} \prod_{j=1}^{10} \Omega_{i.k.j}\, w_j \quad and \quad i = 1.2.3. \ldots . n \tag{5}$$

$$Y = \beta_0 + \beta_1\, X \tag{6}$$

where X_i is the final explanatory variables created for the i^{th} day of sampling and n is the total number of samples, $\Omega_{i.j.k}$ is the $(k^{th}.j^{th})$ element of the encoded matrix obtained from the one-hot encoding of 24 categorical data, and w_j is the weight values assigned for the j^{th} category of the qualitative data.

2-4: Censored Data:

There are 65, 58, 63, 54, and 64 days of sampling from 2014 to 2018 respectively. Values of 10 and 1000 were used in the calculation for results reported as <10 CFU/100 ml/100 ml and > 1000 CFU/100 ml/100 ml respectively. There was no sample below the limits but 29 samples out of 240 samples (12.1%) during the training period (2014-2017) and 7 samples out of 64 samples (11%) during testing (2018) were reported as 1000 CFU/100 ml. Table 2-S2 show the number of samples and percentage of censored data in each year and for each mode constructed from the different training period. Censored data for all the models were less than 15% that we believe did not have too much impact on the results. We have tried to replace the censored data with the values that we got from the uncensored data correlation with the explanatory variables such as rainfall but the results were not very different. It should be mentioned that we still don't know whether these values are censored data or they are measured as 1000 CFU/100 ml.

Table 2-S2: Censored data information

Years	**2014-2015**	**2014-2016**	**2014-2017**	**2015-2017**	**2016-2017**
Number of samples	123	186	240	175	117
Number of censored data	18	22	29	16	11
Percentage	14.6%	11.8%	12.1%	9.1%	9.4%

2-5: Model Selection and the number of variables:

Exhaustive search (all possible combinations through all variables) which has been used previously in the literature in a similar context was used to build the models (Feng et al., 2015; Simmer, 2016). In computer science, it is often referred to as the Brute-Force search. The intent is to examine all possible combinations. For example, for the M series models, we have 17 explanatory variables. Constructing models based on 7 out of these 17 variables resulted in 19,448 M series models for each training period, which were then compared to select the best model. With qualitative weather data, there are 18 explanatory variables and the selection of 7 will result in 31,824 models. Out of these 31,824 models, there are 12,376 models that one of the explanatory variables is weather data. The best model was selected that gives higher R^2, and lower RMSE and AIC.

To select the number of variables in the models, we used Mallows's Cp value and conducted an exhaustive search among all the models with 3, 4, 5 … to the total number of variables (17 for the case without weather and 18 with weather data). Models were compared based on their Cp values. Results showed that the best models which result in the lowest Cp value had between 6-8 number of variables. Based on the above, we chose to go with 7 variables for all the models.

2-6: Multi-Collinearity check:

Each model was checked for multi-collinearity based on the Variance Inflation Factor (VIF) as suggested by USEPA - Virtual Beach V.3 manual (Cyterski et al 2013). During the model selection within the exhaustive model search, multi-collinearity among predictor variables was checked too. Any model containing an explanatory variable with a high degree of correlation with others (as measured by a large VIF value) was removed from consideration during model selection. As suggested the threshold was set to VIF=5 which means that 80% (1 – 1/VIF = 1 – 1/5 = 4/5) of the variability in an explanatory variable can be explained by the other variables in the model. Among all 19,448 models in M5 series that constructed on 2-year data (2016-2017), less than 1.94% of them had VIF higher than the threshold. Models in W series, higher multi-collinearity were observed in all the training period but still, they are less than 2.26% for W5 models. These models were ignored during model selection.

Table 2-S3: Percent of models with multi-collinearity observed during each training period.

models\ training period	**2014-2015**	**2014-2015**	**2014-2017**	**2015-2017**	**2016-2017**
M series	0.04%	0.36%	0.11%	0.18%	1.94%
W series	0.29%	1.29%	0.40%	0.30%	2.26%

2-7: Effect of Inclusion of Qualitative data:

In order to conduct a direct assessment of the value of adding the qualitative weather parameter, an additional analysis was done as follows. A new set of 19,488 (MW series)

models was created by adding qualitative weather data (as 8th variable) to each of the corresponding M series models (7 variables) to maintain a similar model architecture, for each training period. The M and MW series models with similar model architecture were then compared to see how the inclusion of qualitative weather data change the model metrics (R^2, RMSE and AIC). The results are tabulated below and show that in 100% of the models for R^2, 98-100% models for RMSE, and 91-100% for AIC, the addition of quantitative weather data in MW series models improved model performance as compared to the corresponding M series model with similar model structure. The inclusion of additional variables can often lead to higher R^2 or lower RMSE. However, lowering of AIC values, which get penalized for the inclusion of additional variables to a model, is certainly indicative of improvement in model performance. The results show that the inclusion of the qualitative weather data improved the performance of M series models with a similar architecture in 91 – 99% of the models during training periods 1 and 2, and in 100% of the models in training periods 3 – 5 based on AIC comparison.

Table 2-S4: Percentage of the MW series models that have higher R^2, lower RMSE and lower AIC compared to corresponding M series model with similar model architecture.

Training Period\ Comparison Metric	**R^2**	**RMSE**	**AIC**
Training Period 1 (2014-2015)	100%	99.9%	99.1%
Training Period 2 (2014-2016)	100%	98.2%	91.7%
Training Period 3 (2014-2017)	100%	100%	100%
Training Period 4 (2015-2017)	100%	100%	100%
Training Period 5 (2016-2017)	100%	100%	100%

2-8: Length of the Training period:

Similar to what was done for the qualitative weather parameter, an additional analysis was undertaken to compare models with a similar architecture built with 2- (2016-2017), 3- (2015-2017), and 4- (2014-2017) years of monitoring data. For the M-series models, 19448 models for each training set with similar architecture were compared, whereas, for the W-series models, the comparison was amongst the 12,376 models with a similar structure that included the qualitative weather parameter. The results are presented in table 2-S5.

For the M-series, models built with 2 years of data in training period (2016-2017; 2Y_train), showed higher R^2, lower RMSE, and lower AIC than the corresponding models built with 3- (2015-2017; 3Y_train), and 4- (2014-2017; 4Y_train) years, for 87.7%, 96.1%, and 100% of the models respectively.

Similar results were observed with the W series models where models built with 2 years' data (2016-2017; 2Y_train), showed better performance measured by higher R^2, lower RMSE, and lower AIC than the corresponding models built with 3- (2015-2017; 3Y_train), and 4- (2014-2017; 4Y_train) years, for 99.9%, 100%, and 100% models respectively.

Table 2-S5: Percentage of M and W series models that have higher R^2, lower RMSE and lower AIC compared to corresponding M and W series models with similar architecture over training periods of 2, 3 and 4 years.

Metrics \ Models	**M series models**			**W series models**		
	4Y_train	3Y_train	2Y_train	4Y_train	3Y_train	2Y_train
R^2	8.3%	4.0%	87.7%	0.0%	0.1%	99.9%
RMSE	0.0%	3.9%	96.1%	0.0%	0.0%	100.0%
AIC	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%

2-9: PCoA analysis

Principal Coordinates Analysis (PCoA) is a method to explore and to visualize similarities or dissimilarities of data. PCoA converts data from a high dimensional space into a lower-dimensional space without losing too much of the information. It can be used to visualize individuals and/or clusters of data based on their characteristics in 2 or 3 dimensions. Table 2-S6 shows the models and their metrics used in PCoA. Model X represents the ideal model with 100% accuracy (R^2, sensitivity, specificity and AUROC equal to 1 and RMSE =0). The shorter the distance between "X" and the developed model, the better the model is. Before we use them in PCoA analysis we normalized all the metrics to values between 0 and 1. The results of the normalized data are presented in Table 2-S7. A weight value can be applied to emphasize the impact of each metrics (e.g. sensitivity that is important for public health) in model selection but we did treat them equally. Table 2-S8 displays the dissimilarity matrix used in the PCoA. We use Chi-Squared Distance to generate the dissimilarity matrix. Figure 2-S2 and Table 2-S10 show the scree plot and tabulated data that is used to select the two coordinates that explain the must of the variability in the data. Coordinates F1 and F2 can explain the data variability by 67.9% and 23.1% respectively. The first two columns of Table 2-S10 are used to plot PCoA graph. All the analysis and plot were performed using a MATLAB script and we used XLSTAT trial version (https://www.xlstat.com/en/) to check for the accuracy of the code in doing PCoA analysis.

Table 2-S6: Models with their metric. This data reorganized from Table 2-1 in the main text presented in Chapter 2

Models Metrics	**M1**	**M2**	**M3**	**M4**	**M5**	**W1**	**W2**	**W3**	**W4**	**W5**	**P1**	**P2**	**X**
TP	11	9	7	9	10	18	17	16	15	16	9	6	23
TN	33	32	37	37	37	30	33	32	32	34	26	27	41
FP	8	9	4	4	4	11	8	9	9	7	10	12	0
FN	12	14	16	14	13	5	6	7	8	7	19	17	0
R^2	0.23	0.10	0.07	0.07	0.13	0.17	0.14	0.18	0.10	0.13	0.00	0.02	1
R^2Adj	0.21	0.09	0.05	0.06	0.12	0.16	0.12	0.17	0.08	0.11	-0.02	0.00	1
RMSE	0.29	0.35	0.25	0.34	0.34	0.38	0.38	0.39	0.45	0.42	0.62	0.61	0
Accuracy	69	64	69	72	73	75	78	75	73	78	54	53	100
Specificity	80	78	90	90	90	73	80	78	78	83	72	69	100
Sensitivity	48	39	30	39	43	78	74	70	65	70	32	26	100
AUC	0.73	0.73	0.69	0.72	0.75	0.78	0.78	0.76	0.77	0.79	0.57	0.53	1

Table 2-S7: Normalized data from table 2-S6 to values between 0 and 1.

Models Metrics	**M1**	**M2**	**M3**	**M4**	**M5**	**W1**	**W2**	**W3**	**W4**	**W5**	**P1**	**P2**	**X**
TP	0.29	0.18	0.06	0.18	0.24	0.71	0.65	0.59	0.53	0.59	0.18	0.00	1.00
TN	0.47	0.40	0.73	0.73	0.73	0.27	0.47	0.40	0.40	0.53	0.00	0.07	1.00
FP	0.67	0.75	0.33	0.33	0.33	0.92	0.67	0.75	0.75	0.58	0.83	1.00	0.00
FN	0.63	0.74	0.84	0.74	0.68	0.26	0.32	0.37	0.42	0.37	1.00	0.89	0.00
R^2	0.23	0.10	0.07	0.07	0.13	0.17	0.14	0.18	0.10	0.13	0.00	0.02	1.00
R^2Adj	0.23	0.11	0.07	0.08	0.14	0.18	0.14	0.19	0.10	0.13	0.00	0.02	1.00
RMSE	0.47	0.56	0.40	0.55	0.55	0.61	0.61	0.63	0.73	0.68	1.00	0.98	0.00
Accuracy	0.34	0.23	0.34	0.40	0.43	0.47	0.53	0.47	0.43	0.53	0.02	0.00	1.00
Specificity	0.35	0.29	0.68	0.68	0.68	0.13	0.35	0.29	0.29	0.45	0.10	0.00	1.00
Sensitivity	0.30	0.18	0.05	0.18	0.23	0.70	0.65	0.59	0.53	0.59	0.08	0.00	1.00
AUC	0.43	0.43	0.34	0.40	0.47	0.53	0.53	0.49	0.51	0.55	0.09	0.00	1.00

Table 2-S8: Dissimilarity matrix shows the distance between each pair of the models

	M1	M2	M3	M4	M5	W1	W2	W3	W4	W5	P1	P2	X
M1	0.000	0.296	0.603	0.486	0.404	0.540	0.436	0.366	0.384	0.399	1.003	1.170	0.921
M2	0.296	0.000	0.571	0.494	0.480	0.664	0.589	0.526	0.456	0.544	0.782	0.944	1.209
M3	0.603	0.571	0.000	0.212	0.308	1.059	0.858	0.862	0.813	0.763	1.148	1.311	1.223
M4	0.486	0.494	0.212	0.000	0.129	0.888	0.671	0.686	0.634	0.568	1.116	1.301	1.108
M5	0.404	0.480	0.308	0.129	0.000	0.806	0.587	0.599	0.568	0.484	1.147	1.335	0.986
W1	0.540	0.664	1.059	0.888	0.806	0.000	0.261	0.212	0.282	0.366	1.147	1.329	1.001
W2	0.436	0.589	0.858	0.671	0.587	0.261	0.000	0.124	0.179	0.116	1.169	1.370	0.906
W3	0.366	0.526	0.862	0.686	0.599	0.212	0.124	0.000	0.153	0.189	1.098	1.288	0.912
W4	0.384	0.456	0.813	0.634	0.568	0.282	0.179	0.153	0.000	0.190	1.005	1.203	1.038
W5	0.399	0.544	0.763	0.568	0.484	0.366	0.116	0.189	0.190	0.000	1.150	1.356	0.905
P1	1.003	0.782	1.148	1.116	1.147	1.147	1.169	1.098	1.005	1.150	0.000	0.349	1.864
P2	1.170	0.944	1.311	1.301	1.335	1.329	1.370	1.288	1.203	1.356	0.349	0.000	2.040
X	0.921	1.209	1.223	1.108	0.986	1.001	0.906	0.912	1.038	0.905	1.864	2.040	0.000

Table 2-S9: Principal coordinates and their eigenvalues and variability (%)

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Eigenvalue	3.1	1.1	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Variability (%)	67.9	23.1	6.8	1.4	0.6	0.2	0.0	0.0	0.0	0.0
Cumulative %	67.9	91.0	97.8	99.3	99.8	100.0	100.0	100.0	100.0	100.0

Table 2-S10: Principal coordinates for each models. Coordinates F1 and F2 are selected to plot the PCoA graph

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
M1	0.07	0.03	0.14	-0.09	-0.03	-0.01	0.01	0.00	0.00	0.00
M2	-0.19	0.08	0.04	-0.14	-0.07	0.04	0.00	0.00	0.00	0.00
M3	0.00	0.59	-0.06	-0.03	-0.02	-0.04	0.00	0.00	0.00	0.00
M4	0.07	0.43	-0.11	0.04	0.03	0.00	0.00	0.00	0.00	0.00
M5	0.15	0.35	-0.05	0.04	0.02	0.03	0.00	0.00	0.00	0.00
W1	0.17	-0.45	-0.08	-0.05	-0.02	-0.02	0.00	0.00	0.00	0.00
W2	0.26	-0.22	-0.15	0.01	0.00	-0.03	0.00	0.00	0.00	0.00
W3	0.18	-0.25	-0.05	0.00	0.01	-0.01	0.00	0.00	0.00	0.00
W4	0.07	-0.21	-0.14	0.00	0.01	0.04	0.00	0.00	0.00	0.00
W5	0.23	-0.12	-0.16	0.05	0.03	0.01	0.00	0.00	0.00	0.00
P1	-0.89	-0.09	0.07	0.15	-0.09	0.00	0.00	0.00	0.00	0.00
P2	-1.08	-0.10	0.18	-0.04	0.10	-0.01	0.00	0.00	0.00	0.00
X	0.94	-0.04	0.39	0.06	0.01	0.00	0.00	0.00	0.00	0.00

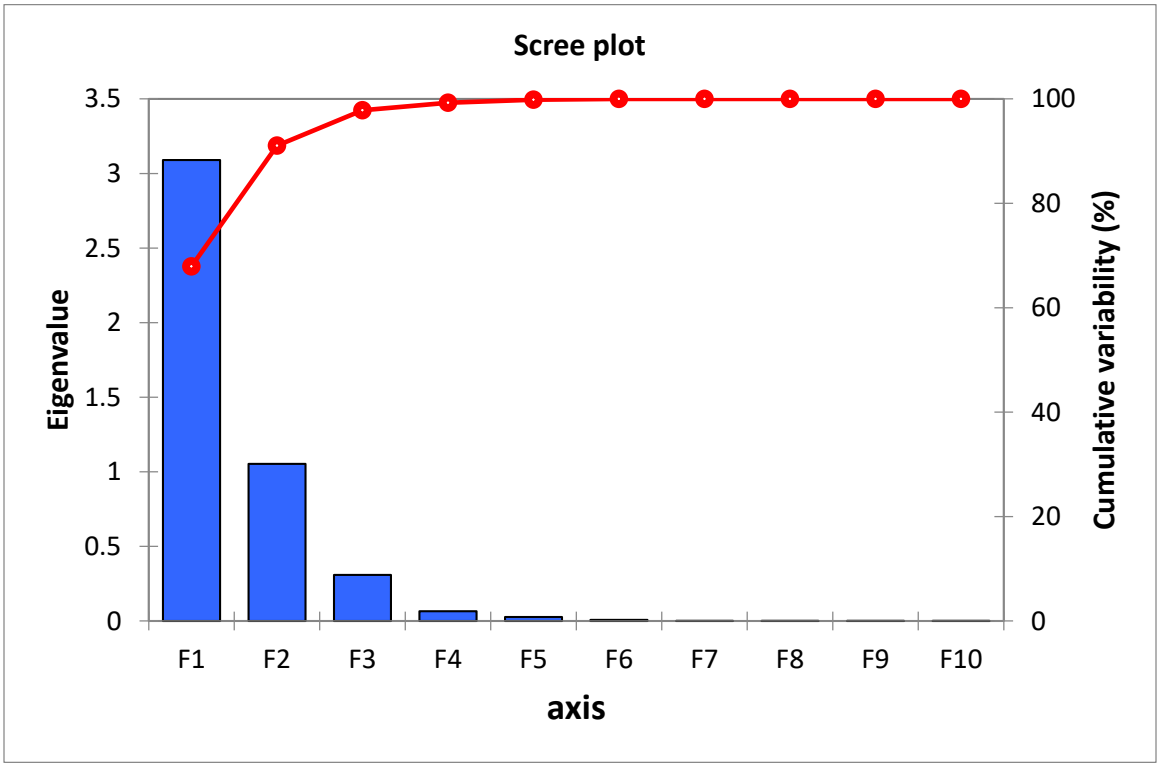

Figure 2-S2: Scree plot that shows the axis and their eigenvalues and cumulative variability.

2-10: Developed Models:

In this section, all the developed M and W series models formula and their estimated coefficient are presented.

M series models:

M1:

Linear regression model:

LOG_ECOLI ~ 1 + RainCombine + Atemp + WSRanked + WaveH + WindS + NBirds + WSPOIns

	Estimate	SE	tStat	pValue
(Intercept)	-3.9862	2.9882	-1.334	0.18484
RainCombine	0.000712	0.000184	3.874	0.000178
Atemp	-0.03398	0.014599	-2.3278	0.021674
WSRanked	0.11	0.049751	2.211	0.029016
WaveH	0.025356	0.01241	2.0431	0.043326
WindS	0.021323	0.011995	1.7777	0.078095
NBirds	0.004014	0.002029	1.9784	0.050278
WSPOIns	-0.06155	0.031252	-1.9694	0.051312

Number of observations: 123, Error degrees of freedom: 115
Root Mean Squared Error: 0.48
R-squared: 0.338, Adjusted R-Squared 0.297
F-statistic vs. constant model: 8.38, p-value = 3e-08

M2:

Linear regression model:

LOG_ECOLI ~ 1 + RainCombine + Turb + WaveH + WSPAVec + NBirds + WDRanked + WSPOIns

	Estimate	SE	tStat	pValue
(Intercept)	-9.3444	3.1136	-3.0011	0.003076
RainCombine	0.000699	0.000137	5.0876	9.15E-07
Turb	0.006768	0.002489	2.7192	0.007192
WaveH	0.016891	0.010765	1.5691	0.11839
WSPAVec	-0.03719	0.013848	-2.6853	0.007932
NBirds	0.005654	0.001571	3.5984	0.000415
WDRanked	1.0842	0.30605	3.5425	0.000506
WSPOIns	-0.05821	0.031577	-1.8435	0.06692

Number of observations: 186, Error degrees of freedom: 178
Root Mean Squared Error: 0.472
R-squared: 0.325, Adjusted R-Squared 0.298
F-statistic vs. constant model: 12.2, p-value = 9.82e-13

M3:

Linear regression model:

LOG_ECOLI ~ 1 + RainCombine + Turb + WSPOVec + NBirds + WDirVec + WDRanked + WSPOIns

	Estimate	SE	tStat	pValue
(Intercept)	-8.4659	2.7991	-3.0245	0.002771
RainCombine	0.001135	0.00017	6.6817	1.73E-10
Turb	0.007543	0.002301	3.2786	0.001204
WSPOVec	0.009933	0.003751	2.648	0.008653
NBirds	0.004556	0.0014	3.2541	0.001307
WDirVec	-0.0298	0.008507	-3.5022	0.000553
WDRanked	1.0517	0.28908	3.6381	0.000338
WSPOIns	-0.04984	0.031542	-1.5801	0.11544

Number of observations: 240, Error degrees of freedom: 232
Root Mean Squared Error: 0.477
R-squared: 0.316, Adjusted R-Squared 0.296
F-statistic vs. constant model: 15.3, p-value = 1.85e-16

M4:

Linear regression model:

LOG_ECOLI ~ 1 + RainCombine + Rain24 + Turb + WSPAVec + NBirds + WDRanked + Wtemp

	Estimate	SE	tStat	pValue
(Intercept)	-7.2808	2.8011	-2.5992	0.01018
RainCombine	0.000754	0.0002	3.7607	0.000234
Rain24	0.008584	0.005036	1.7048	0.090099
Turb	0.009774	0.002646	3.694	0.000299
WSPAVec	-0.03783	0.009999	-3.7838	0.000215
NBirds	0.003212	0.001649	1.948	0.053086
WDRanked	0.80835	0.26363	3.0663	0.002529
Wtemp	0.024918	0.015535	1.604	0.1106

Number of observations: 175, Error degrees of freedom: 167
Root Mean Squared Error: 0.46
R-squared: 0.336, Adjusted R-Squared 0.308
F-statistic vs. constant model: 12.1, p-value = 1.97e-12

M5:

Linear regression model:

LOG_ECOLI ~ 1 + RainCombine + Rain24 + DOY + Wtemp + WSpdVec + WSPOVec + WDRanked

	Estimate	SE	tStat	pValue
(Intercept)	-13.586	4.0257	-3.3747	0.001024
RainCombine	0.001782	0.000588	3.0328	0.003028
Rain24	0.013128	0.006462	2.0317	0.044613
DOY	0.002664	0.00166	1.6052	0.11135
Wtemp	0.043639	0.016763	2.6034	0.010517
WSpdVec	0.22714	0.060165	3.7752	0.000261
WSPOVec	-0.2271	0.061613	-3.6859	0.000357
WDRanked	1.342	0.38471	3.4883	0.000702

Number of observations: 117, Error degrees of freedom: 109
Root Mean Squared Error: 0.43
R-squared: 0.4, Adjusted R-Squared 0.361
F-statistic vs. constant model: 10.4, p-value = 6.42e-10

W Series Models:

W1:

Linear regression model:

LOG_ECOLI ~ 1 + Atemp + WeatherRank + WaveH + WindS + WSPOVec + DOY + WDRanked

	Estimate	SE	tStat	pValue
(Intercept)	-5.0182	5.7454	-0.87343	0.38425
Atemp	-0.04041	0.015635	-2.5847	0.010998
WeatherRank	0.029357	0.007953	3.6913	0.000343
WaveH	0.034347	0.013739	2.4999	0.013832
WindS	0.013618	0.012707	1.0717	0.28611
WSPOVec	-0.00149	0.004139	-0.35914	0.72015
DOY	-3.55E-05	0.001694	-0.02099	0.98329
WDRanked	0.66675	0.55559	1.2001	0.23257

Number of observations: 123, Error degrees of freedom: 115
Root Mean Squared Error: 0.501
R-squared: 0.278, Adjusted R-Squared 0.234
F-statistic vs. constant model: 6.33, p-value = 2.68e-06

W2:

Linear regression model:

LOG_ECOLI ~ 1 + Rain24 + Turb + WeatherRank + WSPAVec + WindS + WSPOVec + WDRanked

	Estimate	SE	tStat	pValue
(Intercept)	-6.0541	3.6107	-1.6767	0.095353
Rain24	0.007587	0.004041	1.8775	0.062088
Turb	0.009207	0.002537	3.6287	0.000372
WeatherRank	-0.01737	0.00602	-2.8854	0.004393
WSPAVec	-0.0362	0.016305	-2.2202	0.027666
WindS	0.006734	0.009887	0.68106	0.49672
WSPOVec	-0.00943	0.004797	-1.9656	0.050899
WDRanked	1.1156	0.34937	3.1931	0.001664

Number of observations: 186, Error degrees of freedom: 178
Root Mean Squared Error: 0.493
R-squared: 0.263, Adjusted R-Squared 0.234
F-statistic vs. constant model: 9.07, p-value = 1.46e-09

W3:

Linear regression model:

LOG_ECOLI ~ 1 + WeatherRank + WaveH + Atemp + WSpdVec + WSPOVec + WDRanked + WSPOIns

	Estimate	SE	tStat	pValue
(Intercept)	-1.5927	2.7842	-0.57204	0.56785
WeatherRank	-0.02515	0.004301	-5.8458	1.70E-08
WaveH	0.019243	0.011936	1.6122	0.10827
Atemp	-0.01145	0.010516	-1.0889	0.27731
WSpdVec	0.1402	0.053166	2.637	0.00893
WSPOVec	-0.14029	0.053963	-2.5997	0.009929
WDRanked	0.83483	0.26079	3.2011	0.001561
WSPOIns	-0.04159	0.033011	-1.2599	0.20896

Number of observations: 240, Error degrees of freedom: 232
Root Mean Squared Error: 0.507
R-squared: 0.23, Adjusted R-Squared 0.206
F-statistic vs. constant model: 9.88, p-value = 8.89e-11

W4:

Linear regression model:

LOG_ECOLI ~ 1 + WeatherRank + Turb + WaveH + DOY + WDirVec + Atemp + WDRanked

	Estimate	SE	tStat	pValue
(Intercept)	-4.7171	4.1536	-1.1357	0.25772
WeatherRank	-0.0379	0.005773	-6.5651	6.29E-10
Turb	0.011498	0.002739	4.1986	4.36E-05
WaveH	0.013377	0.014419	0.92768	0.35491
DOY	-2.85E-06	0.001331	-0.00214	0.99829
WDirVec	-0.02354	0.010068	-2.338	0.020571
Atemp	0.017473	0.012073	1.4473	0.1497
WDRanked	1.2836	0.4259	3.0137	0.002983

Number of observations: 175, Error degrees of freedom: 167
Root Mean Squared Error: 0.47
R-squared: 0.305, Adjusted R-Squared 0.276
F-statistic vs. constant model: 10.5, p-value = 7.37e-11

W5:

Linear regression model:

LOG_ECOLI ~ 1 + WeatherRank + WSRanked + Turb + WSPOVec + WDRanked + WSPAVec + WaveH

	Estimate	SE	tStat	pValue
(Intercept)	-3.1807	4.8651	-0.65378	0.51463
WeatherRank	-0.01911	0.003161	-6.0445	2.13E-08
WSRanked	-0.04726	0.019173	-2.4647	0.015274
Turb	0.008391	0.005781	1.4515	0.14952
WSPOVec	-0.01486	0.005916	-2.5117	0.01348
WDRanked	1.209	0.43826	2.7586	0.00681
WSPAVec	-0.05333	0.020773	-2.5675	0.011599
WaveH	0.004188	0.015023	0.27878	0.78094

Number of observations: 117, Error degrees of freedom: 109
Root Mean Squared Error: 0.446
R-squared: 0.353, Adjusted R-Squared 0.311
F-statistic vs. constant model: 8.49, p-value = 2.9e-08

2-11: False positive and false negative:

False positives or false negatives are site-specific and often also related to the number of exceedances observed. For example, Francy et al. (2003) reported false positive and negative values for six Ohio beaches in the range of 0-27.4% and 2.6-26.3% respectively. False positives and negatives in the current study (see Table 2-S11) were within the range for both the M series models (6.3-14.1% and 18.8-25% respectively), as well as the W series models (10.9-18.8% and 7.8-12.5% respectively). False negatives are of greater concern from a public health perspective, since the recreational water-quality is determined to be acceptable by the model, when in fact the standard was exceeded. There are no advisements from the stakeholder of the study but a target of ≥ 50% for specificity (ability to predict true negative) and ≥ 82% for sensitivity (ability to predict true positive) has previously been targeted in the literature (Francy et al., 2013). The sensitivity of more than 99.6% of all the M series models meet, which selected M1-M5 were among them meet

both of these thresholds. 99.8% of the W series models meet both of these thresholds, which W1-W5 were among them.

Table 2-S11: False positive and false negative of the models

	M1	**M2**	**M3**	**M4**	**M5**	**W1**	**W2**	**W3**	**W4**	**W5**	**PM1**	**PM2**
False positive	12.5%	14.1%	6.3%	6.3%	6.3%	17.2%	12.5%	14.1%	14.1%	10.9%	15.6%	18.8%
False negative	18.8%	21.9%	25.0%	21.9%	20.3%	7.8%	9.4%	10.9%	12.5%	10.9%	29.7%	26.6%

2-12: Wind pattern:

The *E. coli* data>200 CFU/ 100 ml were segregated from data <200 CFU/100 ml for 2018, and the wind pattern for those days is presented in Figure 2-S3.

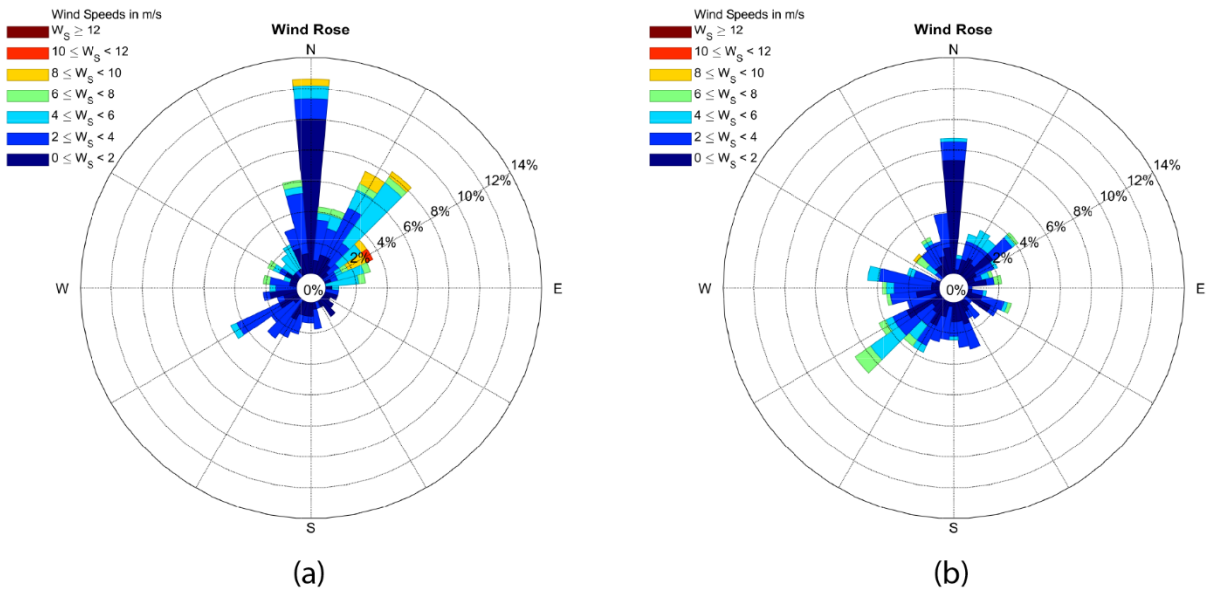

Figure 2-S3. Wind Rose pattern for days that had a) *E. coli* data greater than 200 CFU/100 ml and b) *E. coli* data less than 200 CFU/100 ml in summer 2018

References:

Cyterski, M., Brooks, W., Galvin, M., Wolfe, K., Carvin, R., Roddick, T., Fienen, M., Corsi, S., 2013. Virtual Beach 3: User's Guide. United States Environmental Protection Agency.

Feng, Z.X., Reniers, A., Haus, B.K., Solo-Gabriele, H.M., Wang, J.D., Fleming, L.E., 2015. A predictive model for microbial counts on beaches where intertidal sand is the primary source. Marine Pollution Bulletin 94(1-2) 37-47.

Francy, D.S., Gifford, A.M., Darner, R.A., 2003. *Escherichia coli* at Ohio bathing beaches: Distribution, sources, wastewater indicators, and predictive modeling. US Department of the Interior, US Geological Survey.

Francy, D.S., Stelzer, E.A., Duris, J.W., Brady, A.M., Harrison, J.H., Johnson, H.E., Ware, M.W., 2013. Predictive models for *Escherichia coli* concentrations at inland lake beaches and relationship of model variables to pathogen detection. Appl Environ Microbiol 79(5) 1676-1688.

Simmer, R.A., 2016. Source determination and predictive model development for *Escherichia coli* concentrations at F.W. Kent Park Lake, Oxford, Iowa.

Appendix – 3 Supplementary Information of Chapter 3

3-1: Source of data

Table 3-S1. List of the input/forcing data

#	Input/forcing data	Source
1	Weekly *E. coli* monitoring data	Windsor-Essex County Health Unit (WECHU)
2	30 consecutive days *E. coli* data during late Summer 2010 (August 10 – September 9)	McPhedran et al. (2013)
3	30 consecutive days water temperature data during late Summer 2010 (August 10 – September 9)	McPhedran et al. (2013)
4	Bathymetry	National Geophysical Data Center (https://www.ngdc.noaa.gov/)
5	wind direction, wind speed, air temperature, humidity, atmospheric pressure, rainfall	Windsor A. (Station ID=4716); Windsor Riverside (Station ID=4715) ; Buoy (West Erie) 45005 (41.677 N 82.398 W); Buoy 45147 (42.430N 82.680W); LSCM4 (42.465N 82.755W); Buoy CLSM4 (42.471N 82.877W)
6	solar radiation (SOLAR_RAD and LW_RAD_IN)	Canadian Regional Climate Model (CRCM5) provided by University of Quebec at Montreal (UQAM). Climate data from O. Huziy at Centre ESCER (Étude Simulation du Climat à l'Échelle Régionale), provided 3-hourly data from 1979 to 2012 (Huziy and Sushama, 2017).
7	St. Clair River daily flow	Port Huron station (Station ID= 02GG014)
8	St. Clair River daily water temperature	National Data Buoy Center (NOAA) station at Algonac (Station ID= 9014070)
9	Sydenham River daily flow	Sydenham River at Florence (Station ID= 02GG003) and Bear Creek below Brigden (Station ID= 02GG009)
10	Sydenham River daily water temperature	Provincial Water Quality Monitoring Network (PWQMN) (Station ID= 04002701702)
11	Thames River daily flow data	Thamesville (Station ID= 02GE003)
12	Thames River daily water temperature	PWQMN at the Station ID= 04001305802
13	Flow and water temperatures for the Clinton River	United States Geological Survey (USGS) data site (USGS Station ID= 04165500 Clinton River at Moravian Drive)
14	Detroit River daily flow	Fort Wayne station (Station ID= 02GH015)
15	Satellite-derived lake-wide daily time series of water surface temperature	NOAA (Coast Watch Great Lakes Surface Environmental Analysis, GLSEA: https://coastwatch.glerl.noaa.gov/thredds)

Table 3-S2. Description of the runs.

Run	Grid size	Time step	Domain	Simulation period
1	400 m	300 s	Entire lake	May - September 2010
2	200 m	300 s	Entire lake	May - September 2010
3	100 m	300 s	Entire lake	May - September 2010
4	50 m	60 s	Nested from 200 m grid resolution - Southwestern region of Lake St. Clair	May - September 2010

3-2: Flushing time calculation

Flushing time describe the local water body exchange rate without considering the whole domain concentration distribution (Andutta et al., 2013). In other words, flushing time represents the amount of time needed for the water in one specific area to be replaced by surrounding water. The flushing time is calculated using the Volume Averaged Concentrations (VAC), which is defined as the volume integrated concentration of tracer concentrations over a specific sub-area normalized by the total volume of the sub-area was calculated. Time series of tracer VAC were plotted and the flushing time obtained when concentrations reduced to 1/e (~ 0.37) of its initial concentration (Deleersnijder et al., 2006). The time series of VAC can be approximated by the first order exponential equation, as: $C = C_o \times e^{(-t/T_e)}$ where C is the VAC at time t, C_o is the initial value of VAC and equal to 1.0 at the initialization day, and T_e is the e-folding time.

3-3: Model validation and selection of the bottom boundary layer using ADCP data

An ADCP was deployed on April 27, 2016 at the location indicated in Figure 3-S1. Data was collected every 30-minutes up to November 03, 2016. The simulation was conducted for June to September of 2016. The model output and ADCP data was averaged in a

window of 12 hours and compared using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics. In order to chose a bottom boundary condition, two options were tested: i) turbulent benthic boundary layer (TBBL) and ii) constant drag coefficient (Cd=0.005) for all surface (drag-all). Although drag-all is a common boundary condition in shallow water, TBBL has been successfully applied to shallow waters elsewhere (e.g. Lake Winnipeg by Zhao et al. 2012).

The simulated results for the case of using the TBBL boundary condition appeared to have good agreement with the west-east and south-north components of velocity (Figure 3-S2). In the case of using drag-all boundary condition, the simulated velocity components cannot capture the fluctuation in observed data. The root mean square errors (RMSE) between the simulated and observed west-east and south-north velocity were 0.062 and 0.056 (m/s), respectively. These are higher that the case where TBBL is used (RMSE for west-east and south-north velocity were 0.055 and 0.049 (m/s), respectively). Hence, TBBL was selected because it reproduced the currents well in Lake St. Clair compared to the drag-all option.

In TBBL case, better linear correlation with observed data (Pearson-R of 0.56) was obtained for the west-east component compared to the south-north component of velocity. The simulated south-north component of velocity tends to be over predicted. The model can simulate both low and high velocity and its rapid changes in the west-east direction better than the south-north component. In addition, model accuracy was assessed using Normalized Root Mean Square Difference (NRMSE). NRMSE is used to compare the relative difference between two scalar quantities (west-east and south-north velocity) from the model and ADCP observations (Hamidi et al. 2015). The west-east and south-north velocity components NRMSE are 0.276 and 0.449, respectively.

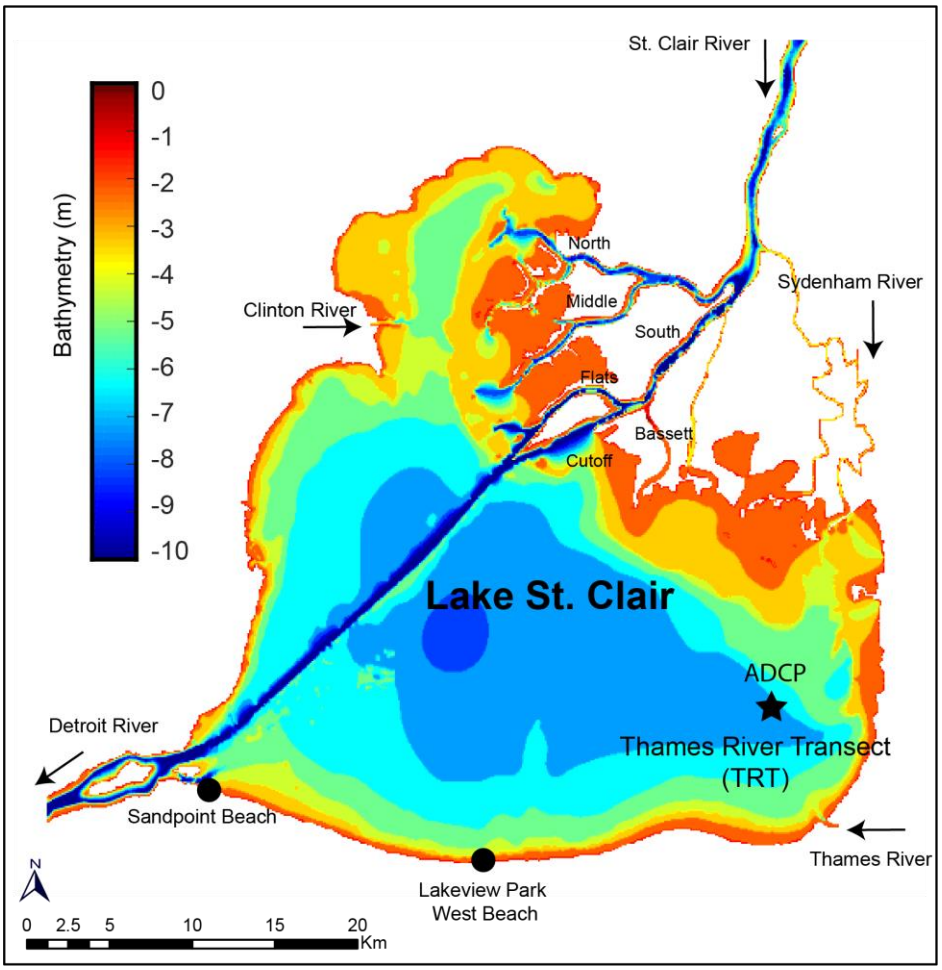

Figure 3-S1. Map of Lake St. Clair and its bathymetry. ADCP location at the Thames River Transect (TRT) (N 42º,22',37.9" and W 82º,30',09.1") indicated as black star on the map.

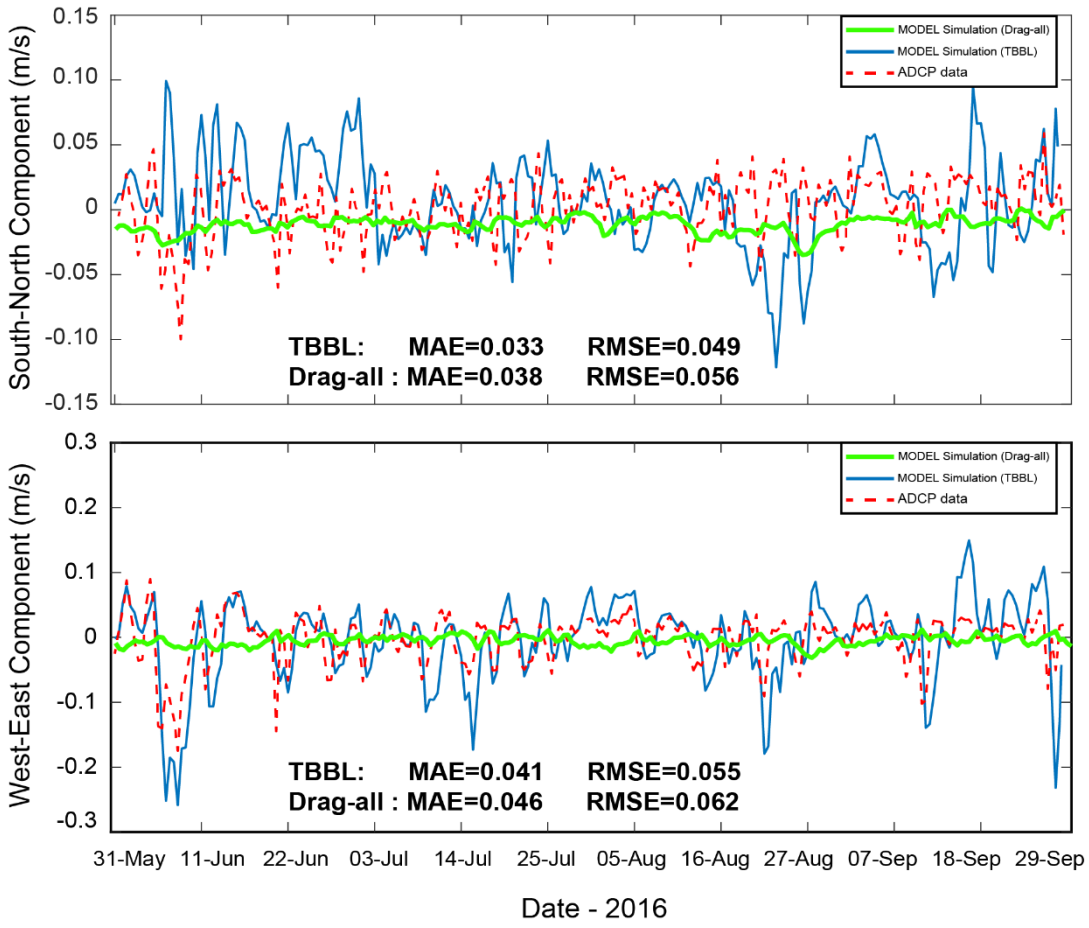

Figure 3-S2. Comparison of observed data (dashed red line) at the location of the TRT ADCP (see Figure 3-S1) with model simulations using TBBL boundary condition (solid blue line), and drag-all boundary condition (solid green line) during 4-month period in 2016.

3-4: Stratification

Model outputs were used to check the occurrence of stratification in the water column of Lake St. Clair. Based on the simulated temperature, the vertical distribution of temperature revealed less than 1 °C variation in the water column more than ~95% of the time. Measured vertical distributions of temperature and water density are not available for Lake St. Clair during the summer of 2010 for comparison. To check for the occurrence of

stratification, additional analysis was done based on the densimetric Froude number (F_{rd}), similar to works done by Loewen et al. (2007) and Ackerman et al. (2001). The Froude number is a dimensionless number that measures the relative strength of inertial to buoyancy forces. It is defined as:

$$F_{rd} = \frac{U}{\sqrt{g\ \Delta h \frac{\Delta \rho}{\rho}}}$$

Where U is the magnitude of depth-average velocity, g is the acceleration due to gravity (~ 9.8 m/s^2), $\Delta\rho$ is the density difference between two points a vertical distance Δh apart, and ρ is the average water density. When F_{rd} is less than unity, then buoyancy forces are stronger than inertial forces and the water column is stratified. In case of $F_{rd} > 1$, inertial forces can overcome the buoyancy forces and the water column remains unstratified.

Figure 3-S3 shows the percentage of the time that stratification can occurred in the water column ($F_{rd} < 1$). In the region near Sandpoint Beach (up to 5.5 km to the east along the south shoreline shown by green rectangle), our model results showed that F_{rd} is greater than unity ~ 94% of the time, indicating that the water column was unstratified. However, along the south shoreline (average depth ~3.5 m) and navigation channel (average depth ~9 m) (see Figure 3-5a and 5c in the main text), the small water temperature variation (<1 °C) can cause stratification ($0 \leq \Delta\rho \leq 0.47$ k$g\ m^{-3}$) which on average occurred 35% and 69% of the time, respectively. There are locations in the middle and south and south-west region of the lake where stratification can occur all the time.

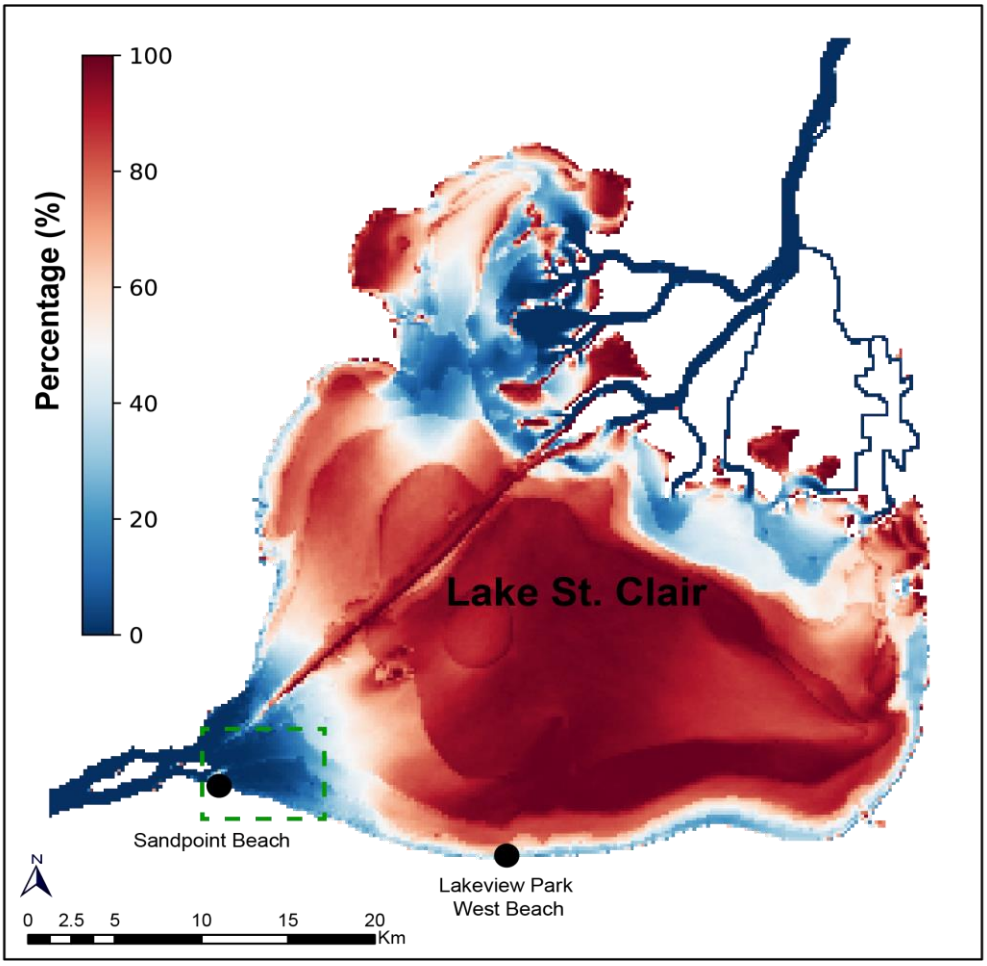

Figure 3-S3. Map of Lake St. Clair showing the percentage of the time that stratification can occur in the water column based on the simulation results calculated densimetric Froude number.

References

Andutta, F.P., Ridd, P.V., Wolanski, E., 2013. The age and the flushing time of the Great Barrier Reef waters. Continental Shelf Research 53 11-19.

Deleersnijder, E., Beckers, J.-M., Delhez, E.J.M., 2006. On the behaviour of the residence time at the bottom of the mixed layer. Environmental Fluid Mechanics 6(6) 541-547.

Huziy, O., Sushama, L., 2017. Lake–river and lake–atmosphere interactions in a changing climate over Northeast Canada. Climate Dynamics 48(9) 3227-3246.

McPhedran, K., Seth, R., Bejankiwar, R., 2013. Occurrence and predictive correlations of *Escherichia coli* and Enterococci at Sandpoint beach (Lake St Clair), Windsor, Ontario and Holiday beach (Lake Erie), Amherstburg, Ontario. Water Quality Research Journal of Canada 48(1) 99-99.

Zhao, J., Rao, Y.R., Wassenaar, L.I., 2012. Numerical modeling of hydrodynamics and tracer dispersion during ice-free period in Lake Winnipeg. Journal of Great Lakes Research 38 147-157.

Appendix – 4 Supplementary Information of Chapter 4

4-1: Sensitivity analysis

For a complex model with long run times and numerous parameters, such as three dimensional high-resolution model for Lake St. Clair based on AEM3D, it is impractical, if not impossible, to run hundreds of thousands of simulations for testing the entire parameter space. First, the sensitivity of *E. coli* concentration to changes in decay rate is presented and then to evaluate how the parameters of decay function influence the calculated decay value Monte Carlo sensitivity analysis was used. The sensitivity of Y in respect to a parameter X , S_X^Y , is defined according to Jørgensen and Bendoricchio (2001) proposed equation as follow

$$S_Y^X = \left.\frac{\partial Y}{\partial X}\right|_{X_0} \times \frac{X_0}{Y_0} \qquad (5)$$

where Y_0 and X_0 are the values for the output and parameter respectively in the base model run. $\left.\partial Y/\partial X\right|_{X_0}$ is the partial derivative of the output Y with respect to an input factor X at the base value.

The sensitivity of *E. coli* concentration to the decay rate is illustrated in Figure 4-S1. Negative values mean that there is a negative correlation between *E. coli* concentration and decay rate. According to Figure 4-S1a, the sensitivity at the southern region of the lake and near the mouth of major tributaries and St. Clair branches is lower than the middle of the lake. These areas are where are expected to have higher *E. coli* concentration because of the added impact of local tributaries. That means that the correlation between sensitivity

and *E. coli* concertation is also negative. Figure 4-S1b shows the profile of the sensitivity in the sites located in the southern region of the lake. Low sensitivity at the shoreline indicated that small changes in the decay rate have a limited effect on the *E. coli* concentration. However, the decay rate can change between 0.5 to 1.3 d^{-1} within the simulation period. Maximum *E. coli* concentration using constant lower-end (0.5 d^{-1}) and high-end (1.3 d^{-1}) decay rate is illustrated as a bar graph in Figure 4-S1c for different site locations. As can be seen, although the sensitivity for sites S135, LS_WTP, RCM RIV, S136, and SP_WTP are lower than -1 the *E. coli* concentration using these two decay rates make differences between 19.7 to 51.3%. The results show that sensitivity to the decay rate is not uniform. Locations closer to sources that dominate the *E. coli* concentrations observed at that location are less sensitive to variation in decay rate than those farther away.

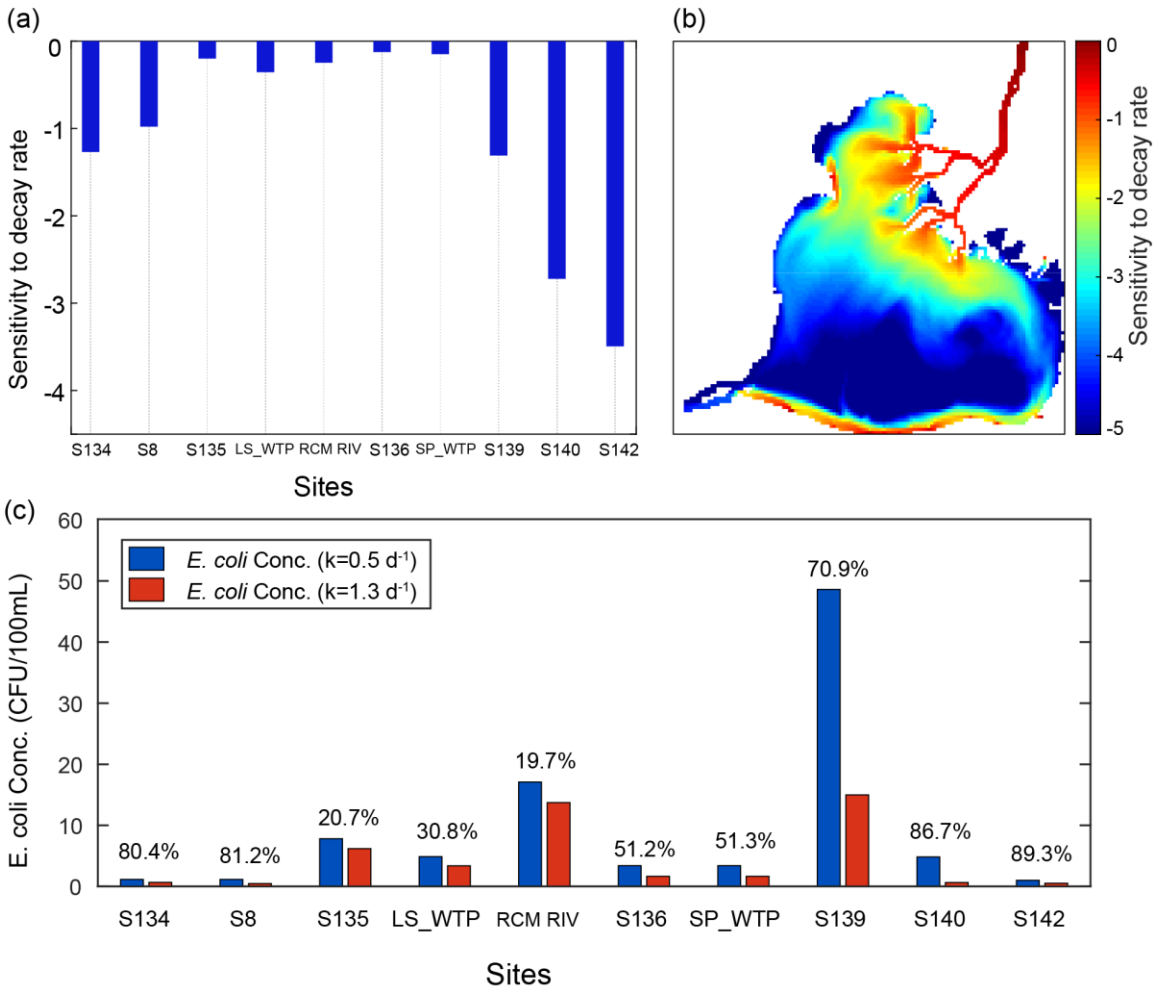

Figure 4-S1: Sensitivity of *E. coli* concentration to decay rate at a) different sites b) the whole lake, and plot presented in panel c) shows mean *E. coli* concentration using lower-end (0.5 d^{-1}) and upper-end (1.3 d^{-1}) constant decay rate. Values above the bar chart show with their relative difference in respect to values obtained in k=0.5 d^{-1}.

Parameters that affect the decay rate are listed in Table 4-S1 which $k_{d\,20}$ is the highest and θ_p is the lowest influential parameter. That can be explained by the higher sum of the square, higher correlation coefficient based on Pearson and Spearman. According to omega squared (ω^2 , a measure of effect size), 75% of the variance in the decay rate is attributable

to the dark mortality rate at 20 $^{\circ}C$ (k_{d20}). The decay rate is very sensitive to the k_{d20} and θ which are parameters in K_T which is the natural mortality or die-off rate due to water temperature according to Eq. 4. The sensitivity of the decay rate to its parameters as a function of temperature was tested, and the results are presented in Figure 4-S2. As can be seen, the sensitivity changed as a function of water temperature, with the most variation observed for θ and k_{d20}. The decay rate was most sensitive to θ and k_{d20} between water temperatures of about 20 – 30 $^{\circ}C$.

Table 4-S1: parameters that were included in the sensitivity analysis of the decay rate are sorted based on their higher degree of sensitivity to the lowest.

	Sum of Square	**F**	**P-value**	**Pearson**	**Spearman**	$\boldsymbol{\omega^2}$
k_{d20}	6.43	4.5E+05	0.0E+00	-0.87	-0.85	0.75
θ	1.35	9.5E+04	0.0E+00	-0.39	-0.38	0.16
$k_b(vis)$	0.32	2.3E+04	0.0E+00	-0.20	-0.19	0.04
ct_1	0.24	1.7E+04	0.0E+00	0.19	0.18	0.03
k_{p20}	0.18	1.3E+04	0.0E+00	-0.19	-0.19	0.02
μ_{max}	0.01	9.4E+02	5.6E-146	0.04	0.03	0.00
$k_b(UVA)$	0.01	4.4E+02	1.5E-81	-0.07	-0.08	0.00
ct_2	0.00	2.8E+02	2.3E-56	0.01	0.01	0.00
$k_b(UVB)$	0.00	6.6E+01	1.5E-15	0.02	0.02	0.00
θ_p	0.00	6.3E+01	6.4E-15	-0.02	-0.01	0.00

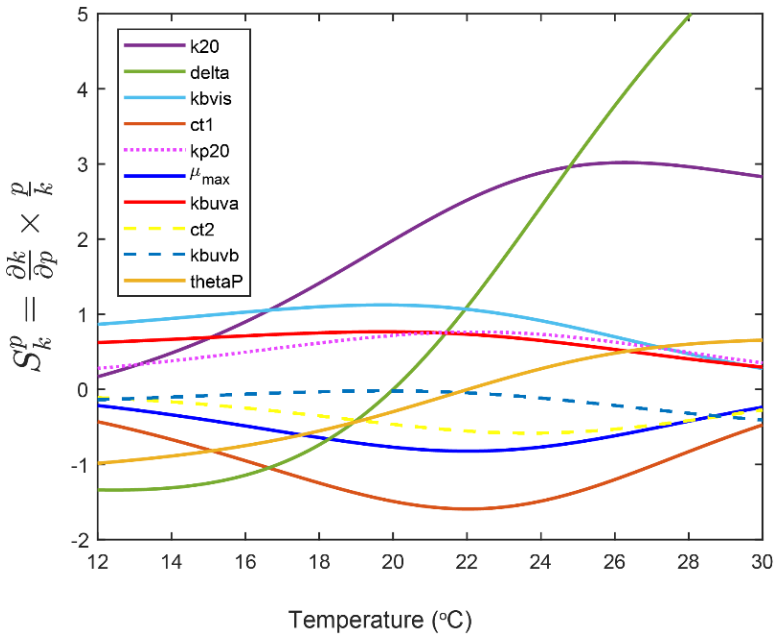

Figure 4-S2: Sensitivity of decay rate to its parameter in the different water temperature

References:

Jørgensen, S.E., Bendoricchio, G., 2001. Chapter 2 Concepts of modelling, Developments in Environmental Modelling. Elsevier, pp. 19-92.

Appendix – 5 Supplementary Information of Chapter 5

5-1: E. coli data

Table 5-S1 presents mean, maximum and the 90th percentile of *E. coli* concentration for different tributaries of Lake St. Clair

Table 5-S1: Descriptive statistics of *E. coli* concentration from the major and small tributaries of Lake St. Clair

Tributaries	90% Percentile	Max	Mean
St. Clair River	86	172	33
Thames River	377	1299	262
Sydenham River	1670	2270	769
Clinton River	445	1801	306
Pike Creek	1000	8100	574
Puce River	732	2200	295
Belle River	788	5600	414
Duck Creek	1300	8550	615
Moison Creek	1250	4600	530
Ruscom River	519	1600	211
Stoney Point	1120	3800	350
Little Creek	222	3000	124

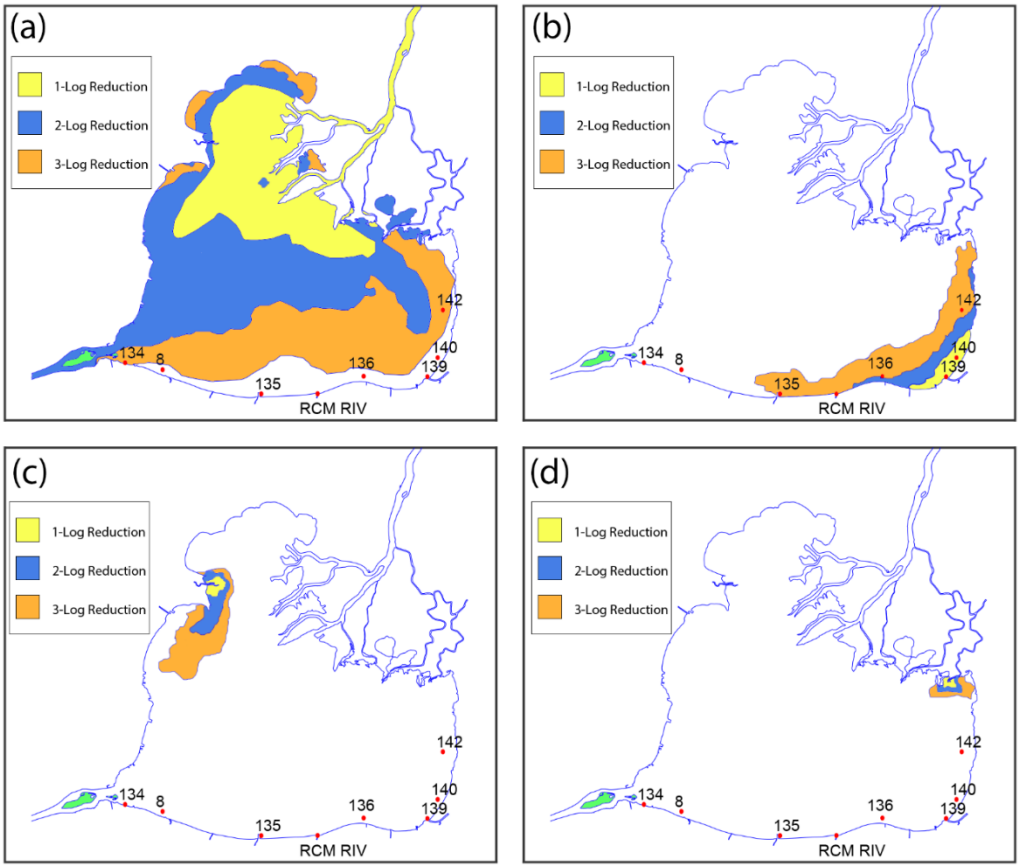

Figure 5-S1. The area in which *E. coli* from major tributaries of Lake St. Clair reduced based on log-reduction. a) St. Clair River b) Thames River c) Clinton River d) Sydenham River

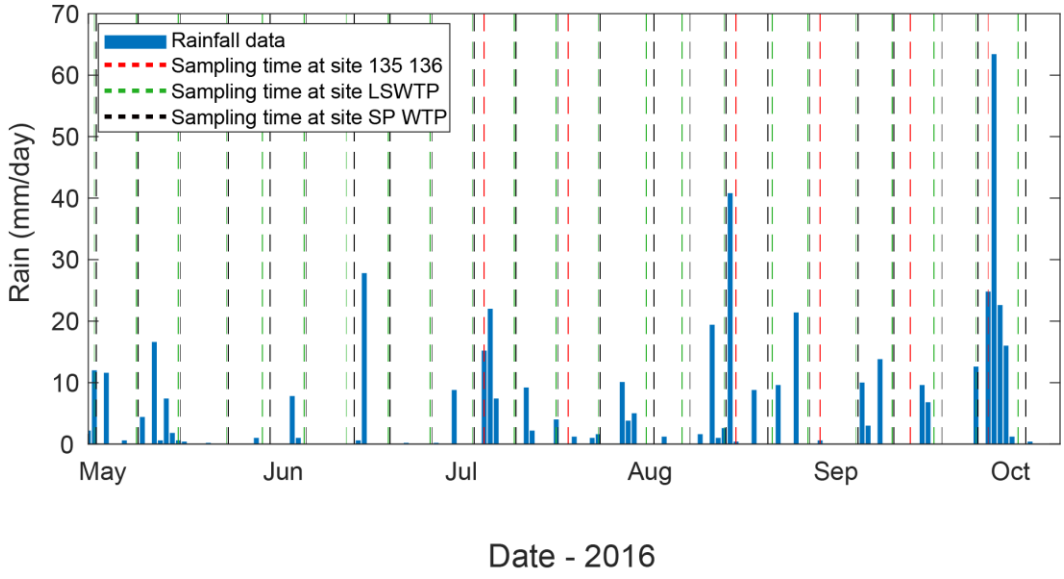

Figure 5-S2. Rainfall data for summer 2016. The vertical lines in different colours show the time of sampling at different sites.

Appendix – 6 Supplementary Information of Chapter 6

6-1: High resolutoin image of Belle River area

Figure 6-S1: Areal image of Lakeview Park West Beach (LP), Belle River and Marina.

VITA AUCTORIS

NAME:	Mohammad Madani
PLACE OF BIRTH:	Saghez, Kurdistan, Iran
YEAR OF BIRTH:	1987
EDUCATION:	• Amirkabir University of Technology, B.Sc. Chemical Engineering, Tehran, Iran., 2009 • Amirkabir University of Technology, B.Sc. Polymer Engineering, Tehran, Iran., 2011 • Amirkabir University of Technology, M.Sc. Chemical Engineering, Tehran, Iran., 2011
PROFESSIONAL EXPERIENCE	• Research Assistant, Dept. of Civil & Environmental Engineering, University of Windsor, Jan 2017 – August 2020 • Hydrology and Ecological Modeller, DHI, Cambridge, April 2020 – Present
LANGUAGES:	English, Kurdish, Persian