

Advancing Microbial Risk Assessments of Subsurface Water Sources

by

Alex Ho Shing Chik

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2020

© Alex Ho Shing Chik 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Ronald W. Harvey
Research Hydrologist
United States Geological Survey

Co-supervisors: Jack F. Schijven
Professor
Faculty of Geosciences,
Utrecht University

Alfred P. Blaschke
Professor
Inst. of Hydr. & Water Resource Mgmt.,
Vienna University of Technology

Monica B. Emelko
Professor
Dept. of Civil & Env. Engineering,
University of Waterloo

Internal Member: Seyed M. Hassanizadeh
Professor
Faculty of Engineering,
University of Waterloo (by cross-appointment)

Internal-External Members: Kirsten M. Müller
Professor
Department of Biology,
University of Waterloo

Seyed M. Hassanizadeh
Professor
Faculty of Geosciences,
Utrecht University

Matthias Zessner
Professor
Inst. of Water Quality & Resource Mgmt.,
Vienna University of Technology

Promotions Committee: Marc F.P. Bierkens
Professor
Faculty of Geosciences,
Utrecht University

Veerle Cnudde
Professor
Faculty of Geosciences,
Utrecht University

Ana Maria de Roda Husman
Professor
Faculty of Veterinary Medicine,
Utrecht University

Caroline P. Slomp
Professor
Faculty of Geosciences,
Utrecht University

Jeannot A. Trampert
Professor
Faculty of Geosciences,
Utrecht University

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Alex Ho Shing Chik was the primary author for all chapters of this thesis. Chapters 1 and 6 were not written for publication. This thesis consists in part of four manuscripts written for publication. Exceptions to sole authorship of material are described below.

Research presented in Chapter 2:

This research was conducted at the University of Waterloo by Alex Chik under the supervision of Dr. Monica Emelko and Dr. Philip Schmidt. Dr. Philip Schmidt, Dr. Monica Emelko, and Alex Chik contributed to study design. Alex Chik, Dr. Philip Schmidt, Dr. Monica Emelko analyzed the data. Alex Chik prepared the draft manuscripts, which all co-authors contributed intellectual input on. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [CRDPJ 484588 - 15] and the City of Calgary. We thank the staff of the City of Calgary for providing the microbial water quality dataset for analysis. Dr. Alfred Paul Blaschke (Vienna University of Technology, Vienna, Austria), Dr. Jack F. Schijven (RIVM & Utrecht Universiteit, Netherlands), and Dr. Norma J. Ruecker (City of Calgary, AB, Canada) and two anonymous reviewers are thanked for their review of this manuscript.

Citation:

Chik, A.H.S., Schmidt, P.J., and Emelko, M.B. (2018) Learning Something From Nothing: The Critical Importance of Rethinking Microbial Non-detects. *Frontiers in Microbiology* 9:2304. doi: 10.3389/fmicb.2018.02304

Research presented in Chapter 3:

This research was conducted at the University of Waterloo by Alex Chik under the supervision of Dr. Monica Emelko, Dr. Philip Schmidt, Dr. Jack F. Schijven, and Dr. Alfred Paul Blaschke. Alex Chik contributed to study design and analyzed the data. Alex Chik prepared the draft manuscripts, which all co-authors contributed intellectual input on.

Citation:

Chik, A.H.S., Schmidt, P.J., Emelko, M.B., Schijven, J.F., and Blaschke A.P. (2020) An End Justified by Means: Ensuring Adequate Treatment of Protozoa by Precise Mean Concentration Estimation. In preparation for submission to *Journal of Water and Health*.

Research presented in Chapters 4 and 5:

Alex Chik and Dr. Monica Emelko conceived this study and designed the study. Alex Chik performed the field study, analyzed the data and prepared the draft manuscripts, which all co-authors contributed intellectual input on. LuminUltra Technologies Ltd. providing the equipment and reagents necessary for the research presented in Chapter 3. For Chapter 4, Kaitlyn E. O’Sullivan helped with field and laboratory work and performed exploratory data analysis. Dr. Domenico Savio, Dr. Andreas H. Farnleitner, Dr. William B. Anderson assisted with the interpretation of the 16S rRNA gene amplicon sequencing data. Dr. Monica Emelko, Dr. Jack Schijven, and Dr. Alfred Paul Blaschke provided project oversight. Alex Chik wrote the first draft of both manuscripts. All coauthors of the respective manuscripts contributed to revisions, read, and approved the manuscripts. Terry Ridgway, Mark Sobon, Dr. Maria Mesquita, Benjamin J.M. Beelen, Kaitlyn E. O’Sullivan, and Caitlin Wong are acknowledged for their help throughout the sampling campaigns. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number EGP 507411-16].

Citations:

Chik, A.H.S., Emelko, M.B., Blaschke, A.P., Schijven, J.F. (2020) Illuminating Subsurface Microbial Water Quality Patterns Using Adenosine Triphosphate and Dynamic Time Warping Approaches. *Groundwater Monitoring and Remediation*. doi: 10.1111/gwmmr.12397

Chik, A.H.S., Emelko, M.B., Anderson, W.B., O’Sullivan, K.E., Savio, D., Farnleitner, A.H., Blaschke, A.P., Schijven, J.F. (2020) Evaluation of Groundwater Microbial Community Composition to Inform Waterborne Pathogen Vulnerability Assessments. Submitted to *Science of the Total Environment*.

As lead author of these four chapters, I was responsible for contributing to conceptualizing study design, carrying out data collection and analysis, and drafting and submitting manuscripts. My co-authors provided guidance during various steps of the research and provided feedback on draft manuscripts. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abstract

Waterborne pathogens are significant, ubiquitous threats to public health. Thus, microbial water quality evaluations comprise a critical component of the multi-barrier approach to delivering safe drinking water. These assessments underpin the selection, design, and management of drinking water treatment processes. However, the selection of the right combination of tools from an ever-expanding repertoire is essential. This dissertation informs prudent water quality monitoring using existing and emerging microbial tools to accurately characterize microbial water quality and associated risks, especially in sources derived from subsurface environments.

Microbial non-detects cannot be directly construed as indicative of low microbial concentrations due to inherent limitations of sampling and analyzing large volumes of water and imperfect microbial analytical methods. Existing conventions to report microbial non-detects as a measured concentration less than one microorganism within the analyzed volume (the purported method detection limit) were demonstrated to be misleading; handling these values as mathematically “censored” concentrations were shown to result in bias (Chapter 2). Appropriate reporting conventions and statistical approaches were recommended to support the accurate portrayal of microbial non-detects. The minimization of microbial sampling effort while maintaining adequate precision is a key consideration for monitoring program design. However, the imprecision of information about concentration associated with small sample sizes is seldom explicitly quantified. Using simulated protozoan monitoring data, the attainable precision of the estimated mean protozoan concentration under different hypothetical sampling scenarios was evaluated. A framework was developed to quantify precision and contrast the relative merits of additional sample collection for protozoan enumeration from the source versus characterization of method analytical recovery (Chapter 3).

In microbial groundwater quality evaluations, the need for sufficient well purging to obtain representative samples of microorganisms suspended in aquifer pore water without artefacts attributable to well-related biofilms has been widely recognized. Adenosine triphosphate (ATP) was therefore evaluated as a rapid, field-ready biochemical indicator of microbial water quality changes (Chapter 4). Supported by concurrently measured microbial water quality parameters, ATP measurements exhibited phenomena reflective of time-limited (bio)particle transport behaviour. Microbial groundwater quality assessments must therefore be designed using approaches that are necessarily different from those used to describe dissolved solute transport behaviour. A subsequent focused investigation of one biomolecular tool—bacterial community analysis based on 16S rRNA gene amplicon sequencing (Chapter 5) demonstrated its utility

for identifying fecal source and surface connectivity indicators (e.g., cyanobacteria). Factors contributing to bacterial community variations were examined. Collectively, these assessments indicated that an appropriate suite of microbiological tools can be concurrently utilized to overcome the challenges of spatial heterogeneity and dynamic hydrogeological conditions to meaningfully characterize microbial water quality at the aquifer scale.

In this dissertation, existing and emerging microbial tools to support groundwater vulnerability assessments to fecal pathogen intrusion were critically examined. The persistent need to consider the “fit-for-purpose” ability of these various tools to support microbial water quality evaluation is emphasized. Moreover, this dissertation underscores the complementary use of these microbial tools to inform groundwater vulnerability assessments; no single tool will entirely capture the elusive, multi-faceted nature of subsurface microbial water quality.

Acknowledgements

I would like to express my sincere gratitude to my co-supervisors Paul, Monica, and Jack for the continuous support of my Ph.D. studies and related research. As I have learned throughout my studies, the search for the right words with the right meanings is often easier said than done. In this case, the official title of “co-supervisor” really fails to capture the patience, motivation, mentorship and support I have received throughout my studies. I am truly blessed to not have had just one, but three, incredible advisors and mentors for my Ph.D. studies.

My sincere thanks also goes to Dr. Andreas Farnleitner, Dr. Domenico Savio, and the broader Interuniversity Cooperation Center for Water and Health team, as well as Dr. Philip Schmidt, Dr. Maria Mesquita, Dr. William B. Anderson, Dr. Kirsten Müller for many thought-provoking and insightful discussions. I thank my fellow lab-/office mates, colleagues, and friends—especially Dr. Margaret Stevenson, Dr. Julia Derx, Dr. Inge van Driezum, Dr. Mike Kittridge, Jill Crumb, Thadsha Chandrakumaran, Dr. Mark Spanjers, Dr. Shoeleh Shams, Dr. Chao Jin, Celine Vandenberg, Elyse Walker-Batista, Amy Yang, Kristina Lee, Dana Herriman—for the inspiring academic (and non-academic) discussions, our shared love for warm/caffeinated beverages and for all the fun we have had over the last five years. A special thanks for the field and laboratory support I’ve received, without which my research would not be possible: Mark Sobon, Terry Ridgway, Dylan Klazinga, Benjamin Beelen, Caitlin Wong, and Kaitlyn O’Sullivan.

I am grateful for all the academic institutions involved: Utrecht University, Vienna University of Technology, and University of Waterloo for agreeing to this one-of-a-kind international collaboration, and especially the University of Waterloo for granting support of parental leave. During my Ph.D. program at University of Waterloo, I was honored to receive the Carl A. Pollock Postgraduate Scholarship and the Queen Elizabeth II Graduate Scholarship in Science and Technology, which are hereby acknowledged. I also thank the Departmental staff of each of the academic institutions who have helped me in various ways during my studies.

I would like to thank my family and friends. I could not have gotten to this point without the immense patience and countless sacrifices from my wife Sonya. My children, Alianna Mei Ning and Brendan Zhi Ho, have continually provided the requisite breaks from academics and the motivation to finish my degree expediently.

Dedication

Dedicated to my wife, Sonya M. Chik,
for your unending understanding of all my flaws,
your support,
and your affection.

And, in honour of my grandfather,
who bestowed on me
his passion for lifelong learning.

Table of Contents

List of Figures	xxi
List of Tables	xxv
1 Introduction	1
1.1 Background	3
1.2 Microbial water quality evaluations of subsurface drinking water sources	4
2 Learning something from nothing: The critical importance of rethinking microbial non-detects	11
2.1 Summary	13
2.2 Introduction	13
2.3 State of scientific practice	15
2.3.1 NDs in analytical chemistry	16
2.3.2 NDs in enumeration-based microbial methods	17
2.4 Results: Evidence that microbial non-detects are not censored data . . .	19
2.4.1 Occurrences of microbial NDs are not solely a function of analyte concentrations	19
2.4.2 Uncertainty in concentration estimates precludes MDL-based interpretation of results	21
2.4.3 Censoring in detection- and enumeration-based microbial methods	21
2.5 Implications for policy and practice	23
2.6 Conclusions	27

3 An end justified by means: Ensuring adequate treatment of protozoa by precise mean concentration estimation **29**

3.1 Summary 31

3.2 Introduction 31

3.3 Methods 33

 3.3.1 Estimation of practically relevant parameter values of hierarchical model using historical protozoan monitoring program data 33

 3.3.2 Bayesian retrospective power analysis 35

 3.3.3 Goals of precision related to the posterior distribution of mean protozoan concentration 35

3.4 Results 37

 3.4.1 Point-based goal of precision based on the mode of the posterior mean concentration distribution 37

 3.4.2 Interval-based goals of precision based on the 95% HDI width of the posterior mean concentration distribution 37

 3.4.3 Impact of analytical recovery on attaining precision mean concentration estimates 43

3.5 Discussion 43

 3.5.1 Relatively few samples may be sufficient for precise mean concentration estimation 43

 3.5.2 Minimum treatment performance targets can effectively alleviate implications associated with low, highly uncertain mean concentration estimates 46

 3.5.3 Point- and interval-based goals reflect different aspects of precision 47

 3.5.4 A broadly applicable framework for evaluating the level of precision attainable based on monitoring program design to support QMRA efforts 47

 3.5.5 Sample size enlargement can enhance precision in mean concentration estimates more than performing additional internally seeded matrix-spikes to quantify analytical recovery 48

3.6 Conclusions 49

4	Illuminating subsurface microbial water quality patterns using adenosine triphosphate and dynamic time warping approaches	51
4.1	Summary	53
4.2	Introduction	53
4.3	Materials and methods	56
4.3.1	Study site	56
4.3.2	Field and laboratory analyses	57
4.3.3	Determination of ATP concentrations	58
4.3.4	Non-metric multidimensional scaling of ATP time series using DTW distances	59
4.4	Results	61
4.4.1	Physical and chemical water quality characteristics	61
4.4.2	Adenosine triphosphate measurements	61
4.4.3	Additional microbial water quality parameters	65
4.4.4	Non-metric multidimensional scaling of ATP time series based on DTW distances	68
4.5	Discussion	69
4.5.1	Physical and chemical indicators do not reflect purging adequacy for microbial water quality evaluations	69
4.5.2	ATP measurements can be useful for indicating aquifer-representative microbial water quality	70
4.5.3	Possible insights related to the spatio-temporal scales of ATP patterns and associated subsurface biofilm behaviour	72
4.6	Conclusions	73
5	Evaluation of groundwater bacterial community composition to inform waterborne pathogen vulnerability assessments	75
5.1	Summary	77
5.2	Introduction	77
5.3	Materials and methods	79

5.3.1	Site location	79
5.3.2	Experimental design	80
5.3.3	Sample collection	81
5.3.4	16S rRNA gene amplicon sequencing	82
5.3.5	Microbiome data analysis	83
5.4	Results	85
5.4.1	The core microbial community: dominant taxonomic groups observed in groundwater samples	85
5.4.2	Seasonal/event-based influences and spatial heterogeneity as key drivers of bacterial community diversity	86
5.4.3	Identification of relevant OTUs through differential abundance testing	88
5.5	Discussion	91
5.5.1	Bacterial diversity and core community analysis as initial indicators for surface-water groundwater interactions	91
5.5.2	<i>Escherichia</i> sequence reads from 16S rRNA gene sequencing require judicious interpretation	92
5.5.3	Anaerobic taxa sequences as potential vulnerability indicators within an aerobic environment	93
5.5.4	<i>Betaproteobacteria</i> sequences as potential indicators of seasonal subsurface processes	94
5.5.5	<i>Bacillales</i> and <i>Sphingomonadales</i> sequences as potential indicators of adequate well purging	95
5.5.6	Scarce <i>Cyanobacteria</i> and <i>Flavobacteriia</i> sequences as potential lines of evidence of limited surface water-ground-water bacterial transport processes	96
5.6	Conclusions	97
6	Synthesis of findings and recommendations	99
6.1	Synthesis	101

6.2	Key contributions	102
6.2.1	Handling microbial non-detects without incurring bias	102
6.2.2	Source water monitoring program design to ensure adequate treatment of protozoan pathogens	103
6.2.3	Microbial water quality observed during purge water abstraction from monitoring wells are subject to time-limited (bio)colloid mobilization and transport mechanisms	105
6.2.4	The use and limitations of “next-generation sequencing” biomolecular tools for supporting vulnerability assessments of subsurface water sources to pathogen intrusion	107
6.3	General conclusion and research outlook	108
	References	111
	Appendices	131
A	Supporting information for <i>Learning something from nothing: The critical importance of rethinking microbial non-detects</i>	133
A1	Calculation of the probability of a non-detect observation	134
A2	Evaluation of the posterior probability density function of the microbial concentration based on a single observation	135
A3	Fitting a Poisson log-normal distribution to <i>Giardia</i> cyst raw data assuming 100% analytical recovery	136
B	Supporting information for <i>An end justified by means: Ensuring adequate treatment of protozoa by precise mean concentration estimation</i>	139
B1	Hierarchical model description	140
C	Supporting information for <i>Illuminating subsurface microbial water quality patterns using adenosine triphosphate and dynamic time warping approaches</i>	141
C1	Groundwater levels	142

C2	Illustrative example of dynamic time warping (DTW) distance determination	143
C3	Background physical and chemical water quality	148
C4	Supporting microbial water quality parameters evaluated	150
D	Supporting information for <i>Evaluation of groundwater bacterial community composition to inform waterborne pathogen vulnerability assessments</i>	153
D1	Groundwater levels	154
D2	Water quality parameters evaluated along with 16S rRNA gene amplicon sequencing samples	154
D3	Quality control samples	155
D4	Bacterial core community analysis	157
D5	Alpha-diversity measures associated with 16S rRNA gene amplicon sequencing samples	159
D6	Constrained analysis of principal coordinates	159
D7	Identification of potentially relevant indicator taxa	160
	List of Publications	167

List of Figures

1.1	Conceptual model for the determination of minimum drinking water monitoring and treatment requirements for subsurface water source . . .	6
2.1	Probability of a non-detect observation as a function of organism concentration and (a) sample volumes, and (b) analytical recovery	20
2.2	Posterior probability density function (PDF) characterizing uncertainty in the true concentration given (a) an ND observation, and (b) an observation of two microorganisms, each based on a 1.0-L sample	22
3.1	Posterior modes and 95% highest density intervals (error bars) of mean concentration across 100 simulations	38
3.1	(continued)	39
3.2	Estimated 95% Bayesian HDI widths as a function of sample size with paired recovery data.	42
3.3	Histogram of posterior mode of the mean concentration and 95% HDI width estimated from 100 simulations	45
4.1	Study site and monitoring well locations	56
4.2	Stabilization history of groundwater quality parameters monitored, expressed as the coefficient of variation evaluated for running windows of three consecutive measurements throughout purging	62
4.3	Stabilization of ATP concentrations throughout purging	63
4.4	Box-and-whisker plots of relative differences between WO77 and WO78 paired ATP measurements throughout concurrent purging activities . .	64

4.5	Normalized ATP concentrations observed throughout well purging activities	67
4.6	(Dis)similarities between all ATP time series portrayed through non-metric multidimensional scaling (NMDS)	68
5.1	Total abundance of OTU sequence reads across all samples plotted against their prevalence	86
5.2	Summary of relative abundance of bacterial taxa on phylum level, by well membership (WO77 vs. WO78) organized by sampling occasion. . .	87
5.3	Non-metric multidimensional scaling (NMDS) ordination of purge water samples using Bray-Curtis dissimilarities	87
5.4	Differential abundance of various OTUs by class, in order of increasing \log_{10} -fold differences in comparison to the base level of each factor across (a) the well factor, (b) the purging stage factor (early vs. intermediate vs. late stages), and (c) the seasonal factor	89
5.4	(continued)	90
B1	Hierarchical model used to estimate posterior distribution of mean protozoan concentrations (μ)	140
C1	Groundwater levels measured manually in monitoring wells WO77 and WO78 and continuously in well WO75	142
C2	Optimal dynamic time warping path based on constraints imposed (i.e., slanted band, window size of 3, open-ended alignment, symmetric2 step pattern).	145
C3	Example of time series aligned by dynamic time warping	147
C4	Example of IRB-BART TM observed from sampling occasion conducted April 24, 2018 (WO78)	150
C5	Example of SLYM-BART TM observed from sampling occasion conducted April 24, 2018 (WO77)	150
D1	Differential abundance of various OTUs by class, in order of decreasing \log_{10} -fold differences in comparison to the negative controls as the base level	155

D2	Empirical taxa abundance distribution to determine the core community across all purge water samples	157
D3	Summary of relative abundance of bacterial taxa on phylum level, core community only	157
D4	Summary of <i>Proteobacteria</i> taxa on family level.	158
D5	Summary of <i>Proteobacteria</i> taxa on family level, core community only (present in at least 91% of samples).	158
D6	Constrained analysis of principal coordinates (CAP) ordination of purge water samples using Bray-Curtis dissimilarities	159

List of Tables

- 2.1 Summary of raw water samples analyzed for *Giardia* cysts from City of Calgary, AB, Canada - October, 2012. 24
- 2.2 Comparison of *Giardia* cyst concentration statistics obtained using various approaches for handling microbial NDs. 26
- 3.1 Estimated power to achieve goal of precision based on the posterior mode of the mean concentration, as a function of sample size 40
- 3.2 Estimated power to achieve goal of precision based on 95% Bayesian HDI widths, as a function of sample size 41
- 3.3 Estimated power to achieve goal of precision as a function of sample size for the *Giardia* concentration parameters, with different mean recoveries simulated 44
- 4.1 Paired t-test results comparing ATP concentrations between WO77 and WO78 64
- C1 Adenosine triphosphate concentrations measured in piezometer WO77, April 2018 144
- C2 Matrix of Euclidean distances between each pair of normalized ATP concentrations 146
- C3 Summary statistics of background physical and chemical water quality characteristics 149
- C4 Summary statistics of flow cytometry and BARTTM results 151
- D1 Summary statistics of physical and chemical water quality characteristics (subset of parameters from Chik *et al.*, [in review]) 154

D2	Summary statistics for α -diversity measures	159
D3	Top 20 OTUs ranked in order of decreasing significance of the well membership factor coefficient	161
D4	Top 20 OTUs ranked in order of decreasing significance of the first purging stage factor coefficient	162
D5	Top 20 OTUs ranked in order of decreasing significance of the second purging stage factor coefficient	163
D6	Top 20 OTUs ranked in order of decreasing significance of the first seasonal factor coefficient	164
D7	Top 20 OTUs ranked in order of decreasing significance of the second seasonal factor coefficient	165

Chapter 1

Introduction

1.1 Background

Advances in the biological sciences made over the past century—and the technological progress that made them possible—have profoundly influenced the water industry in our pursuit of delivering microbiologically safe drinking water. From most-probable-number (MPN) culture-based methods to enzyme-reaction based ColilertTM tests, and from early polymerase chain reaction (PCR) technology to next generation sequencing (NGS) techniques, how microbial water quality evaluations are performed in practice continue to evolve along with approaches for interpreting and analyzing data emanating from these tools. Consequently, practitioners responsible for making decisions related to treatment infrastructure design, operations, and management are often left with the overwhelming task of choosing and deploying the right combination of tools from an ever-expanding repertoire.

In this thesis, fresh perspectives and insights are provided concerning how we, as an industry, must adapt in light of these advances to facilitate evidence-based decision making. The continued need for critical evaluations of the use of microbial data, as well as the tools and approaches generated therefrom, is underscored. Their incorporation into source water quality monitoring programs must be fit-for-purpose and contextualized based on known and/or theorized scientific phenomena. This research addresses potentially overlooked but critical issues, such as the number of samples required for accurate microbial concentration estimation, and the handling of microbial non-detects. Opportunities where emerging microbial tools (e.g., 16S rRNA gene amplicon sequencing) and methods borrowed from other disciplines (e.g., dynamic time warping) can support conventional source microbial water quality assessments were also explored. Their use to complement existing tools—rather than outright replacing them—is emphasized. While elements of this thesis is focused on subsurface water sources, most of the concepts discussed in this work are broadly applicable and transferable to the evaluation of microbial water quality for any source water. In this age of “big data” where more data can be generated than can be meaningfully interpreted, opportunities to advance microbial water quality monitoring and ensuing risk assessments are met with cautious optimism.

1.2 Microbial water quality evaluations of subsurface drinking water sources

Although microbial health risks associated with drinking water drawn from subsurface sources are generally perceived to be lower than those derived from surface sources (WHO, 2017), disease outbreaks associated with drinking water drawn from groundwater sources and tragedies like that in Walkerton, Ontario, Canada (2000) are stark reminders that waterborne pathogens in these water sources can still pose a significant, ubiquitous threat to public health (O'Connor, 2002). This year marks the 20th anniversary of the Walkerton tragedy; both science and policy in Canada and internationally have since evolved substantially. Although Canada remains without legally enforceable national drinking water standards, provincial legislators have responded with new regulations, technological changes, and investments intended to better safeguard public health (Benidickson, 2017). The imminent promulgation of the new Terms of Reference for the *Determination of Treatment Requirements Well-Based Municipal Systems* to replace the *Groundwater-Under-Direct-Influence of Surface Water* Terms of Reference (2001) in Ontario reflects an industry-wide paradigm shift from site vulnerability assessments based primarily on hydrogeological features/indicators (e.g., setback distances away from surface water features, unconfined aquifer settings, etc.) towards assessments prioritizing microbial water quality (e.g., presence of waterborne pathogens or fecal indicators).

While drinking water obtained from surface water sources intuitively requires an additional level of precautionary protection (i.e., a question of “when” rather than “if” contamination events occur), the presence of microbial pathogens in *any* source water—be it groundwater or riverbank filtrate or a “pristine” stream—underpins the decision to provide appropriate and adequate forms of treatment. Accordingly, the vulnerability of a drinking water source to pathogen intrusion must be evaluated. These evaluations are herein collectively referred to in this dissertation as “pathogen vulnerability assessments”. Pathogen vulnerability assessments are typically conducted by the water purveyor to demonstrate that adequate and appropriate treatment is provided for the range of microbial source water qualities observed and anticipated. Due to the inherent limitations of microbial water quality monitoring, selected “index pathogens” representing waterborne pathogenic viruses, bacteria, and enteric protozoa are often monitored instead (Schijven et al., 2011). When these pathogens are not detected, extra vigilance is also warranted for sources that are susceptible to fecal contamination. Upon establishing either or both of these microbial lines of evidence, hydrogeological

indications of source water vulnerability to pathogen intrusion becomes a peripheral consideration to the determination of treatment requirements, owing to relative uncertainty of pathogen transport processes in the subsurface (e.g. Yates et al., 1988; Ginn et al., 2005; Tufenkji, 2007; Emelko and Tufenkji, 2010; Bradford et al., 2015). Given that a minimum level of disinfection is mandated in Ontario (Ontario Ministry of Environment, Conservation and Parks, 2019) that inherently mitigates public health risks attributable to a range of bacterial and viral pathogens, the determination of *additional* treatment requirements beyond a baseline level of disinfection would necessarily imply the implementation of proactive intensive monitoring or treatment processes appropriate and proportionate to address the protozoan pathogen risk at hand.

It is within this decision-making framework that key research questions of practical importance were identified and explored in this thesis (Figure 1.1). The rest of this thesis comprises of four (4) manuscripts currently submitted or published in peer-reviewed journals that addresses the identified research questions (RQ):

- RQ1. How have microbial non-detects been reported, interpreted, and analyzed, and do some approaches lead to bias?
- RQ2. What level of precision in the estimation of mean microbial concentrations is attainable and which aspects of microbial monitoring program design are most influential to improving or compromising precision?
- RQ3. What additional considerations of the subsurface environment are necessary to indicate that representative microbial water quality sampling of aquifer water quality is achieved, and what tools are available to support these considerations?
- RQ4. How can bacterial community analysis using emerging biomolecular tools such as 16S rRNA gene amplicon sequencing be used to inform vulnerability assessments?

First, microbial non-detects and their use to support concentration estimation were critically examined. A critical review of current approaches used to report and handle non-detects was performed. It was investigated whether some of the approaches would lead to bias. Specifically, non-detects are commonly reported as values below a detection threshold (e.g. <1 microorganism/L) and are interpreted as measured concentrations. Given this interpretation, “censored-data” analysis approaches have become increasingly adopted for handling of microbial non-detects. Notably, a markedly different set of approaches has also been suggested for the handling of microbial non-detects. These approaches were examined in Chapter 2 to provide guidance on appropriate handling of microbial non-detects, therefore answering RQ1.

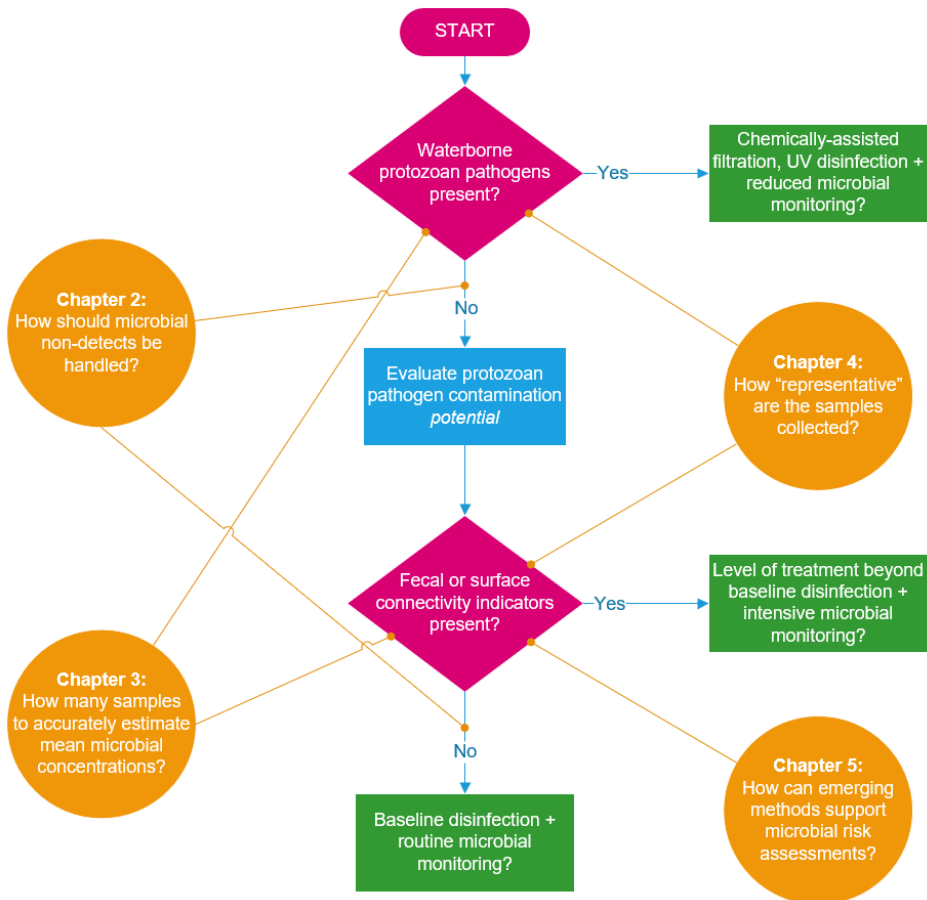


Figure 1.1: Conceptual model for the determination of minimum drinking water monitoring and treatment requirements for subsurface water sources. Green rectangles indicate treatment/monitoring decisions that will vary depending on jurisdictional drinking water policy contexts.

The attainable precision of microbial concentration estimates (and particularly those of protozoan pathogens) is often limited by the monitoring program design. The minimization of sampling effort without substantially compromising accuracy is often a key consideration for many monitoring programs and policy requirements. However, the degree to which precision has been compromised is often unknown, but can be quantified retrospectively using simulated data generated from estimated parameters. Statistical power was evaluated based on the proportion of simulations successfully rendering adequately accurate estimates of mean concentration. To answer RQ2, a statistical framework for exploring the relative merits of enlarging sample size and the quantification of analytical recovery (with the latter a key consideration of methods used to enumerate protozoan pathogens) was applied. Through a proof-of-concept demonstration of this type of analysis to an extensive dataset from the City of Calgary, the utility of such an approach to inform microbial monitoring program design on the basis of statistical power is investigated in Chapter 3.

In Chapters 4 and 5, additional microbial lines of evidence to support evaluations of (subsurface) source water vulnerability to pathogen intrusion were investigated using emerging microbial tools and approaches. Adenosine triphosphate (ATP) patterns were monitored throughout well purging activities and investigated as an indicator of microbial water quality change and the sufficiency of purging in Chapter 4. The presence of subsurface biofilms in the vicinity of the well are known to potentially introduce microbial artefacts that can influence the acquisition of—and obscure the subsequent interpretation of—microbial water quality samples representative of the aquifer. A novel, parameter-free approach of time series analysis borrowed from speech recognition was used to gain additional insights about spatio-temporal dynamics of the ATP patterns, along with implications for microbial water quality monitoring program implementation and interpretations. In this work to answer RQ3, the inadequacy of physical and chemical water quality parameters for establishing the sufficiency of well purging effort for microbial water quality evaluation is highlighted.

Next Generation Sequencing (NGS) biomolecular methods are another set of emerging tools that have been increasingly advocated to support source microbial water quality evaluations. Their reduction in costs over the past decades have made them attractive tools to consider in support of conventional enumeration- and/or culture-based methods. Specifically, 16S rRNA gene amplicon sequencing for microbial community analysis was investigated as a potential tool to establish lines of evidence for either microbiological source (i.e., indicators linked to fecal contamination), or transport (i.e., indicators that reflect surface connectivity). The sequence counts emanating from 16S rRNA gene amplicon sequencing analysis was systematically examined using a nega-

tive binomial generalized linear model (NB GLM) along major gradients to identify microbial taxa of relevance. The use and limitations of this biomolecular method are presented and discussed to answer RQ4 in Chapter 5.

In summary, the goals of this thesis were to:

1. Address handling of microbial non-detects:
 - Critically review current approaches used to report and handle non-detects
 - Investigate whether some approaches lead to bias
 - Provide guidance on how to handle microbial non-detects
2. Evaluate the influence of microbial monitoring program design factors to the estimation of mean microbial concentrations:
 - Develop a framework to quantify how sampling effort (i.e., number of samples and number of matrix spike samples) influence the estimation of mean microbial concentrations
 - Evaluate the relative merits of increasing sampling intensity and quantifying analytical recovery
3. Investigate ATP as a rapid sentinel of microbial water quality change in the subsurface
 - Examine the behaviour of ATP concentrations throughout well purging activities at field scale
 - Compare ATP fluctuations to the stabilization of physical and chemical water quality parameters
 - Explore ATP patterns to elucidate potential subsurface microbial dynamics
4. Examine the utility of 16S rRNA as a rapid biomolecular tool to complement existing culture- and enumeration-based microbial methods in pathogen vulnerability assessments
 - Determine the influence of key spatio-temporal gradients upon results emanating from field application of this tool
 - Corroborate results with known microbial indicators and screen for potentially relevant indicators of microbial water quality change
 - Interpret results collectively given the hydrogeological conditions at the site

These topics collectively advance microbial water quality assessments of subsurface water sources. Chapter 6 provides a brief summary of the various aspects covered in this dissertation and presents major conclusions and implications of this work for practice.

Chapter 2

Learning something from nothing: The critical importance of rethinking microbial non-detects

Alex H.S. Chik
Philip J. Schmidt
Monica B. Emelko

Reprinted from *Frontiers in Microbiology* (2018); 9:2304, under Creative Commons Attribution License from the copyright holders Chik, Schmidt, and Emelko

2.1 Summary

Accurate estimation of microbial concentrations is necessary to inform many important environmental science and public health decisions and regulations. Critically, widespread misconceptions about laboratory-reported microbial non-detects (NDs) have led to their erroneous description and handling as “censored” values. This ultimately compromises their interpretation and undermines efforts to describe and model microbial concentrations accurately. Herein, these misconceptions are dispelled by 1) discussing the critical differences between discrete microbial observations and continuous data acquired using analytical chemistry methodologies and 2) demonstrating the bias introduced by statistical approaches tailored for chemistry data and misapplied to discrete microbial data. Notably, these approaches especially preclude the accurate representation of low concentrations and those estimated using microbial methods with low or variable analytical recovery, which can be expected to result in non-detects. Techniques that account for the probabilistic relationship between observed data and underlying microbial concentrations have been widely demonstrated, and their necessity for handling non-detects (in a way which is consistent with the handling of positive observations) is underscored herein. Habitual reporting of raw microbial observations and sample sizes is proposed to facilitate accurate estimation and analysis of microbial concentrations.

2.2 Introduction

Whether describing pathogens in water or the density of red blood cells, the concentration of discrete objects cannot be measured directly. In these cases, concentration is estimated by enumerating or detecting the objects in finite sample portions (e.g., volumes); such approaches are used extensively in health, food, and water applications. These estimates are required for decision making, during which they are typically evaluated against concentration-based criteria or targets (Dickey et al., 1999; Lund et al., 2000; Havelaar et al., 2001; Gerba and Rose, 2003; Gracias and McKillip, 2004; John et al., 2007; Schijven and de Roda Husman, 2011; Davis, 2014; WHO, 2017). This underscores the importance of accurate representation and analysis of detection- and enumeration-based data, especially where the protection of public health is at stake.

Regardless of application area, concentration estimates derived from non-detects (NDs) or low counts are widely perceived to be more uncertain and less reliable than those based on higher counts. This has often led to a desire to quantify enough of

these objects by modifying the enumerated sample portion so that the count falls in a range that is deemed acceptable (Emelko et al., 2008; US FDA, 2017; APHA et al., 2017). When this is not possible, resulting NDs are widely reported as being less than a detection limit (e.g., <1 per analytical sample size) and used as a statement about true source concentration. This convention has been widely implemented and deemed precautionary because it usually leads to higher (i.e., conservative) mean concentration estimates. Approaches for handling this type of ND data are often developed out of computational convenience, though more elaborate approaches also continue to be developed. One important reason for the development of more complex approaches arises from the recognition that true microbe concentrations are imperfectly estimated by the analytical methodologies used to obtain counts from samples (Nieminiski et al., 1995; Allen et al., 2000). For example, the impact of measurement error (i.e., random sampling error and imperfect and/or variable analytical recovery) on microbial concentration estimates has been widely demonstrated and thoroughly discussed (Nahrstedt and Gimbel, 1996; Schmidt and Emelko, 2011; Gonzales-Barron and Butler, 2011; Gronewold et al., 2008; Commeau et al., 2012; Pouillot et al., 2013; Duarte et al., 2015). Measurement error applies universally to all microbial detection and enumeration methods and refers to the random discrepancy between the actual concentration in the presumably homogeneous source and the concentration estimate obtained from a sample (Emelko et al., 2010). Failure to account for measurement error properly has been shown to bias concentration estimates and associated risk estimates, sometimes by orders of magnitude (Pouillot et al., 2013; Schmidt et al., 2013). In contrast, the implications of interpreting and handling NDs using approaches that mishandle measurement error have not been thoroughly discussed. Current reporting conventions for NDs frequently obfuscate their interpretation, so data analysis approaches have been tailored to how these data are reported rather than what the NDs truly represent.

Here, methods used to characterize microbial concentrations from detection- and enumeration-based data are reviewed, and common misconceptions associated with the reporting and handling of NDs are discussed. Examples that draw upon conventions and standards in the drinking water industry are provided to demonstrate why common approaches that treat NDs as censored data are incorrect and lead to bias in interpretation. Finally, recommendations to facilitate standardized reporting and analysis of such data are provided.

2.3 State of scientific practice

Microbial concentrations in food and water are often estimated using detection- and enumeration-based methods. A detection test produces either an ND or positive (≥ 1 microorganism) result. With a series of repeated presence/absence tests (e.g., Colilert Quanti-Tray®) and assumed Poisson-distributed numbers of microorganisms in each test (as a function of aliquot size and shared source concentration), the most probable number (MPN) approach yields a maximum likelihood estimate of concentration (Pouillot et al., 2013). In these detection methods, reporting of raw aliquot sizes and presence/absence results is necessary for concentration estimation. Enumeration-based methods are distinguished from detection-based methods because they yield a whole number count of target microorganisms within an analytical sample size. These include cultivation plate counts of colonies or virus plaques and cell counts obtained using microscopy or flow-/solid-phase cytometry. We suggest that the concepts addressed in this paper also apply to increasingly common biochemical molecular methods (e.g., qPCR, 16S rRNA gene sequencing); however, such methods are excluded from the scope of this work due to additional assumptions and complexities in the inference of concentrations using these methods, which remain hotly debated (Keer and Birch, 2003).

Although many of the aforementioned microbial enumeration methods have been standardized, protocols for the representation, reporting, and analysis of resulting data remain largely inconsistent. Standard microbiological methods, such as those stipulated within *Standard Methods for the Examination of Water and Wastewater*, Part 9000 (APHA et al., 2017), ASTM D5465-16 (ASTM, 2016), ISO 8199:2005 and ISO 7218:2007/2013 (ISO, 2005, 2013), advise that observations should be reported as a count per analytical sample size (e.g., volume). These data (count and sample size) are *raw* in the sense that the original information pertaining to the precision of the count has not been lost, whereas neither the count nor sample size can be deduced when only a concentration estimate is reported (e.g., 1 microorganism in 64.4 litres [L] is more informative than just a reported concentration estimate of 0.0155 microorganisms/L).

In many cases, counts beyond certain thresholds are considered unreliable and avoided if possible. For example, when counting colonies in plating protocols (APHA et al., 2017; ASTM, 2016), an upper bound is often reasonably suggested because of overcrowding and difficulty in distinguishing between individual colonies. In these cases, an upper threshold is often applied beyond which a result of “too numerous to count” (TNTC) is reported. Notably, many conventions related to lower thresholds and NDs also exist. For example, some methods (e.g., in which counts are obtained from a

dilution series) suggest that NDs should be omitted in concentration estimation (United States Pharmacopeial Convention, 2014). It has been a common convention to report NDs as <1 microorganism per analytical sample size (Forum on Environmental Measurements (FEM) Microbiology Action Team, 2016; ISO, 2013; US FDA, 2017), which is the purported method detection limit (MDL). This is frequently interpreted at face value as the *de facto* concentration for statistical analyses and regulatory compliance—despite the recently stipulated caveat that MDLs are inapplicable to “methods that do not produce results with a continuous distribution such as [...] presence/absence methods, and microbiological methods that involve counting colonies” (US EPA, 2016).

2.3.1 NDs in analytical chemistry

To understand the widespread convention of reporting NDs as values below MDLs in microbiology, it is important to understand the origin and motivation behind the concept of an MDL. The MDL (also known as the “limit of detection”) was developed as a performance criterion for chemical analyses (Glaser et al., 1981). This concept has remained largely unchanged since its original conception (Currie, 1999). Although slight variations of this concept exist, the MDL can be operationally defined as the minimum measurement of concentration of a substance that can be reported with a high degree of confidence (commonly 95 or 99%) that the concentration is actually greater than zero (Armbruster and Pry, 2008) (i.e., that the measurement is unlikely to be just random noise despite actual absence of the substance). In stark contrast to the field of microbiology where NDs reflect the inability to observe a single microorganism in a particular analysis, analytical chemistry results are much less susceptible to influence by small numbers of analyte particles—signals obtained for quantification arise from the collective effect of very large numbers of atoms/molecules/ions per mole (e.g., 6.022×10^{23}). In fact, merely 50 nanograms of lead in a litre of water (a detection limit attainable by current lead analysis methods) is comprised of more than 1.45×10^{14} lead atoms due to the magnitude of Avogadro’s number. In chemistry, random sampling errors associated with specific numbers of analyte particles in a well-mixed sample is largely insignificant compared to errors introduced through the application of the analytical method itself—the accuracy of the measurement is limited by the precision of the measurement instrument. The construct of the MDL is intended to reflect these method-specific errors to facilitate comparisons of data generated using different analytical methods for the same analyte at the lower end of concentration ranges.

Although the MDL construct can be useful, concentration observations falling be-

low these thresholds are not devoid of meaning and it has been recommended that these data should be reported as measured chemical detections. They are still valid observations from which true concentrations can be estimated (albeit with greater uncertainty) by applying appropriate statistical approaches (e.g., that make relevant assumptions concerning randomly distributed error, unbiased analytical methodology, and interference effects) (Analytical Methods Committee, 1987). However, some policies require substances to be described as “absent, present in only a limited number of samples, or present in less than a specified number or amount of a given quantity” (National Research Council (US) Subcommittee on Microbiological Criteria, 1985) in regulatory and contractual frameworks, leading to the adoption of reporting limits (i.e., a value below which data are not reported) by many laboratories.

While these reporting conventions are not themselves problematic, they become problematic if these data are incorrectly interpreted or statistically analyzed. The implications of NDs in environmental chemistry have long been recognized (Analytical Methods Committee, 1987; Lambert et al., 1991). Unaltered zero concentrations preclude the calculation of geometric means and cannot be fit by many continuous distributions (without their explicit accommodation through a zero-inflated model). Values reported as below detection or reporting limits have commonly been either omitted or substituted with a function of the limit (Helsel, 2006) to facilitate computationally convenient analysis. These approaches are deemed conservative, but sacrifice information about data reliability and uncertainty that may be critical in decision making. Chemical concentration data reported as less than a detection limit are an example of censored continuous measurements (where the exact measured value within the specified interval is unknown), for which appropriate statistical approaches do exist (Helsel, 2005).

2.3.2 NDs in enumeration-based microbial methods

The direct application of analytical chemistry MDL concepts and associated censoring conventions to microbial enumeration data has inflicted similar challenges for statistical analysis in microbiology. Taking NDs as zeros and weighing them with other non-zero counts based on their respective analytical sample sizes is sufficient for the simple calculation of mean concentrations provided the microorganisms are randomly dispersed and a representative sample was obtained (i.e., from a source where the spatial distribution of the analyte is not heterogeneous); however, this approach is insufficient for fitting concentration distributions and quantifying data reliability or uncertainty in the calculated mean (Parkhurst and Stern, 1998). Commonly used

omission and substitution methods borrowed from analytical chemistry for summarizing and reporting mean microbial concentrations in water introduce bias; substitution methods have been demonstrated to be increasingly biased with greater proportions of NDs in both chemical and microbial data (Parkhurst and Stern, 1998; Roser and Ashbolt, 2005; Helsel, 2005). While the bias introduced using substitution methods can offer a substantial safety factor when harmful microorganisms are rare (by considering them to be present when they are not or they have not been detected), it is critical to note that this bias offers no factor of safety when it is most needed (e.g., when pathogens are routinely observed) (Parkhurst and Stern, 1998).

The acknowledgement that “[data reported as censored] cannot be treated statistically without modification” (APHA et al., 2017) and the growing need to quantify uncertainty in the concentration estimate have led to the development of various statistical tools for analyzing these data. Critically, non-detect microbial data are in fact observed counts of zero commonly misrepresented as censored data. Their misrepresentation has led to the adoption of censored data approaches for handling microbial NDs (Lorimer and Kiermeier, 2007; Busschaert et al., 2010; Williams and Ebel, 2012). While many statistical analyses have assumed that microbial concentrations are measured directly and precisely, markedly different statistical methods have been developed that acknowledge the probabilistic relationship between actual observed data (including NDs) and the underlying microbial concentrations by accounting for measurement error (Nahrstedt and Gimbel, 1996; Schmidt and Emelko, 2011; Gonzales-Barron and Butler, 2011; Gronewold et al., 2008; Commeau et al., 2012; Pouillot et al., 2013; Duarte et al., 2015). As would be expected, different approaches for handling microbial NDs can result in substantially different outcomes. Specifically, the statistical analysis of inappropriately censored NDs may lead to erroneous microbial concentration estimates and subsequent interpretations—this is demonstrated by the examples below. It is critical to recognize that data for which *both* raw counts and sample sizes are *known* are not censored—these include NDs that are based on counts of zero in known sample sizes. These data are not censored and must not be statistically treated as such.

2.4 Results: Evidence that microbial non-detects are not censored data

2.4.1 Occurrences of microbial NDs are not solely a function of analyte concentrations

All microbial concentration estimates are imprecise, not only non-detects. An ND can arise when either the concentration is truly zero or when target microorganisms are present in the source but not successfully detected. Because of the latter case, it is commonly understood that an ND does not necessarily imply that the concentration is truly zero. Indeed, consistent with the aphorism “absence of evidence is not evidence of absence,” a concentration of zero cannot actually be proven by an ND for this reason.

Figure 2.1 examines factors leading to ND results at non-zero concentrations (derivation in Appendix A1). Figure 2.1a depicts the probability of observing an ND as a function of the true concentration and the sample volume assuming Poisson-distributed organism counts and a method with 100% analytical recovery. Probability of ND profiles are presented for volumes of 0.010 L, 1.0 L and 100 L to illustrate the impact of hundredfold increases in the analytical sample size. Common sample volumes for total coliform/*E. coli* and protozoan (oo)cyst analyses are 0.100 L and 100 L, respectively. Intuitively, the probability of an ND observation from a single sample increases with decreasing concentration and analytical sample size. In practice, the occurrence of random NDs can be reduced by increasing sample size.

Building upon the previous example, Figure 2.1b addresses the occurrence of NDs given a 1.0 L sample volume and various analytical recovery profiles. The bold curve in Figure 2.1b is identical to the one in Figure 2.1a, but plotted on a linear concentration scale. It represents 100% analytical recovery, whereas the second curve addresses the scenario of a constant analytical recovery of 40% (i.e., the probability of observation for each microorganism initially gathered is 40% in any sample). Logically, the probability of NDs increases as microorganisms are more likely to be lost during sample processing. The remaining curve retains a mean recovery of 40%; however, substantial variation in recovery among samples is described by a beta distribution. This further inflates the probability of an ND observation because some samples would have relatively low recovery. Clearly, the occurrence of NDs is sensitive not only to the source concentration and the analytical sample size, but also the analytical recovery profile of the method for the particular sample matrix.

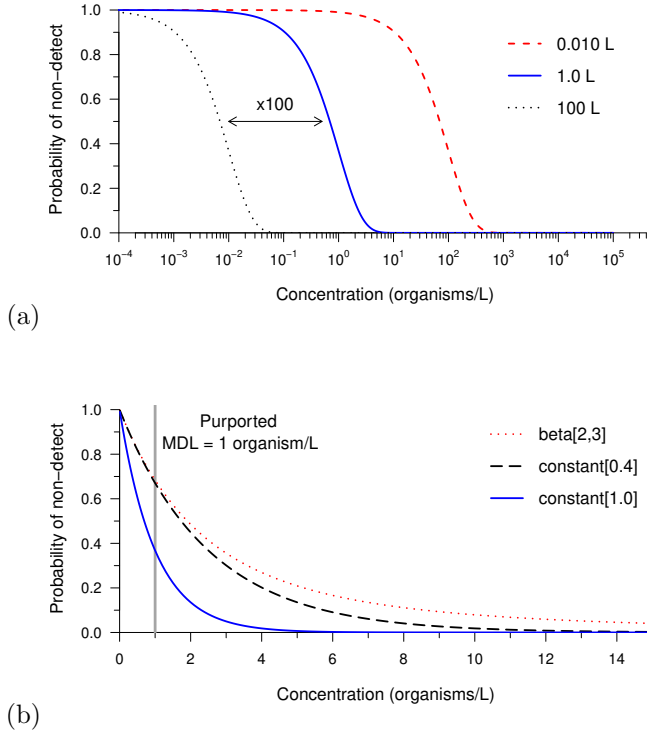


Figure 2.1: Probability of a non-detect observation as a function of organism concentration and (a) various analytical sample volumes given 100% analytical recovery, and (b) various analytical recovery profiles given a 1.0-L sample, each assuming Poisson random sampling error. The constant [0.4] and beta-distributed [beta(2,3)] recovery profiles share a mean of 40% analytical recovery, but the latter is more variable.

It may be useful to consider the concentration beyond which NDs become improbable (e.g., probability $<1\%$) when comparing alternative methods, choosing a target sample volume, or determining the appropriateness of a method for a particular application. We propose that this threshold may be called a method sensitivity limit (MSL) because sensitivity is the probability of detection when the target microorganisms are actually present in the source. Considering the examples in Figure 2.1b, the scenario with 100% analytical recovery has an MSL of 4.6 organisms per litre. With 40% analytical recovery, the MSL increases to 11.5 organisms per litre. The MSL is 32.5 organisms per litre in the final scenario, illustrating the pronounced effect of variability in analytical recovery upon sensitivity of microbial analytical methods. While

this calculated value could be useful, it is important to note that it is sensitive to uncertainty in the parameters and shape of the analytical recovery distribution (where low recovery values are common), and would not be practical to evaluate for every method and sample matrix.

2.4.2 Uncertainty in concentration estimates precludes MDL-based interpretation of results

The statistical analysis of inappropriately censored microbial data ultimately leads to erroneous concentration estimates and subsequent interpretations. Bayesian techniques (Gelman et al., 2014) provide a means of demonstrating the uncertainty surrounding the concentration estimate obtained from microbial enumeration data (Grone-wold et. al., 2008; Schmidt and Emelko, 2011; Gonzales-Barron and Butler, 2011; Duarte et al., 2015). Accounting for measurement error, these methods describe the relative probability of alternative values of the true microbial concentration given the count observation obtained from the analytical sample and a prior representing beliefs about the plausible values of concentration before data analysis. Figure 2.2 illustrates what a single ND observation (Figure 2.2a) and an observation of two microorganisms (Figure 2.2b) within a 1.0 L sample volume imply about concentration assuming perfect analytical recovery and using a relatively uninformative semi-infinite uniform prior (derivation in Appendix A2).

When an ND is observed (Figure 2.2a), there is still a large probability ($\approx 37\%$ in this example) that the actual concentration exceeds the purported MDL, therefore invalidating the assertion that an ND means that the actual concentration is $<MDL$. Conversely, a count of two organisms (Figure 2.2b) leads to a considerable probability ($\approx 8\%$ in this example) that the actual concentration could still be less than the purported MDL. This simple demonstration shows that the interpretation of NDs as censored data below the purported MDL is inappropriate, and further underscores that point estimates of concentration ought not be treated as exact measurements.

2.4.3 Censoring in detection- and enumeration-based microbial methods

Although NDs in microbial detection and enumeration methods are not censored data, there are scenarios when certain microbial methods yield truly censored data.

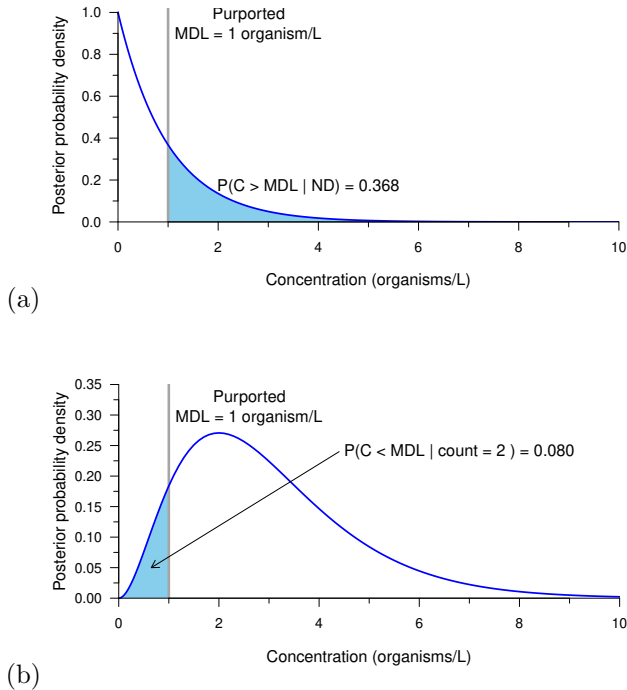


Figure 2.2: Posterior probability density function (PDF) characterizing uncertainty in the true concentration given (a) an ND observation, and (b) an observation of two organisms, each based on a 1.0-L sample, 100% analytical recovery, and a semi-infinite uniform prior. The purported MDL of 1 organism/L is shown with the probability of the true concentration exceeding or falling short of the purported MDL, respectively.

Censored data occur when there is incomplete knowledge about an observed measurement above, below or between specified values (Millard et al., 2012). Such censoring can be inherent to the method or imposed deliberately by the analyst as exemplified below. In either case, censoring applies to raw measurements (e.g., count test results) rather than calculated values that are not measured directly (e.g., concentration estimates).

An ND in presence-absence tests (e.g., Colilert®[®], Rapid Hi-Coliform™[™], Aqua-CHROM™[™]) implies a count of zero within the associated analytical sample volume. A positive test result can be construed as an inherently censored count of at least one microorganism because the method cannot reveal the exact number of microorganisms leading to detection. For a series of presence-absence tests, MPN approaches implicitly reflect censored data analysis by using the cumulative probability of all non-zero counts

(i.e., the complement of the probability of a non-detect) to represent a positive test result in the likelihood function.

When using culture-based methods, counts beyond an upper limit are conventionally reported as TNTC (e.g., 150 or 200 colony forming units (cfu) for spread plates, 80 cfu for membrane filtration, and 300 cfu for pour plates [APHA et al. 2017; ASTM 2016]). If a specific observed count is replaced with TNTC, then this is an example of imposed censoring. In contrast, censoring is inherent if counting is terminated upon reaching the limit or is not attempted because the count would clearly exceed it. Such truly censored observations may be incorporated into the Bayesian method described previously (or any likelihood-based method) using cumulative density for the censored range of counts rather than just probability density associated with particular observed counts. Some standards (APHA et al., 2017) recommend completing a new analysis with dilution to *replace* TNTC results. We suggest that TNTC results should be *retained* in subsequent statistical analyses by using likelihood-based methods that allow inference from both the TNTC result and the count obtained through re-analysis of the sample. This would enable more accurate description of knowledge about the concentration by harnessing all of the available information rather than omitting inconvenient data.

2.5 Implications for policy and practice

Many environmental science and public health decision-making and regulatory frameworks rely upon the accurate evaluation of microbial concentrations and comparison with concentration-based criteria. For example, evaluation of source water pathogen concentrations is used to determine minimum treatment infrastructure requirements in the provision of safe drinking water (US EPA, 2006; Alberta Environment and Sustainable Resource Development, 2012). Here, concentration estimates that bias high may lead to misallocation of resources including costly infrastructure investments and operational adjustments.

Giardia has been the most commonly reported intestinal protozoan in North America and worldwide; it is also likely the most common cause of surface water-borne infectious disease outbreaks (Adam et al., 2016; Efstratiou et al., 2017). A set of eight source water *Giardia* cyst counts (Table 2.1) from the larger City of Calgary database were used to illustrate the potential impacts of various ND data analysis approaches. The impact of inappropriate ND data analysis approaches will vary in accordance to characteristics of each dataset; greater bias can be expected when NDs constitute a

larger proportion of the dataset and perfect analytical recovery is not attainable. However, more detailed analysis regarding the scale of implications associated with such characteristics (e.g., number of samples, proportion of zeros, distribution of positive detections, etc.) was beyond the scope of the present investigation. Consistent with current practice and interpretation of the regulations, the raw data were not adjusted for viability or infectivity, with 100% analytical recovery assumed.

Table 2.1: Summary of raw water samples analyzed for *Giardia* cysts from City of Calgary, AB, Canada - October, 2012.

Sample	Raw count	Volume processed [L]	Data reported [cysts/100 L]
1	1	64.4	1.6
2	0	50.2	<2.0
3	0	50.0	<2.0
4	0	53.2	<1.9
5	0	50.2	<2.0
6	0	50.4	<2.0
7	0	50.4	<2.0
8	2	50.7	3.9

The data were used to obtain maximum likelihood estimates (MLE) of the mean and standard deviation of *Giardia* cyst concentrations (US EPA, 2006; Alberta Environment and Sustainable Resource Development, 2012) assuming log-normally distributed concentrations and independence among sampling events. NDs were omitted in Approach A and substituted with the MDL and half the MDL in Approaches B and C (approaches critiqued by Helsel [2005]), and were handled as censored data in Approach D (Busschaert et al., 2010; Williams and Ebel, 2012). The purported MDL of 1 cyst per volume analyzed is critical for substitution and censored data methods. For Approach D, the cumulative density between zero and the purported MDL was used for NDs (Busschaert et al., 2010). MLE was applied for Approaches A–D using the *fitdistrplus* package (v. 1.0-9) (Delignette-Muller et al., 2017) in *R*. In Approach E, a Poisson distribution was used to account for random sampling error with log-normally distributed concentrations, using the *poilog* package (v. 0.4) (Grøtan and Engen, 2008) in *R* (details provided in Appendix A3). It is important to note that Approaches A–D are based only on reported concentration estimates whereas Approach E (and other approaches that account for measurement error) necessitates the reporting of raw data.

Statistics from this analysis are summarized in Table 2.2.

As would be expected, Approaches A–D yielded substantially higher mean *Giardia* cyst concentrations relative to Approach E because the NDs were omitted or represented as non-zero values. Omission and substitution approaches are known to lead to biased mean concentration estimates relative to methods appropriate for censored data (Helsel, 2005). However, the types of microbial ND data considered herein are fundamentally not censored, as discussed above. There is a critical difference between censored data approaches (Approach D) and those that actually incorporate NDs as legitimate, discrete observations by accounting for measurement error (e.g., Approach E). In this example, the parasite concentrations were overestimated relative to Approach E by a factor of 2.1 to 3.9 when NDs were inappropriately handled (Approaches A–D).

Given sufficient and suitable information, the MLE approach incorporating random sampling error (Approach E) can be extended to account for analytical recovery. However, model fitting by MLE becomes more difficult with increasing model complexity—numerical integration required for evaluating the resulting likelihood function becomes practically intractable in many cases. Bayesian methods can be used to fit more complex probabilistic models to data, but also suffer from substantial parametric uncertainty where insufficient data and/or data that are relatively uninformative about model parameters are available (Gleit, 1985; Helsel and Cohn, 1988). Indeed, small statistical sample sizes are often inevitable when using time- and/or resource-intensive microbial analytical methods (such as those for protozoan (oo)cyst enumeration [US EPA 2005, 2012]). For example, utilities undertaking minimum source water monitoring requirements for the determination of drinking water treatment targets (Alberta Environment and Sustainable Resource Development, 2012) would be determining running mean *Giardia* cyst concentrations based on monthly samples collected over the course of two years (i.e., $n=24$). The impact of small statistical sample sizes on concentration distribution parameter estimates is exacerbated when all of the data available are non-detects, in which case statistical analysis is not possible without strongly subjective priors.

Although treatment requirements are not typically determined based on mean *Giardia* cyst concentrations estimated from a handful of samples, these data may exemplify monitoring results from utilities that draw upon high quality source waters. Such systems, especially those that have limited treatment, operational, and/or monitoring capacity, are particularly vulnerable to the implications associated with overestimated mean concentrations. As demonstrated in this analysis, concentration estimates may be biased high by a factor of two or more just by handling NDs as censored. This could lead to operational and maintenance costs/adjustments (e.g., energy for UV disinfection).

Table 2.2: Comparison of *Giardia* cyst concentration statistics obtained using various approaches for handling microbial NIDs.

Approach/Model	Treatment of NIDs	$\hat{\mu}$	$\hat{\sigma}$
A) log-normal	omitted	0.0276	0.0136
B) log-normal	substituted with MDL	0.0216	0.0055
C) log-normal	substituted with $\frac{1}{2}$ MDL	0.0139	0.0067
D) log-normal	censored data ($<$ MDL)	0.0149	0.0100
E) Poisson log-normal	zeros with random sampling error	0.0071	0.0071

$\hat{\mu}$: MLE of the mean cyst concentration [cysts/L]

$\hat{\sigma}$: MLE of the standard deviation [cysts/L]

tion, alteration of design flow rates) (Cotton et al., 2001) that are inordinate given the levels of pathogens actually present in the source. Thus, such bias can also inappropriately affect assessments of water treatment plant “firm capacity,” which indicates pathogen treatment capacity in absence of one key treatment barrier and therefore informs infrastructure needs. While application of these approaches may result in bias that invokes more conservative levels of treatment (Parkhurst and Stern, 1998), it is better to analyze microbial concentrations accurately and apply consistent safety factors, regardless of the data than to apply flawed data analysis approaches with unspecified safety factors attributable to preventable bias. This precludes the universal and equitable application of microbial standards, and ultimately undermines the consistent level of public health protection that the industry strives to maintain.

2.6 Conclusions

- Non-detect microbial detection and enumeration data are fundamentally not censored data and should not be reported or analyzed as such.
- MDLs are not intended to be used for, and have therefore been misapplied in, detection- and enumeration-based methods that count discrete microorganisms.
- The convention of reporting NDs as censored values relative to an MDL is misleading when using enumeration-based methods and has resulted in the misuse of censored data statistical approaches for microbial data analysis.
- It is inconsistent to consider the uncertainty in non-detects by representing them as censored data while ignoring the inherent uncertainty in all non-zero counts.
- Censored data approaches should be reserved for data correctly interpreted as being censored, such as TNTC plate counts where the actual count is known only to exceed a specified threshold.
- This work re-emphasizes that raw microbial data must be reported to facilitate proper statistical analysis approaches that account for measurement error.

Chapter 3

An end justified by means: Ensuring adequate treatment of protozoa by precise mean concentration estimation

Alex H.S. Chik
Philip J. Schmidt
Monica B. Emelko
Alfred P. Blaschke
Jack F. Schijven

In preparation for submission to *Journal of Water and Health*

3.1 Summary

Quantification of risks and determination of drinking water treatment needs for enteric protozoan pathogens relies critically on the characterization of their source water concentrations. This involves the use of time-consuming and labour-intensive methods to enumerate protozoan pathogens in raw water samples as well as the evaluation of analytical recovery to correct for losses from the application of the method. The minimization of these efforts without substantially compromising precision is a key consideration for monitoring program design, especially for systems serving small populations or that are far from laboratories offering these types of analyses. In this study, the precision of estimated mean source protozoan concentrations is evaluated under different hypothetical monitoring scenarios using simulated data. The probability that the mean concentration is precisely estimated—the statistical power of the sampling scenario—is represented by the proportion of repeated simulations successfully rendering sufficiently precise mean concentration estimates. For the scenarios considered, even low numbers of samples (e.g., $n=4$) consistently yielded mean concentration estimates within 0.5-log_{10} . A more stringent goal of precision to attain 95% highest density interval widths of $<1.0\text{-log}_{10}$ required larger numbers of samples (e.g. 24 or 52 samples). The impact of reducing efforts to characterize recovery upon achieving monitoring goals was comparatively minor. This work provides proof-of-concept for Bayesian retrospective power analysis as a tool to explore the relative merits of various protozoa monitoring strategies to support quantitative microbial risk assessments.

3.2 Introduction

Whether quantitative microbial risk assessments (QMRA) are conducted to evaluate regulatory compliance, inform treatment process design, or for quantifying and characterizing risk as part of a greater water safety plan approach, accurate drinking water risk estimates begin with exposure assessment and more specifically, microbial hazard identification and quantification (WHO, 2017). Given the inherent limitations of monitoring pathogen concentrations in the treated drinking water (Regli et al., 1991; Allen et al., 2000; Petterson et al., 2007; Smeets et al., 2010), the evaluation of microbial pathogen risk relies on the estimation of pathogen concentrations in the raw source water (Teunis et al., 1997; Schijven et al., 2011). This typically involves the enumeration of pathogens in raw water samples as well as the evaluation of analytical recovery to correct for losses during sample processing and errors in counting. Current

methods available for protozoan pathogen concentration estimation are particularly costly, time-consuming and labour-intensive. In particular, the merit of protozoan source water monitoring programs have often been scrutinized when low source concentrations are compounded by the impacts of imperfect analytical recovery to yield highly variable results comprising many non-detects (Chik et al., 2018). The mean concentrations estimated therefrom are often perceived to be highly uncertain. However, this presumption of imprecision in mean concentration estimates has not been corroborated.

Existing works have addressed sampling effort based on the analytical sample volume; Emelko, Schmidt, and Roberson (2008) demonstrated that the best estimates of source water pathogen concentrations arise from the collection and analysis of sample volumes sufficient to have an average of at least 10 organisms in each sample. Depending on source water quality and the analytical recovery anticipated, sample volumes well over 1000 L may be necessary, although practical sample volumes range from 10-20 L (e.g. Reynolds et al., 1999; US EPA, 2005, 2012; Ruecker et al., 2007) to 100 L (e.g. Rochelle et al., 1999; Quintero-Betancourt et al., 2002). Despite the costs and resources that can be allocated based on the number of samples collected for 1) protozoan enumeration and 2) internally-seeded matrix-spikes of non-native protozoan pathogen surrogates (e.g., ColorSeedTM) to enable quantification of analytical recovery, the influence of these critical factors on protozoan concentration estimation has not been explicitly quantified.

Undoubtedly, enhanced sampling effort would yield increasingly precise pathogen concentrations. However, as the mean is the most broadly used mathematical measure of average and the mean pathogen concentration estimate directly relates to the calculation of mean risk (which is often interpreted/used as a target for regulatory compliance), a focused evaluation of the precision attainable for this measure was performed. Accordingly, an approach for estimating the level of precision attainable for mean protozoan concentrations associated with a specific monitoring program design is presented. To account for bias attributable to method losses, a hierarchical model that incorporates consideration of analytical recovery developed by Schmidt and Emelko (2011) was used. The model accounts for random measurement errors and integrates available recovery information in temporally distributed protozoan enumeration data. First, practically relevant parameter values (including the mean protozoan concentration) for this model are estimated from a historical protozoan monitoring dataset. These values are subsequently used to repeatedly simulate data sets that would emanate from them based on hypothetical monitoring scenarios. Goals related to the precision of mean concentration estimates are evaluated for each simulated dataset;

the proportion of repeated simulations for which goals of precision are achieved represent the probability that the desired level of precision would be attained using that particular monitoring program design.

This overall approach described is consistent with a Bayesian retrospective power analysis, which facilitates the exploration of the influence of the crucial factors on the precision of mean protozoan concentration estimation. Given that treatment needs are often described in terms of decimal reductions and regulatory treatment log-credits are typically assigned to the nearest 0.5-log_{10} , mean concentration estimates within half an order-of-magnitude ($\pm 0.5\text{-log}_{10}$) of the true mean were herein investigated as the goal of precision. The level of precision evaluated herein also implies accuracy provided that enumerated protozoa are not subject to any other source of bias. Bias attributable to species, subtypes, degree of viability and infectivity of the protozoa enumerated (Ruecker et al., 2007; Efstratiou et al., 2017) was outside the scope of the present work, as these factors require more extensive laboratory analyses that are not widely evaluated. Two goals related to precision were evaluated: 1) the true mean and the mode of the posterior mean concentration distribution differs by less than half an order-of-magnitude ($\pm 0.5\text{-log}_{10}$), and 2) the 95% Bayesian highest density interval (HDI) is less than an order-of-magnitude (1.0-log_{10}) wide. This study provides proof-of-concept for the broader consideration of statistical power in the design and implementation of source water protozoan monitoring programs.

3.3 Methods

3.3.1 Estimation of practically relevant parameter values of hierarchical model using historical protozoan monitoring program data

The City of Calgary has an extensive historical protozoan (*Giardia* and *Cryptosporidium*) monitoring dataset spanning from 2012 to 2019 with internally seeded matrix-spikes for the evaluation of sample specific analytical recovery. These data were analyzed using the hierarchical model described in Schmidt and Emelko (2011), which is graphically presented in Appendix B1 and briefly summarized here. Specifically, the number of (oo)cysts enumerated (y_i) in a particular analyzed volume of water (V_i) is Poisson distributed with the mean equal to the product of the source protozoan concentration (C_i), volume, and sample-specific analytical recovery (p_i). Temporal protozoan concentrations are modelled as gamma distributed with mean (μ)

and standard deviation (σ). The number of microorganisms recovered (n_i) from those seeded (S_i) is binomially distributed conditional on the sample-specific recovery (p_i). This analytical recovery is regarded as the probability that each microorganism will be observed and is assumed equal for seeded and indigenous microorganisms in any given sample. It varies randomly among samples according to a beta distribution with parameters α and β .

A Bayesian approach was used to estimate these parameters rather than maximum likelihood estimation because the likelihood function cannot be evaluated explicitly (Gelman and Hill, 2007; Gelman et al., 2014). Informative normal priors on the \log_{10} -mean and \log_{10} -standard deviation of the concentrations were chosen to represent a wide range of concentrations of practical relevance. The mean concentration is further bound between 0.001 organisms/100 L and 1000 organisms/100 L. Additionally, the variability of concentration was constrained to within two orders of magnitude of the mean concentration for numerical stability when implemented in RunJAGS (Denwood, 2016) in the R environment. The parameters of the gamma distribution for concentration variability (μ , σ) and the beta distribution for analytical recovery (α , β) were estimated using the modes of the respective posterior distributions.

Gibbs sampling implemented in JAGS (Plummer, 2016) through RunJAGS was deployed to obtain the posterior distribution of the mean concentration. For the estimation of the posterior parameter distributions, 3 chains were deployed to ensure apparent convergence and parameter mixing. After a burn-in of 1,000 iterations and 2,000 adaptive iterations, every 5th of 25,000 iterations across 3 chains were used to evaluate the posterior distribution of the mean concentration. Computational underflows in JAGS are known to occur when sampling from a gamma distribution with its shape parameter less than 1 and when sampling from beta distributions with α or β parameter values close to 0 and 1. Moreover, the choice of samplers (algorithm used to generate Markov Chains) deployed is imposed by JAGS/RunJAGS. Given these known software limitations, priors on α and β were set to be uniform distributions bound by 0.01 and 100, while the gamma distributed concentrations was truncated to exclude near-zero concentrations. A handful of simulations from RunJAGS were also subsequently corroborated with results obtained in OpenBUGS (Spiegelhalter et al., 2007) (that allows the specification of the sampler used) to verify the impact of these adjustments.

3.3.2 Bayesian retrospective power analysis

The probability that a particular sampling scheme would yield adequately precise mean concentration estimates can be evaluated by means of a Bayesian retrospective power analysis (Gelman and Hill, 2007; Kruschke, 2014). It is retrospective in the sense that the power evaluated reflects the ability of the hypothetical monitoring scenarios investigated to precisely evaluate the historical parameter estimates. In such an analysis, the posterior estimates of the model parameters obtained from historical *Cryptosporidium* (n=229) and *Giardia* (n=234) data are assumed to be collectively representative and sufficiently describe the source water pathogen concentrations and anticipated analytical recovery. The analyses conducted herein were performed using the parameters derived based on these two protozoan data sets.

The *Cryptosporidium* historical parameters are consistent with parameter values that reflect low mean concentrations ($\mu=0.8$ oocysts/100 L, $\sigma=1.2$ oocysts/100 L, $\alpha=4.6$, $\beta=2.8$) whereas the *Giardia* historical parameters are consistent with those that reflect higher mean concentrations ($\mu=19.4$ cysts/100 L, $\sigma=25.3$ cysts/100 L, $\alpha=5.3$, $\beta=3.5$). Inferences drawn from this analysis are strictly based on differences in the parameter values rather than the characteristics of the protozoan pathogens themselves. These historical parameters were used to simulate data sets representing different hypothetical sampling scenarios stochastically. When the simulated data are then analyzed with the same Bayesian hierarchical model as the original historical data, the posterior distribution of the parameter(s) of interest can be examined for whether the specified goals of precision are achieved (Kruschke, 2014). The long-term probability of successfully attaining established goals is an explicit representation of statistical power associated with the hypothetical monitoring scenario. While the extrapolation of results from a retrospective power analysis to future monitoring conducted would require the presumption of stationarity, this process offers a pragmatic starting point that can be used adaptively to evaluate the efficiency of monitoring strategies in relation to attaining the desired goal(s) of precision.

3.3.3 Goals of precision related to the posterior distribution of mean protozoan concentration

For all hypothetical monitoring scenarios, both point- and interval-based goals of precision were evaluated for the posterior distribution estimated from each simulation:

1. the mode of the posterior distribution of mean concentrations to be estimated

within 0.5-log_{10} of the true mean concentration

2. the width of the 95% Bayesian highest density interval (HDI) on the posterior mean concentration to be at most 1.0-log_{10}

The first goal is a point-based criterion that makes use of the posterior mode, which is a special case of the Bayesian maximum likelihood estimate when prior knowledge of the parameter values are equally likely (Cousineau and Helie, 2013). The posterior mode criterion effectively allows a range of acceptable concentrations spanning 1.0-log_{10} . The second goal invokes the use of the 95% Bayesian HDI, which is the range of values from the posterior distribution of mean concentrations that are most credible. A narrow interval is intuitively more desirable and reflects higher precision in the parameter estimated. While both measures of precision are equally valid, they reflect different aspects of precision in the mean concentration estimate. The posterior mode relative to the true data-generating mean is an explicit quantification of precision. On the other hand, precision is implicit to the computation of the HDI width using the simulated data for each trial that does not rely on knowledge of the true data-generating mean. Both types of criteria were examined in this work to elucidate their use and limitations in evaluating statistical power.

The collection of 4, 8, 12, 24, and 52 samples (that correspond to quarterly, biquarterly, monthly, biweekly, and weekly monitoring frequencies if conducted over an annual time frame) with internally seeded matrix-spikes were first explored to determine the influence of the monitoring effort necessary to attain a mean concentration estimate within the specified targets. As long as the samples are collected as part of a routine monitoring program and collectively representative of the conditions spanning an entire year, the mean concentration estimated would be useful for quantitative microbial risk assessments that are performed based on mean annual risk. Next, the effort associated with internally seeded matrix-spikes was investigated. In this regard, the contribution of analytical recovery to overall goals of precision was also investigated by randomly removing a quarter, a half, and three quarters of the internally seeded matrix-spikes (as though the spikes were only conducted for a subset of all source water samples). Finally, the scenario of poorer analytical recovery was investigated, which provides an example of how this framework can be used to evaluate the relative merits of increasing efforts related to native protozoan enumeration versus improving analytical recovery. For each hypothetical monitoring scenario, 100 simulations were performed.

3.4 Results

3.4.1 Point-based goal of precision based on the mode of the posterior mean concentration distribution

For both sets of low and high historical concentration parameters, the influence of the informative but relatively flat prior centred at 10 microorganisms/100 L was discernible when relatively few samples ($n=4$) were collected. The posterior modes estimated from the simulation of the *Cryptosporidium* parameters tended to be inflated compared to the historical mean. Conversely, the posterior modes from simulations of the *Giardia* parameters tended to be reduced compared to the data-generating mean. The influence of the prior becomes less discernible with additional samples, as reflected by the tightening of the range of posterior modes when four additional samples were collected ($n=8$). This will converge asymptotically upon the data-generating mean with more samples collected.

A large proportion of the posterior modes estimated will fall within 0.5-log_{10} of the true mean concentration even with very few samples ($n=4$) and minimal recovery data (i.e., a single recovery estimate) (Table 3.1, Figure 3.1). A higher probability of attaining a posterior mode within 0.5-log_{10} of the true mean for the set of *Cryptosporidium* parameters (which reflect a lower historical mean concentration) compared to the *Giardia* parameters (which reflect a higher historical mean concentration) used. As additional points of reference, the probability of attaining a posterior mode within 1.0- and within 0.3-log_{10} of the true mean was also presented. It is all but certain that posterior modes will fall within a full order-of-magnitude even with only four samples for these scenarios investigated. Posterior modes that are within 0.3-log_{10} of the true mean denotes the ability to capture a mean concentration that is either half or double the true mean concentration. As expected, more samples are typically required to achieve this higher level of precision.

3.4.2 Interval-based goals of precision based on the 95% HDI width of the posterior mean concentration distribution

95% HDIs were estimated through sampling the posterior distribution of the mean concentration by updating the prior with a truncated gamma likelihood distribution (Table 3.2, Figure 3.1). There is a clear trend in the reduction of the 95% HDI width

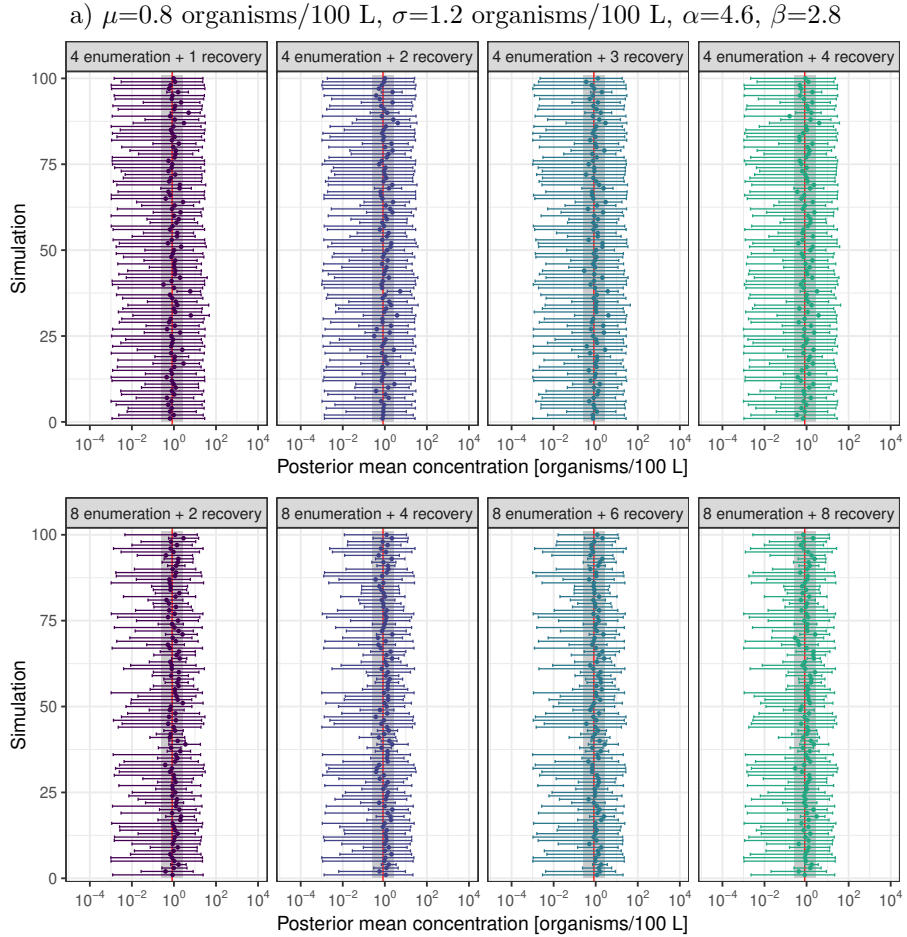


Figure 3.1: Posterior modes (dots) and 95% highest density intervals (error bars) of mean concentration across 100 simulations for a) *Cryptosporidium* (i.e., low mean concentrations) and b) *Giardia* (i.e., high mean concentrations) historical parameters. Individual facets within each subplot are labelled with the number of 100 L samples collected from the source for enumeration and the number of internally seeded matrix-spikes performed. The red line denotes the true mean concentration used to generate the simulated data; the grey region denotes mean concentrations that are within 0.5-log_{10} of the true mean concentration.

b) $\mu=19.4$ organisms/100 L, $\sigma=25.3$ organisms/100 L, $\alpha=5.3$, $\beta=3.5$

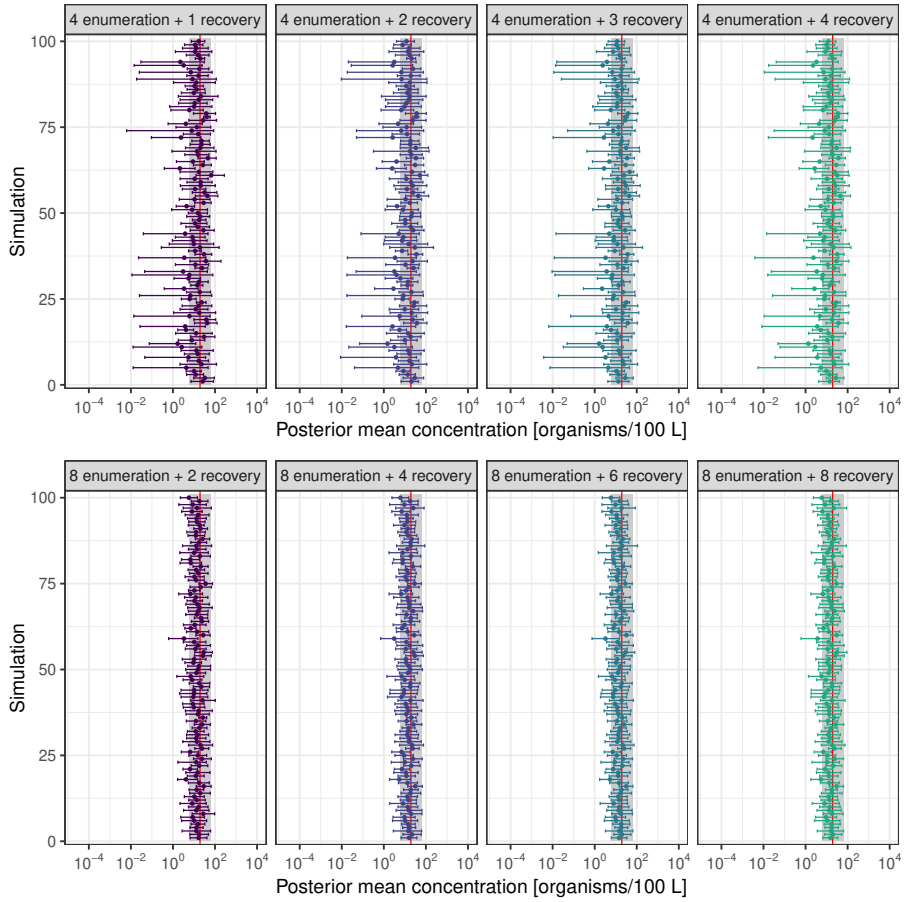


Figure 3.1: (continued)

Table 3.1: Estimated power to achieve goal of precision based on the posterior mode of the mean concentration, as a function of sample size reflecting a) *Cryptosporidium* (i.e., low mean concentrations) and b) *Giardia* (i.e., high mean concentrations) historical parameters. All sample volumes are assumed to be 100 L and each paired with an internally seeded matrix-spike count. Values in tables are the estimated proportion of 100 simulations for which the modal posterior estimate of the mean lies within the specified range of the historical mean concentration (with 95% HDI lower and upper bounds on the estimated proportion in brackets, respectively)

No. of samples	a) <i>Cryptosporidium</i> $\mu=0.8/100$ L, $\sigma=1.2/100$ L $\alpha=4.6$, $\beta=2.8$			b) <i>Giardia</i> $\mu=19.4/100$ L, $\sigma=25.3/100$ L $\alpha=5.3$, $\beta=3.5$		
	posterior mode					
	$\pm 1.0\text{-log}_{10}$	$\pm 0.5\text{-log}_{10}$	$\pm 0.3\text{-log}_{10}$	$\pm 1.0\text{-log}_{10}$	$\pm 0.5\text{-log}_{10}$	$\pm 0.3\text{-log}_{10}$
4	1.0 (0.97, 1.0)	0.96 (0.91, 0.99)	0.76 (0.67, 0.84)	0.99 (0.95, 1.0)	0.82 (0.74, 0.89)	0.67 (0.58, 0.76)
8	1.0 (0.97, 1.0)	0.99 (0.95, 1.0)	0.84 (0.76, 0.90)	1.0 (0.97, 1.0)	0.96 (0.91, 0.99)	0.83 (0.75, 0.90)
12	1.0 (0.97, 1.0)	0.99 (0.95, 1.0)	0.93 (0.87, 0.97)	1.0 (0.97, 1.0)	0.97 (0.92, 0.99)	0.85 (0.77, 0.99)
24	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	0.90 (0.83, 0.95)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	0.96 (0.91, 1.0)
52	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	0.91 (0.96, 0.99)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)

Table 3.2: Estimated power to achieve goal of precision based on 95% Bayesian HDI widths, as a function of sample size reflecting a) *Cryptosporidium* (i.e., low mean concentrations) and b) *Giardia* (i.e., high mean concentrations) historical parameters. All sample volumes are assumed to be 100 L and each paired with an internally seeded matrix-spike count. Values in tables are the estimated proportion of 100 simulations for which the 95% Bayesian HDI width is attained (with 95% HDI lower and upper bounds on the estimated proportion in brackets, respectively)

No. of samples	a) <i>Cryptosporidium</i> $\mu=0.8/100$ L, $\sigma=1.2/100$ L $\alpha=4.6, \beta=2.8$			b) <i>Giardia</i> $\mu=19.4/100$ L, $\sigma=25.3/100$ L $\alpha=5.3, \beta=3.5$		
	95% Bayesian HDI width					
	2.0- \log_{10}	1.0- \log_{10}	0.5- \log_{10}	2.0- \log_{10}	1.0- \log_{10}	0.5- \log_{10}
4	0.06 (0.02, 0.12)	0.01 (0, 0.05)	0 (0, 0.03)	0.80 (0.72, 0.87)	0.32 (0.23, 0.41)	0.04 (0.01, 0.09)
8	0.39 (0.30, 0.49)	0.05 (0.02, 0.11)	0 (0, 0.03)	1.0 (0.97, 1.0)	0.65 (0.55, 0.74)	0.01 (0, 0.05)
12	0.59 (0.49, 0.68)	0.08 (0.04, 0.14)	0 (0, 0.03)	1.0 (0.97, 1.0)	0.93 (0.87, 0.97)	0.07 (0.03, 0.13)
24	0.95 (0.90, 0.98)	0.59 (0.49, 0.68)	0 (0, 0.03)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	0.48 (0.38, 0.58)
52	1.0 (0.97, 1.0)	0.98 (0.94, 1.0)	0.42 (0.33, 0.52)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)

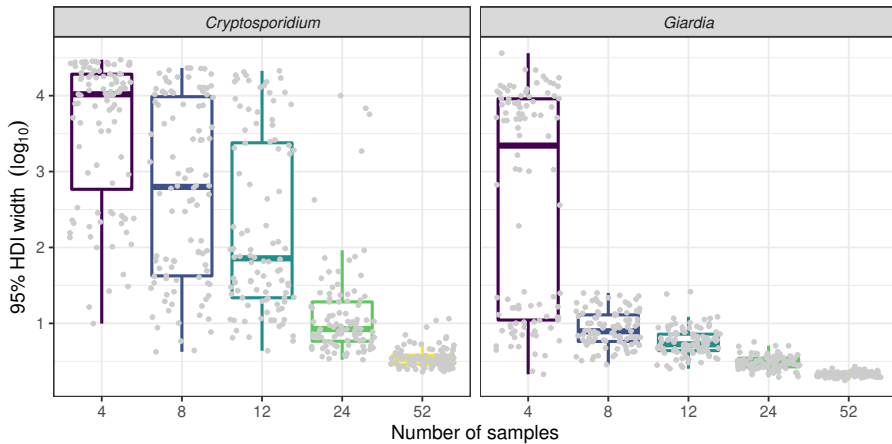


Figure 3.2: Estimated 95% Bayesian HDI widths as a function of sample size with paired recovery data.

with an increased number of samples collected for a given set of historical mean concentrations (Figure 3.2). HDI widths of approximately 1.0-log_{10} or less are generally attainable with 52 and 24 samples for the low and high historical mean concentration sets of parameters, respectively. Predictably, more samples were necessary to attain the same HDI width for the *Cryptosporidium* parameters used than the *Giardia* parameters used. This is largely attributable to the higher prevalence of relatively uninformative non-detects to the precision of the mean concentration given that the historical mean concentration simulated is over an order of magnitude lower. This leads to a wider HDI that reflects the uncertainty in the mean concentration estimate. Notably, some 95% HDI widths were further inflated (e.g., 95% HDI widths estimated to be approximately 4-log_{10} for the scenario with 4 samples using the *Giardia* historical parameters, Figure 3.2). These are recognized to be a computational artefact owing to the truncation of the gamma likelihood distribution that was necessary at near-zero concentrations to overcome computational underflows using the conjugate gamma sampler imposed by JAGS when using RunJAGS. This artefact may be remedied through re-evaluating the posterior mean distribution using OpenBUGS (that does not impose the choice of samplers and necessitate the truncation of the gamma likelihood on concentration). However, OpenBUGS is manually cumbersome for extensive computational iterations that are required for Bayesian power analysis and therefore was only used to confirm the influence of this artefact.

3.4.3 Impact of analytical recovery on attaining precision mean concentration estimates

Despite the rather variable recovery profiles that are estimated based on characteristics of the City of Calgary protozoan dataset (e.g., mean recovery \pm standard deviation $63 \pm 15\%$), a reduction in the number of internally seeded matrix-spikes performed did not substantially reduce the probability of achieving either point- or interval-based criteria for a specified sample size. This was reflected by the similarities amongst the different facets presented in Figure 3.1. This is expected as the uncertainty in concentration estimates is minimally affected by improving method recovery and estimates thereof when low organism counts are observed (Emelko et al., 2008). The *Giardia* recovery parameters were further adjusted to investigate the influence of poorer mean recovery but with a similar degree of variability ($\alpha=2$, $\beta=5$, mean recovery = $20 \pm 16\%$, Figure 3.3). All specified goals of precision did not appear to be substantially impacted (Table 3.3). However, it can be noted that the poorer recovery has moderated the tendency for inflation of 95% HDI widths owing to additional non-detects observed when sample size was low ($n=4$).

3.5 Discussion

3.5.1 Relatively few samples may be sufficient for precise mean concentration estimation

While it is intuitive that larger sample sizes will reduce the uncertainty surrounding both point and interval estimates of mean concentration, there are diminishing returns that are intricately linked to the source concentration and the recovery anticipated. Although the uncertainty can be rather large with a minimal number of samples ($n=4$), the probability of attaining a posterior mode mean concentration that differs by more than 0.5-log_{10} from the true source concentration is practically insignificant for the range of scenarios investigated. While this may suggest that utilities providing treatment much beyond what is deemed necessary at low protozoan pathogen levels can forego a more intensive sampling program to minimize costs, key caveats concerning stationarity, sampling program duration, sampling based on routine (non-temporally auto-correlated) monitoring are underscored. This retrospective power analysis can only be interpreted in a prospective sense if the historical parameters do not change in the future. In some environments (e.g., a deep subsurface aquifer that is not hy-

Table 3.3: Estimated power to achieve goal of precision as a function of sample size for the *Giardia* concentration parameters, with different mean recoveries simulated. All sample volumes are assumed to be 100 L with internally seeded matrix-spike data simulated to evaluate recovery. Values in tables are the estimated proportion of 100 simulations for which the specified goal of precision is attained (with 95% HDI lower and upper bounds on the estimated proportion in brackets, respectively).

No. of samples	a) high mean recovery (60%) $\mu=19.4/100$ L, $\sigma=25.3/100$ L $\alpha=5.3, \beta=3.5$		b) low mean recovery (20%) $\mu=19.4/100$ L, $\sigma=25.3/100$ L $\alpha=2, \beta=5$	
	posterior mode within $\pm 0.5\text{-log}_{10}$	95% HDI width $\leq 1.0\text{-log}_{10}$	posterior mode within $\pm 0.5\text{-log}_{10}$	95% HDI width $\leq 1.0\text{-log}_{10}$
4	0.82 (0.74, 0.89)	0.32 (0.23, 0.41)	0.72 (0.63, 0.80)	0.18 (0.11, 0.26)
8	0.96 (0.91, 0.99)	0.65 (0.55, 0.74)	0.85 (0.77, 0.91)	0.60 (0.50, 0.69)
12	0.97 (0.92, 0.99)	0.93 (0.87, 0.97)	0.98 (0.94, 1.0)	0.84 (0.76, 0.90)
24	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)
52	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)	1.0 (0.97, 1.0)

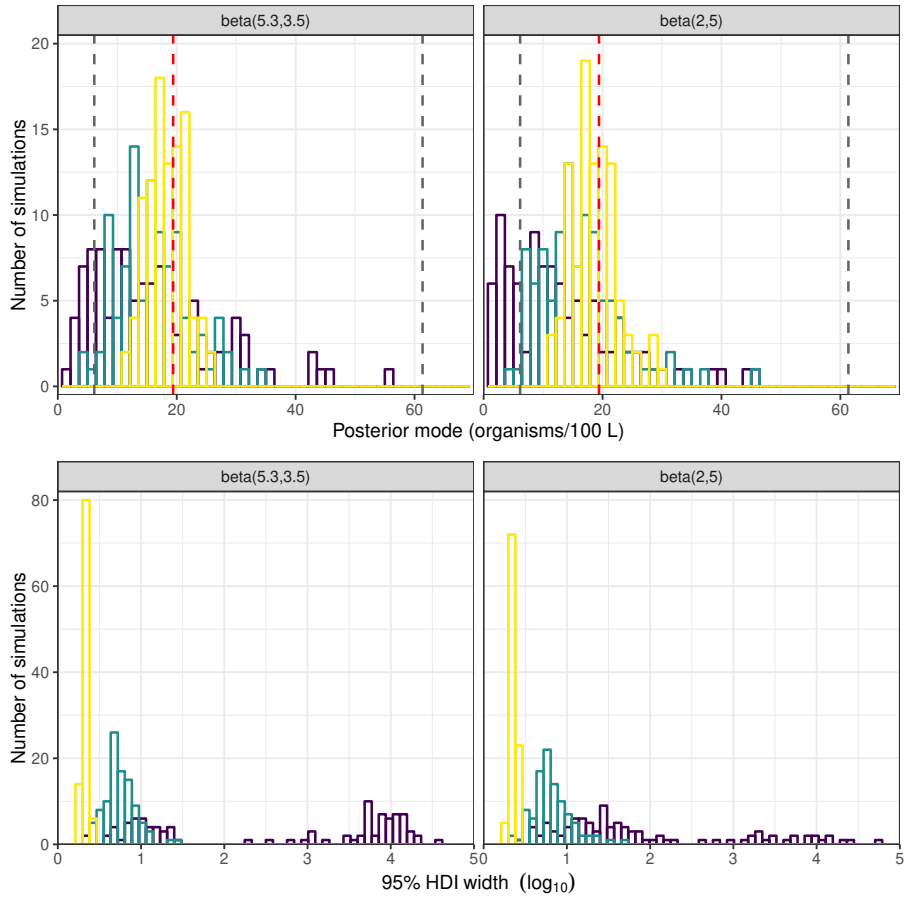


Figure 3.3: Histogram of posterior mode of the mean concentration (top) and 95% HDI width (bottom) estimated from 100 simulations each of the *Giardia* parameters with 4 (purple), 12 (green), and 52 (yellow) samples with paired recovery data simulated with a mean recovery of 60% (left, $p_{recovery}$ beta[5.3,3.5]) and 20% (right, $p_{recovery}$ beta[2,5]). The red dashed line represents the reference mean concentration value used to simulate the data; grey dashed lines indicate $\pm 0.5\text{-log}_{10}$ from the reference value.

draulically connected to the surface) this caveat may hold; however, a minimal routine monitoring program would still be prudent to allow for the possibility of microbial source water quality changes. The sampling program duration must adequately capture a range of representative conditions, including high and low concentrations driven by seasonal generation and hydrologic patterns (Efstratiou et al., 2017; Ongerth, 2017), yet consideration should also be given to avoid temporal auto-correlation during sample collection. A final caveat is that a small sample size is not likely adequate if concentrations are severely non-gamma distributed (e.g., bi-modal); in these cases more samples may be required as consistent with the Central Limit Theorem.

3.5.2 Minimum treatment performance targets can effectively alleviate implications associated with low, highly uncertain mean concentration estimates

While overestimation of mean source concentrations bears the implication of costs associated with excessive treatment, the underestimation of mean source concentrations bears the potential consequence of inadequate protection of public health. However, the prescription of minimum treatment requirements at low mean source concentrations (e.g., as represented by *Cryptosporidium* historical parameters used in this work) can effectively alleviate implications associated with the underestimation of low protozoan pathogen levels that are generally more uncertain compared to mean estimates that typically yield higher counts. This is consistent with the recommendations of the World Health Organization for providing QMRA-based performance targets as requirements for the reduction of enteric protozoa. Health Canada currently recommends a minimum 3-log_{10} reduction requirement for protozoan pathogens. The US EPA generally stipulates minima of 2- and 3-log_{10} reductions for *Cryptosporidium* and *Giardia* respectively. Neither the European Union nor the Australian National Health and Medical Research Council have established minimum treatment requirements for protozoa to date (Health Canada, 2019). However, the increasing adoption of QMRA frameworks across a number of these jurisdictions (Bichai and Smeets, 2013) inevitably allows for the stipulation of tailored minimum system-specific treatment requirements for protozoan pathogens.

3.5.3 Point- and interval-based goals reflect different aspects of precision

Both point- and interval-based goals of precision in mean concentration estimation were evaluated. The use of the posterior mode relative to the true mean concentration as a goal of precision appears to be more readily achieved with fewer samples for the simulations based on the lower mean concentration parameters (*Cryptosporidium*) than for the higher mean concentration parameters (*Giardia*). The goal of precision related to the 95% HDI width exhibits an apparent trend to the contrary; it is more readily achieved with fewer samples for *Giardia* parameters than for *Cryptosporidium* parameters. Considering that the historical analytical recovery parameters simulated and estimated for both protozoa are not substantially different, this apparent discrepancy is essentially an effect of the difference in magnitude between the protozoan concentrations. It underscores that while both goals involve an allowable precision that is equivalent to a 1.0-log_{10} span, they do not represent precision in the same way; the goals of precision are not equally stringent. The point-based criterion is an explicit representation of precision only made possible through simulations as the true mean concentration is known, as in the case of the retrospective power analysis approach applied herein. The posterior mode provides the mean protozoan concentration that is most likely given the data. Even if this value is adequately precise (i.e., $\pm 0.5\text{-log}_{10}$ of the true mean), a substantial fraction of the posterior distribution of mean concentrations may be much beyond the precision limit selected. On the other hand, the 95% HDI represents the range of mean protozoan concentrations that are most probable (with 95% probability), and therefore is a much more stringent goal of precision. Aside from the more intuitive interpretation of the HDI width, it is also determinable when the true mean is unknown as it is in practice. Nonetheless, the 95% HDI width evaluated based on sampling data requires careful implementation and computations in a probabilistic framework. Artefacts associated with the implementation of the model in the software used (e.g., inflated HDI widths as previously noted) must be taken into consideration.

3.5.4 A broadly applicable framework for evaluating the level of precision attainable based on monitoring program design to support QMRA efforts

Amongst the range of hypothetical scenarios investigated, monitoring programs with 52 and 24 samples collected from the source will practically ensure that 95% HDI

widths of at most 1.0-log_{10} for *Cryptosporidium* and *Giardia* mean concentration distributions are attained, respectively. Practically, this corresponds to a weekly/biweekly monitoring program for reliable estimation ($\pm 1.0\text{-log}_{10}$) of mean annual concentrations. In fact, 95% HDI widths of at most 0.5-log_{10} can consistently be achieved when higher mean concentrations are observed, as is the case for the *Giardia* historical parameters. It should again be noted that the stringency of the approach used herein focuses on limiting the worst precision (i.e., the specified HDI width will be attained in essentially all simulations within each sampling scenario) rather than average precision (i.e., all simulations of each sampling scenario will attain the specified HDI width, on average) or another measure of precision (e.g., posterior mode compared to the true data-generating mean). This implies that utilities implementing protozoan monitoring programs with this sample size can be highly confident in the precision of their mean source protozoan concentration estimates. Notably, comparable sample sizes have already been adopted into decision-making frameworks of several jurisdictions. For example, *Cryptosporidium* compliance monitoring for utilities serving over 10,000 people in the United States require the collection of monthly samples for 24 months. In the Dutch Inspectorate Guideline 5318 (translated and referenced within Schijven and de Roda Husman, 2011), a range of sampling frequencies spanning those examined (6, 13, 26) was mandated for index pathogens (including *Cryptosporidium* and *Giardia*). Therefore, this work provides a framework that can be broadly applied to other historical data sets. This facilitates quantification of the level of precision attained to inform sampling program design to support QMRA efforts.

3.5.5 Sample size enlargement can enhance precision in mean concentration estimates more than performing additional internally seeded matrix-spikes to quantify analytical recovery

In the absence of an independent recovery study (Schmidt et al., 2010), a handful of internally seeded matrix-spikes may be sufficient to characterize the recovery profile and supplement the native protozoan monitoring program without substantially compromising estimates pertaining to the mean concentration. This holds true even at substantially poorer mean recovery values; when counts are low, additional source water quality samples (and thereby the cumulative volume analyzed) are more informative to mean concentration estimation than additional characterization of the recovery profile (Emelko et al., 2008). Indeed, at levels of analytical recovery typified by the historical dataset (mean recovery $\approx 60\%$) and simulated in this work, the contribution

of analytical recovery to uncertainty in the mean concentration would be less than an order of magnitude. While we agree with Petterson et al. (2007) that internal recovery estimates for every protozoan assay is not likely necessary to limit uncertainty in mean concentration estimation, a much smaller recovery dataset than the 20 recommended by Petterson may be sufficient, assuming that the recovery analyses performed are adequately representative across a range of water matrices. A critical difference in the approach used was the handling of analytical recovery as a random variable independent of the observed counts. Approaches that adjust microbial concentration estimates by independent, random recovery values have been demonstrated to introduce bias that can exceed an order of magnitude (Schmidt et al., 2013). This has been demonstrated to bias concentration estimates, sometimes exceeding an order of magnitude when low analytical recovery values are common (Schmidt et al., 2013).

3.6 Conclusions

To quantify the attainable precision associated with mean protozoan concentration estimates, a range of protozoan sample sizes and number of internally seeded matrix-spikes to quantify analytical recovery was investigated using simulated data. The following conclusions can be drawn from this analysis:

- For the scenarios considered, mean concentration estimates within 0.5-log_{10} of the true data-generating mean was attainable even with a low number of samples (e.g., $n=4$).
- Although as few as four samples may provide adequately precise estimates of the mean protozoan pathogen concentrations, practical considerations of stationarity, sampling program duration, lack of temporal auto-correlation between samples, and deviations from gamma-distributed concentrations must be considered.
- A more stringent goal for attaining 95% HDI widths of $<1.0\text{-log}_{10}$ on the posterior mean concentration was attained on the order of 24 to 52 samples for the scenarios considered.
- As the true mean protozoan concentration is ultimately unknown in practice, the HDI width provides an intuitive representation of the attainable precision for a given sample data set. However, the computed HDI width can be sensitive to the implementation of the probabilistic model and tools applied and requires judicious interpretation.

- Underestimation of mean source concentrations bears the potential consequence of inadequate protection of public health (through prescription of inadequate treatment).
- A minimum treatment target at low mean concentrations (as can be determined through a reverse QMRA) can effectively alleviate the increased uncertainty associated with the characterization of these low mean concentrations.
- When enumeration of native protozoa from source water quality samples yield low counts, additional source samples (and thereby the cumulative volume analyzed) are more informative to mean concentration estimation than additional characterization of the analytical recovery profile.

Chapter 4

Illuminating subsurface microbial water quality patterns using adenosine triphosphate and dynamic time warping approaches

Alex H.S. Chik
Monica B. Emelko
Alfred P. Blaschke
Jack F. Schijven

Preprint accepted for publication in *Groundwater Monitoring and Remediation* (2020)

4.1 Summary

Aquifer microbial water quality evaluations are often performed by collecting ground-water samples from monitoring wells. While samples collected from continuously pumped sources are seldom disputed as representative of the aquifer, natural biofilm present in the vicinity of well screens may introduce unwanted microbial artefacts in monitoring wells that are only periodically sampled. The need for well water purging to obtain samples void of these artefacts has been widely recognized. However, purging methods are not standardized; many approaches presume that physico-chemical water quality stability achieved through the removal of 3 to 5 well volumes is indicative of the stability of target analytes. Using a dataset collected from a shallow unconfined aquifer in Southern Ontario, Canada, the need for using dedicated approaches that account for the time-dependent nature of microbial water quality changes was demonstrated. Specifically, the utility of adenosine triphosphate (ATP) as a rapid, field-ready biochemical indicator of microbial water quality stability was investigated. This work shows that ATP concentrations reflect time-limited (bio)colloid transport processes that are consistent with other microbial water quality parameters monitored, but different from commonly measured physical and chemical water quality indicators of adequate well purging. ATP concentrations occasionally fluctuated even after 3 or 4 hours of purging, indicating that microbial artefacts attributable to subsurface biofilms in the vicinity of the well screen can still persist. The recurrence of characteristic ATP patterns in each well were systematically examined through the novel application of dynamic time warping (DTW), a non-parametric time series analysis approach. These patterns are believed to be linked with seasonal hydrogeological conditions, which warrant consideration in the design and interpretation of subsurface microbial water quality investigations.

4.2 Introduction

Representative samples of aquifer microbial water quality—specifically, the composition and concentrations of suspended microorganisms in the aquifer pore water—are crucial for aquifer source vulnerability assessments to pathogen/fecal contamination. However, the ubiquity of microorganisms presents a significant challenge to ensuring the integrity and representativeness of samples. For example, Harter et al. (2014) demonstrated the need for extended well purging to control for significant fecal microbial indicator contamination of the well-head, within the well casing and within the

immediate aquifer vicinity of the well screen. Although there is general alignment to suggest that extended purging/continuous operation of wells would all but ensure that a sample representative of the microbial composition in the aquifer can be collected (Cullimore, 2007; Harter et al., 2014), it is often impractical or infeasible. Therefore, evaluations of subsurface microbial water quality parameters can greatly benefit from an inexpensive, sensitive field-ready tool that can support the contextualization of microbial results and aid in determining the minimal level of purging.

Owing to the contrast in hydrogeochemical conditions between a well and the surrounding aquifer, naturally occurring conglomerates of sessile microorganisms and their metabolites—collectively known as biofilm—form distinctive patterns on solid surfaces in the vicinity of a well (i.e., well casing, well screen, and sediment surfaces). These subsurface biofilms are known to underpin microbial water quality when water is initially pumped (Cullimore, 2007). The dynamic hydraulic conditions invoked by groundwater pumping can cause biofilm mobilization and entrainment (Cullimore, 2007; Smith and Comeskey, 2009), therefore potentially introducing unwanted microbial artefacts into a water sample. Intuitively, aquifer water quality during extended purging is less impacted by biofilm and increasingly representative of suspended microorganisms (Cullimore, 2007; Smith and Comeskey, 2009; van Driezum et al., 2017; Korbel et al., 2017). The biofilms can be characterized throughout well purging activities by examining “zones of interrogation projections” in wells—increasingly sparse time-based intervals—as proposed by Cullimore (2007). These zones feature irregularly spaced sampling intervals that focus on capturing and describing microbial water quality shifts at key points in time during purging and in absence of continuous monitoring.

Accordingly, it is both desirable and pragmatic to establish the minimum length of time—using relevant microbial indicator(s)—required for well purging activities to limit artefacts attributable to subsurface biofilms. Amongst a multitude of potential microbiologically relevant indicators, adenosine triphosphate (ATP) is a biochemical measurement of microbial activity suggested to be useful in the delineation of biofilm expanse and microbial densities (Webster et al., 1985; Jensen, 1989; Metge et al., 1993; Hammes et al., 2010). ATP is ubiquitous in living cells and lost rapidly from dead cells; it is present in fairly constant concentrations in microorganisms and can be measured rapidly with high sensitivity using bioluminescence when assayed with the firefly enzyme luciferase (McElroy, 1947). In highly oxic conditions, elevated ATP concentrations reflect elevated microbial densities (McCarthy, 1991; Abelho, 2005; van der Kooij et al., 2017) or enhanced metabolism of fast-growing microorganisms without substantial nutrient limitations (Howsam, 1988; Knezev and van der Kooij, 2004). The relative microbial activity as manifested by ATP fluctuations monitored throughout

well purging activities has been regarded as an indicator for characterizing subsurface microbial biofilms (Cullimore, 2007). These measurements hold particular promise as they are field-ready and have been recently adapted in laboratory instrumentation for near real-time monitoring applications (e.g., Hach[®] EZ7300 microbial load ATP analyzer).

ATP behaviour during well purging activities has not been reported; the evolution of its behaviour over spatial and seasonal time scales has also not been documented. Therefore, the main goal of this work was to examine the potential of ATP measurements as an indicator of purging sufficiency for collecting aquifer-representative microbial water quality samples. We examine ATP concentrations in groundwater samples collected sequentially throughout concurrent well purging activities at two locations developed in the same shallow unconfined aquifer, exhibiting practically equivalent physical and chemical water quality characteristics. The following questions were therefore addressed: (i) how do ATP concentrations fluctuate throughout well purging and (ii) how do they compare with other physical and chemical water quality parameters measured concurrently? (iii) Will ATP concentrations measured simultaneously in these two wells converge to suggest microbial water quality conditions that are increasingly representative of the aquifer? Lastly, we explore (iv) spatio-temporal scales over which ATP patterns are observed to gain insights related to the dynamics of subsurface biofilms. For this purpose, the novel application of a parameter-free approach that leverages information within sequential ATP measurements was demonstrated. While ATP concentrations may be analyzed solely based on their magnitude and summary statistics, additional insights arising from the relative microbial activity throughout purging may be overlooked unless underlying features captured by the sequentially collected data are considered. However, given the irregularly-spaced sample collection intervals applied throughout purging that preclude parametric time series analysis approaches, a non-parametric time series analysis approach—dynamic time warping (DTW)—was applied. This approach facilitates the comparison of time series of differing lengths and irregular sampling intervals by generating a metric to describe the (dis)similarity of any features exhibited. This metric, the DTW distance, was calculated for all pairs of time series and subsequently ordinated using non-metric multidimensional scaling to portray the collective relationships between ATP patterns and explore potential spatio-temporal influences.

4.3 Materials and methods

4.3.1 Study site

The study site (Figure 4.1) is situated near drinking water production wells six kilometers south of Woodstock, Ontario, Canada. The topography of the site comprises rolling hills and drumlin features. This site was selected because this glacial outwash aquifer system features extremely high groundwater velocities and is highly aerobic (Devlin et al., 2012). Therefore, it provides ideal conditions that strengthen the aforementioned relationships between ATP concentrations and microbial densities (McCarthy, 1991; Abelho, 2005; van der Kooij et al., 2017). The geochemistry within this aquifer system is known to be rather uniform (Critchley et al., 2014). Regional groundwater flow is generally in a southeasterly direction (Critchley, 2010).

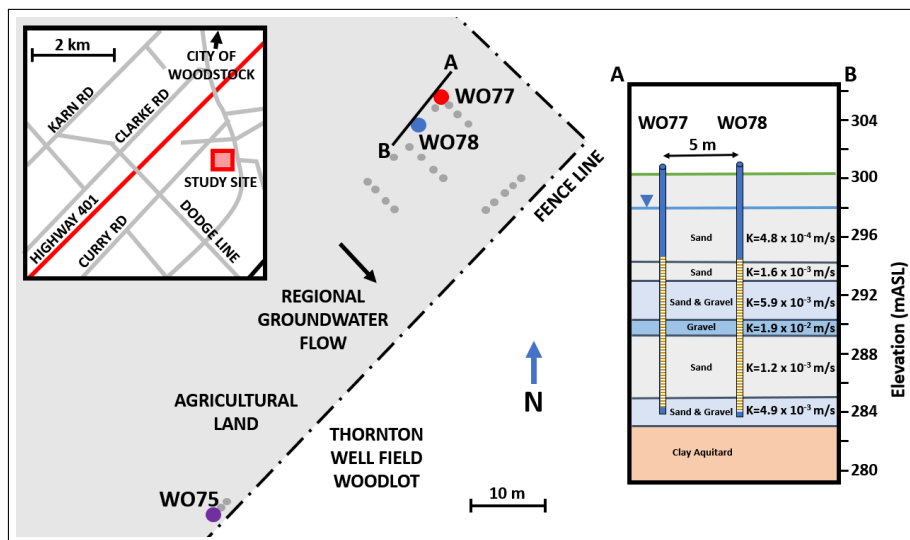


Figure 4.1: Study site and monitoring well locations. Monitoring wells used only for regional hydraulic gradient estimation in this study are represented by grey circles. Cross-section of transect A-B shows the hydrogeological conceptual model previously developed at the site (adapted from Critchley, 2010).

The monitoring wells (WO77, WO78) of interest are developed in the shallowest unconfined aquifer of the regional aquifer system, comprising primarily of sand and gravels interspersed with discontinuous silty till aquitard units (Haslauer, 2005; Critchley, 2010; Critchley et al., 2014) (inset transect cross-section, Figure 4.1). These

wells are situated 5 meters apart in a direction perpendicular to regional groundwater flow. Both monitoring wells (50.8 millimeter [mm] diameter, polyvinyl chloride [PVC] construction) were screened at similar depths, between 5.91 m to 16.57 m below ground surface (mBGS) and vented to the atmosphere. The average water table elevation occurred at approximately 298 meters above sea level [mASL] (≈ 3 mBGS), which was consistently above the top of the screened interval. The aquifer hydraulic conductivity in the screened intervals was estimated to range from 4.8×10^{-4} to 1.9×10^{-2} m/s (based on grain size analysis, flowmeter test, point velocity probes [PVP] measurements (Devlin et al., 2012), tracer tests, and three dimensional finite difference model calibration). Horizontal velocities were estimated to be between 1 m/day and about 13 m/day using a solute tracer study. Hydraulic conductivity profiles generated between these two wells exhibited remarkable similarities (Critchley, 2010). Additional well installation details are provided in Critchley (2010).

Groundwater levels were regularly monitored in 28 additional piezometers within a 500 m radius of the two wells of interest; the mean regional groundwater gradient over the study period was estimated to be 0.0069 using HydrogeoEstimatorXL (Devlin and Schillig, 2017). Groundwater levels were highest in the spring months (April/May) and lowest during late fall/early winter (December) (Appendix C1). Annual recharge at this location is estimated to be 396 mm/year (Koch, 2009). Local infiltration travel time through the unsaturated zone to the water table is estimated to be on the order of 2.8 to 5.6 years (Sousa, 2013). These observations are generally consistent with the seasonal hydrogeological fluctuations observed historically at this site (Haslauer, 2005; Koch, 2009; Christie et al., 2009; Critchley, 2010; Brook, 2012). Water level measurements do not support vertical flow at this site to be a significant factor (Critchley, 2010).

4.3.2 Field and laboratory analyses

Twelve groundwater sampling events occurred between May 2017 and May 2018. A minimum of two weeks between sampling events was used in this study to limit interferences from irreversible subsurface biofilm perturbations associated with the preceding sampling event(s) (Lundkvist et al., 2007; Tolhurst et al., 2008; Worley-Parsons, 2015). During each event, a series of samples was collected from one or both monitoring wells throughout continuous well purging from quiescence to over three hours time. All samples were collected using dedicated pumps with PVC tubing that were pre-sterilized in the laboratory by soaking and pumping a dilute bleach solution (0.6% sodium hypochlorite) through the tubing for a minimum of one hour followed by a sterile deionized water flush. Sterile deionized water was also used on site to prime the

pumps (Flojet model No. 4105 Series diaphragm pump, Irvine, California, USA) as necessary. Water levels were measured prior to pumping and the rate of water extraction from the wells was tracked periodically (average pumping rate = 5.8 litres/minute [L/min]). The pumping rates applied are consistent with the goal of minimizing draw-down within the well as in low-flow purging to minimize artificial mobilization and entrainment of particulates (Puls and Barcelona, 1996; Barcelona et al., 2005); however, low levels of shear stress under laminar flow regimes are still expected to result in some degree of biofilm detachment, mobilization and entrainment processes (Rittman, 1982). Temperature, dissolved oxygen, pH and electrical conductivity were measured on-site throughout the well purging process using dedicated portable multiparameter meters (YSI Quattro Professional Plus, YSI Inc./Xylem Inc., Ohio, USA). Turbidity was also monitored using a portable turbidimeter (Hach®, Colorado, USA). All probes and meters were calibrated using standards as per the manufacturers' instructions.

Additional parameters relevant for microbial water quality were measured for selected samples. Flow cytometry was deployed for the evaluation of microbial cell densities using the FACSCaliburTM flow cytometer (BD Biosciences, New Jersey, USA). Culture-based methods, in the form of Biological Activity Reaction Tests (BARTTM, Droycon Bioconcepts Inc., Regina, Canada) for iron-related bacteria (IRB BARTTM), sulfate reducing bacteria (SRB BARTTM), slime-forming bacteria (SLYM BARTTM), and algae (ALGE BARTTM including grass-green algae, blue-green algae, desmids, diatoms and euglenoids) were also deployed in select samples to corroborate the relative abundance of specific groups of microorganisms. A molecular method for microbial community analysis (16S rRNA gene amplicon sequencing) was performed concurrently with some samples. As these analyses were the focus of another study (Chik et al., 2020a), only key findings relevant for the interpretation of the ATP data are presented in Appendix C4 and discussed. The parameters were first monitored in WO78 for four additional sampling campaigns prior to initiating sample collection from WO77 on the fifth sampling event.

4.3.3 Determination of ATP concentrations

All ATP measurements were determined using the Lumitester C-110 luminometer (Kikkoman Food Products Company, Tokyo, Japan) with the Quench-Gone Aqueous (QGATM) test kit (LuminUltra, Fredericton, New Brunswick, Canada) in accordance to the manufacturer's procedures (compliant with ASTM Standard D4012) in the laboratory. Briefly, 120 mL of sample was collected and passed through a 25 mm diameter glass microfiber syringe filter with a 0.7 micron (μm) nominal pore size (WhatmanTM

GD/X, Florham Park, New Jersey, USA). Internal tests conducted by the manufacturer have ascertained the ability of these depth filters to capture microorganisms down to $0.2 \mu\text{m}$. The larger analytical volume was used compared to that recommended by the manufacturer (25 to 50 mL for this type of source water matrix) to improve method sensitivity, which would therefore allow for the quantification of microbial activity levels anticipated to be near the nominal method detection limit. The syringe filter was first removed from the syringe barrel to avoid the application of excessive negative pressure to the filter membrane upon removal of the syringe plunger, and reattached to the barrel. One (1) millilitre (mL) of UltraLyse 7 was added to the barrel and passed through the filter and collected in a new 9 mL UltraLute (dilution) tube. The sample was capped and inverted three times to mix the contents. Dedicated pipette tips were used to transfer $100 \mu\text{L}$ of the contents of the dilution tube and $100 \mu\text{L}$ of luminase enzyme to a new assay tube. The tube was gently swirled five times and immediately inserted into the luminometer for measurement. The RLU (relative light units) associated with the cellular ATP is measured using the luminometer, recorded and converted to cellular ATP concentration in units of picograms/mL (pg-ATP/mL). The luminometer was calibrated for each set of samples analyzed as recommended by the manufacturer’s instructions. All ATP measurements were conducted on the same day of sample collection. Duplicate aliquots were collected every fifteen samples to characterize intra-sample variation; negative controls were also measured to evaluate potential cross-contamination throughout sample collection and analysis. A total of 20 ATP time series between two wells were collected over a year at near-monthly intervals; the collection of additional time series would be necessary to facilitate more rigorous statistical tests and was beyond the scope of this initial proof-of-concept demonstration.

4.3.4 Non-metric multidimensional scaling of ATP time series using DTW distances

Dynamic time warping (DTW) was originally developed for speech recognition to facilitate pairwise comparisons between sequences of different lengths with irregularly spaced observations (Sakoe and Chiba, 1978; Mueen and Keogh, 2016). DTW has been recently used to elucidate spatial, temporal and/or seasonal dynamics for phosphorus transport in watersheds (Dupas et al., 2015), geophysical seismic images to delineate geological strata (Hale, 2013), and sewer flow monitoring (Dürrenmatt et al., 2013). Here, the novel application of time series analysis using DTW to investigate ATP patterns is demonstrated. This technique realigns the most similar features of each time series to those of another based on imposed constraints, before a distance metric

representing the (dis)similarity between the time series is computed. This temporal realignment allows for time series of differing lengths and irregular sampling intervals to be compared, based on morphological features of the biofilm as manifested by the ATP time series. During a given purging sequence, peaks in microbial parameters (such as ATP) have been interpreted as the location(s) of elevated microbial biomass attributable to enhanced biofilm growth in the vicinity of the well from which water is abstracted (Cullimore, 2007). As the *relative* microbial activity is of interest, each ATP time series was normalized by re-scaling each ATP concentration measurement between the minimum and maximum values for each well and sampling event. Thus, the maximum ATP value within a time series would be indicative of the foci of biomass from subsurface microbial biofilms while lower values can be expected with extended purging (Cullimore, 2007). Details of the calculation of DTW distances are provided in Appendix C2. In this work, DTW was implemented using the *dtw* package (v1.20-1) in R (Giorgino, 2009) with local Euclidean distances. These distances were computed for all pairs of time series and compiled in a distance matrix.

Non-metric multidimensional scaling (NMDS) is a robust, indirect ordination approach that can be applied to any (dis)similarity or distance matrix involving quantitative, semi-quantitative, or qualitative variables (Kenkel and Orloci, 1986). A low-dimensional portrayal of ATP time series relationships was generated using NMDS, which iteratively places each time series in a position that reflects the order of the pairwise distances calculated. Accordingly, the scale of the axes and the ordination of the plot are arbitrary and do not reflect the magnitude of the pairwise distances. The emergence of patterns in the ordination allows for the corroboration of existing knowledge, generation of hypotheses, or design of further sampling campaigns to target any observed variation(s) (Kenkel and Orloci, 1986). It also provides an exploratory tool to contextualize the strength of the (dis)similarities between the objects ordinated. NMDS solutions with stress values above 0.20 should be interpreted with caution and those with stress above 0.30 are highly suspect (Buttigieg and Ramette, 2014). The NMDS ordination approach has been discussed extensively elsewhere (Kenkel and Orloci, 1986; Buttigieg and Ramette, 2014).

4.4 Results

4.4.1 Physical and chemical water quality characteristics

Physical and chemical water quality parameters (Appendix C3) were generally consistent with those observed historically at this site (Critchley et al., 2014). In accordance with common practice for groundwater chemistry evaluations, purging was considered adequate when pH stabilized within 0.1 Standard Units (SU), electrical conductivity fluctuated by less than approximately 5%, dissolved oxygen stabilized within 0.2 mg/L or 10% saturation (whichever is greater) and/or turbidity either stabilized or fell below 10 Nephelometric Turbidity Units (NTUs) (Striggow, 2017). Temperature is subject to rapid changes when collected for parameter measurement and therefore was not typically used for determining well purging adequacy (Striggow, 2017). In all sampling campaigns, these criteria were all achieved within the time taken to purge 3 to 5 well volumes, which corresponds to approximately 10 to 15 minutes depending on exact purging flow rate.

To facilitate comparisons between the various parameters measured, the coefficient of variation was evaluated for a running window of three consecutive measurements for each parameter throughout purging (Figure 4.2). The apparent fluctuations exhibited by turbidity are attributable to a large standard deviation relative to the low turbidity values measured (i.e., consistently less than 10 NTUs). Meanwhile, the maximum coefficients of variation observed for pH, electrical conductivity (EC), dissolved oxygen (DO), and temperature were 3%, 6%, 30%, and 8%, respectively. The largest fluctuations in dissolved oxygen concentrations were traceable to measurements taken when the flow-through cell was disturbed during purge water sampling.

4.4.2 Adenosine triphosphate measurements

Groundwater microbial ATP concentrations ranged between 0.046 and 58.6 pg-ATP/mL. Median ATP concentrations across all sampling events were 0.27 pg-ATP/mL (n=126) and 0.38 pg-ATP/mL (n=190) for wells WO77 and WO78, respectively. Across sampling events, ATP concentrations exhibited a slight decreasing trend from the end of summer (sampling event 5) to late spring (sampling event 12) (Figure 4.3). For sampling events during which ATP concentrations were measured concurrently in both wells, ATP concentrations were generally higher in WO78 than in WO77, presumably due to local subsurface biofilm heterogeneities. A notable exception was the

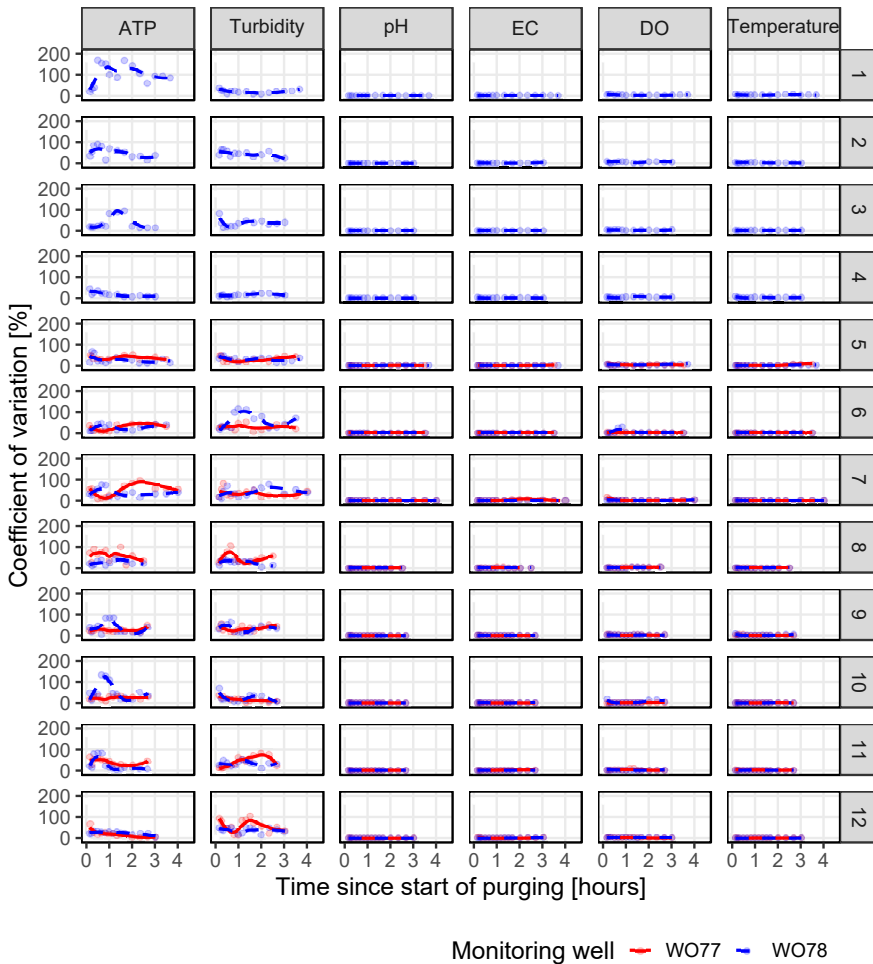


Figure 4.2: Stabilization history of groundwater quality parameters monitored, expressed as the coefficient of variation evaluated for running windows of three consecutive measurements throughout purging. Red and blue markers/lines are used to denote monitoring wells WO77 and WO78, respectively. Each row of the plot denotes a sampling event (corresponding dates provided in Figure 4.5). The panel columns correspond to the following parameters: adenosine triphosphate (ATP), turbidity, pH, electrical conductivity (EC), dissolved oxygen (DO), and temperature.

ATP concentrations observed in WO77 during sampling event 8 (Figure 4.3), which was likely attributable to a higher average flow rate (12 L/min) sustained by the al-

ternate pump used (Simer 2825SS, Delavan, WI, USA). Within a purge sequence, the coefficient of variation evaluated for three consecutive ATP measurements fluctuated as high as 168%, but was usually less than 50% towards the termination of most sampling events (Figure 4.2). Ninety-five percent confidence intervals on the paired differences of ATP concentrations between wells all contained zero, regardless of the extent of purging achieved (Table 4.1). However, the margins of error calculated for these confidence intervals progressively reduced with extended purging, decreasing by nearly an order of magnitude (0.70 to 0.08 pg-ATP/mL) amongst samples collected in the first 30 minutes of purging ($n=35$) compared to samples collected after 1.5 hours of purging ($n=39$). This observation is consistent with the decreasing median of relative differences between ATP concentrations in both wells with extended purging (Figure 4.4).

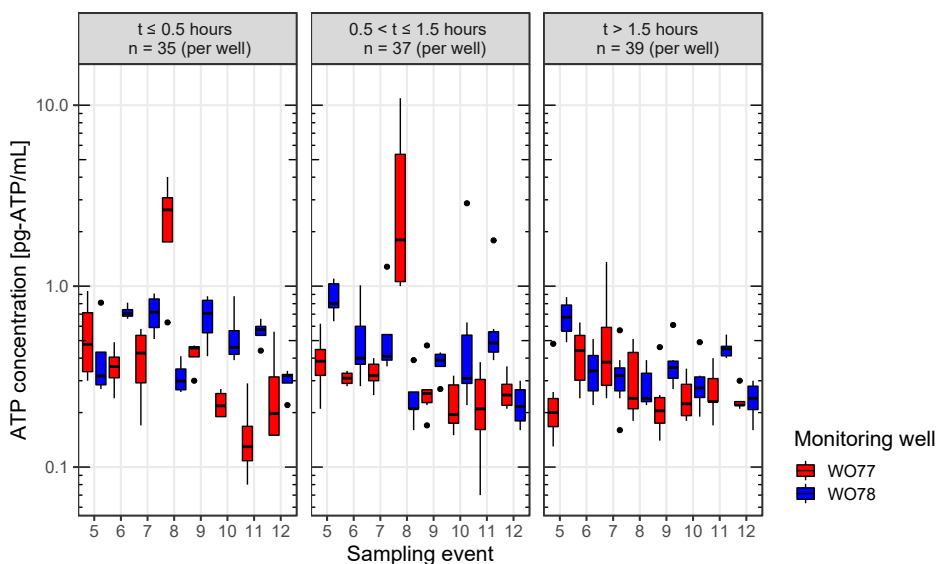


Figure 4.3: Stabilization of ATP concentrations throughout purging. Red and blue fill are used to denote monitoring wells WO77 and WO78, respectively. Each column panel denotes the stage of purging attained. The dates corresponding to each sampling event are provided in Figure 4.5.

The relative microbial activity (i.e., normalized ATP concentrations) in each well was used to characterize the foci of biomass from subsurface microbial biofilms during each sampling event (Figure 4.5). This normalization effectively addresses potential biases attributable to the individual sampling campaign and well-specific factors (e.g.,

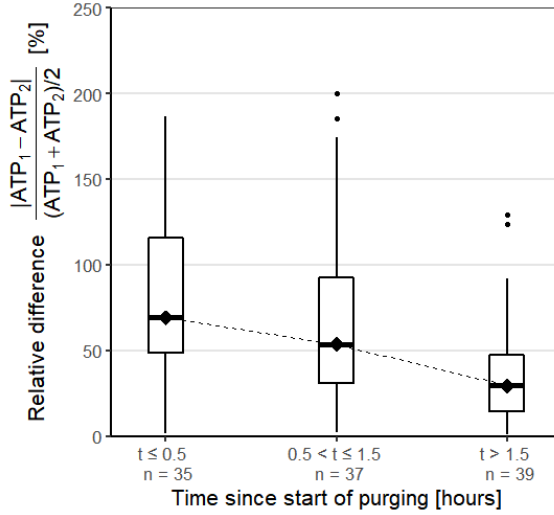


Figure 4.4: Box-and-whisker plots of relative differences between WO77 (ATP_1) and WO78 (ATP_2) paired ATP measurements throughout concurrent purging activities, where t represents the time elapsed since start of purging in hours and n reflects the number of paired samples falling within the specified period. The medians are marked by diamonds, the boundaries of the box indicate the 25th- and 75th-percentile, and the whiskers indicate the highest and lowest values of the results excluding extreme values.

Table 4.1: Paired t-test results comparing ATP concentrations between WO77 and WO78. Tests were performed against the alternative hypothesis that the true difference in means is not equal to zero.

Time since start of purging [hours]	n	mean difference [pg-ATP/mL]	standard error [pg-ATP/mL]	95% confidence interval [lower, upper bound]	t-value	p -value
$t \leq 0.5$	35	0.29	0.36	[-0.40, 0.99]	0.86	0.40
$0.5 < t \leq 1.5$	37	0.04	0.18	[-0.31, 0.39]	0.22	0.83
$t > 1.5$	39	0.01	0.04	[-0.06, 0.09]	0.36	0.72

steady-state flow rate attained, generally higher ATP concentrations in WO78). Comparisons of these patterns were made visually. A generally monotonic, decreasing trend was sometimes observed; however, elevated ATP concentrations can also be observed at later stages of purging. This indicates that the highest levels of microbial activity (and by association, densities) are not always highest in the immediate vicinity of the well casing. Recurring ATP patterns were noted in each well over consecutive sam-

pling campaigns in September and November (Figure 4.5, sampling events 6 and 7) and again in April 2018 (sampling events 10 and 11). During these periods, WO77 ATP patterns typically exhibited two peaks during the latter stages of purging, while a single prominent peak was exhibited within the first hour of purging in WO78. These peaks reflect order-of-magnitude changes in microbial activity. The ATP patterns observed on August 30, December 20, and May 30 (sampling events 5, 8, and 12, respectively) within each well also bear some resemblance to each other, albeit with some visually erratic fluctuations. Similar ATP patterns were also noted during the sampling event on January 30, 2018 (sampling event 9) *between* the wells. This event coincided with the first major increase in groundwater elevation (+20 cm) captured during the study period. To confirm that the variation in these data is not inherent to the method of analysis, additional microbial water quality analyses were conducted—these are described below and detailed in Table C4 of Appendix C.

4.4.3 Additional microbial water quality parameters

Microbial cell densities estimated using flow cytometry generally yielded sample concentration estimates (mean \approx 500 cells/mL, median \approx 60 cells/mL) below the limit of reliable quantification using this technique (i.e., \approx 1000 cells/mL) (Hammes and Egli, 2010) (Appendix Table S4). These estimated densities are consistent with those that can be estimated using ATP concentrations as microbial equivalents (by assuming an average of 0.001 pg-ATP/microorganism as per QGATM test kit manufacturer’s recommendations), which confirmed that cell densities were largely below this threshold. Despite the majority of flow cytometry cell densities falling below the limit of reliable quantification, a weak positive correlation between ATP concentrations and sample microbial cell concentration estimates was noted (Spearman’s $\rho=0.147$, $p=0.03$). 16S rRNA gene sequences attributable to bacterial taxa (e.g., *Paenibacillus* and *Rhizobia*) that are known to produce extracellular polymeric substances (EPS) and are prevalent in the biofilm of reactors mimicking aerobic groundwater conditions (Ross et al., 2001) were generally higher in abundance during the early and intermediate stages of purging (<60 minutes, corresponding with upwards of 10 well volumes purged, Chik *et al.*, [2020a]). Sequences linked with *Sphingomonadales*, which are known to co-aggregate with other bacteria and play a quantitatively important role in freshwater biofilm communities (Rickard et al., 2002), were also in higher abundance during the early and intermediate stages of purging. The vast majority of microbial taxa did not significantly increase after an hour of purging (Chik et al., 2020a). This is consistent with more prominent expression of BARTTM reactions observed for iron-related and slime-

forming bacteria conducted for samples collected during the early stages of purging. Low levels of sulfate reducing bacteria and algae were consistently observed (Appendix C4).

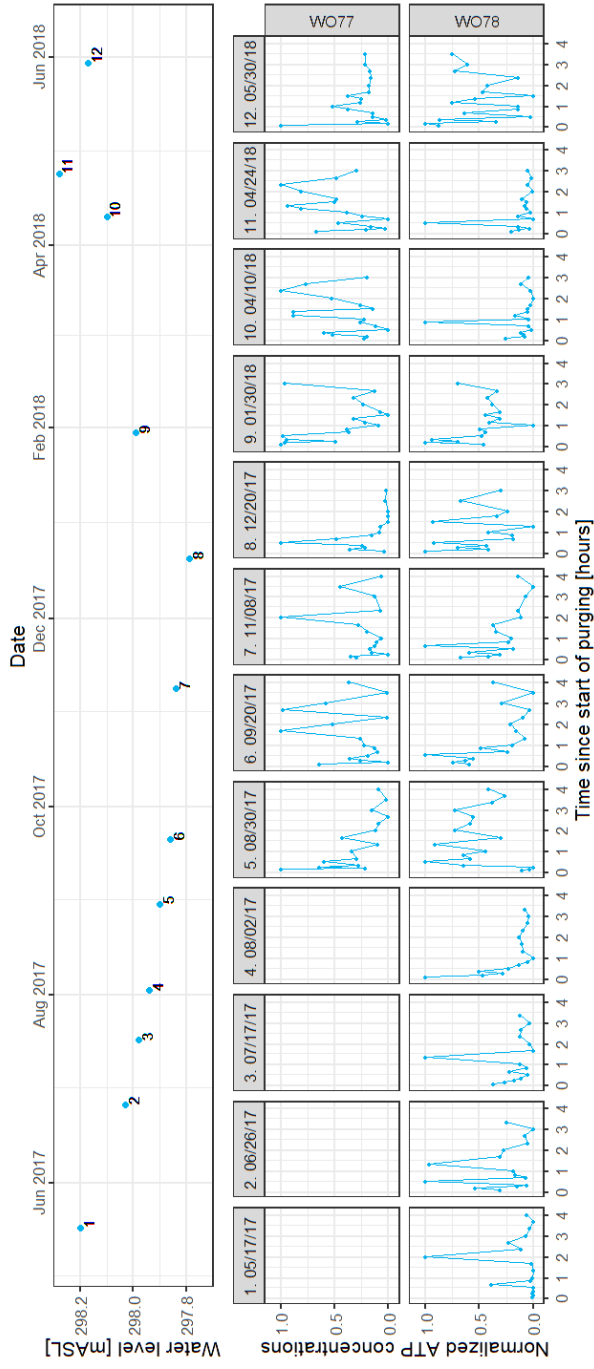


Figure 4.5: Normalized ATP concentrations observed throughout well purging activities. The positions of groundwater levels observed in WO78 during each sampling event is indicated above the plots of normalized ATP concentrations. Continuous groundwater levels monitored in a nearby well WO75 are summarized in Appendix C1.

4.4.4 Non-metric multidimensional scaling of ATP time series based on DTW distances

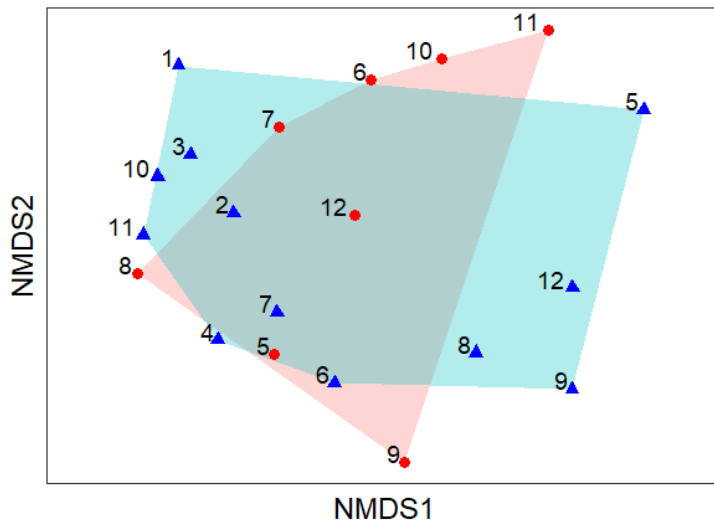


Figure 4.6: (Dis)similarities between all ATP time series portrayed through non-metric multidimensional scaling (NMDS). Each red circle or blue triangle represents an ATP time series corresponding to wells WO77 and WO78, respectively, and is labelled sequentially in order of sampling events. Markers appearing closer together reflect greater similarities between the time series. The shaded regions are consistent with the marker colors used for each well. This NMDS solution resulted in a stress value of 0.139, indicating that this solution captures the order of (dis)similarities reasonably well.

Pairwise (dis)similarities between ATP time series calculated as DTW distances were compiled in a matrix and ordinated using NMDS to generate a two-dimensional representation of these relationships (Figure 4.6). ATP time series that have similar features will typically exhibit a lower DTW distance metric than those that do not. Accordingly, NMDS uses the order of the distance metrics between ATP time series (each represented by a single marker) to position similar ATP time series closer together. ATP time series associated with each well and sampling event is represented by a colored marker (red circles for WO77 and blue triangles for WO78) denoted with the corresponding sampling event number. To aid visualization and interpretation, the regions represented by each of the wells were also shaded consistently with the marker colors.

Notably, the markers denoting ATP time series collected from the two wells during the same sampling campaign generally occupied different quadrants of this ordination, with the exception of the ATP time series collected from both wells during the sampling event on January 30 (Figure 4.6, sampling event 9). ATP time series collected during the same pairs of consecutive events (i.e., sampling events 6/7, 10/11) in each well can also be noted. These observations of (dis)similarities are consistent with those identified through a visual inspection of normalized ATP patterns (Figure 4.5). Notably, not all (dis)similarities can be perfectly portrayed by NMDS and DTW distances. For example, events 5, 8, and 12 were not clustered in this particular ordination. The inability of this ordination to perfectly represent all time series (dis)similarities is reflected by the stress value of this NMDS solution (0.139), which indicates their good, but not excellent portrayal.

4.5 Discussion

4.5.1 Physical and chemical indicators do not reflect purging adequacy for microbial water quality evaluations

As would be expected, ATP does not follow the same types of trends as other physical and chemical water quality parameters as it is directly and exclusively linked to the presence of microorganisms. Fluctuations in this biochemical must be interpreted with consideration of the method used to measure it. Hammes et al. (2010) demonstrated a strong, significant correlation between microbial ATP and estimated cellular densities using flow cytometry across a range of different source waters (including groundwater). While the same extent of correlation was not observed for this study (due to microbial densities falling below the threshold deemed reliably quantifiable by flow cytometry), a statistically significant correlation was nonetheless observed between these parameters. Microbial ATP is measured only after its extraction from a discrete number of living microorganisms—that are in essence, particles—in the sample. Strategies for informing microbial water quality that rely on ATP must therefore reflect particle transport behaviour rather than that of dissolved solutes. Particle and (bio)colloid transport through the subsurface are subject to advection, dispersion, physico-chemical filtration (attachment/detachment), size-exclusion, and straining processes (Ginn et al., 2005). Subsurface biofilms are conglomerates of microorganisms and their metabolites (Palmer and White, 1997). Thus, they are also subject to these processes once detached and entrained in pore water during purging (Liu et al., 2018). Accordingly, the break-

through of these conglomerates, and ATP by association, is likely to exhibit extended tailing resulting from attachment/detachment processes. Moreover, the transport of these conglomerates can lead to increased variability in microbial observations (i.e., overdispersion) relative to that which would be expected if microorganisms were not clustered. Patterns of ATP concentrations were consistent with the occurrence of these phenomena, thereby indicating its utility for signifying fluctuations in microbial water quality.

The use of continuous physical and chemical water quality measurements for the determination of purging adequacy has rarely been disputed for evaluations of chemical water quality sampling. However, this work underscores an important caveat: the use of these conventional indicators of purging adequacy are not likely appropriate for *biochemical* water quality parameters related to microbiological water quality. This disparity between the transport of dissolved-phase substances in contrast with the mobilization and transport of (bio)particles has been a widely observed phenomenon (e.g., Schijven and Hassanizadeh, 2000; Ginn et al., 2005; Bradford and Torkzaban, 2008; Emelko and Tufenkji, 2010; Molnar et al., 2015) and critically underscores the need to consider dedicated purge volume-based and purging time-based approaches for chemical and microbial water quality evaluation, respectively. Purge volume-based approaches presume that the substance measured in a sample occupies negligible volume within the voids of subsurface sediments in order to evaluate the position from which the sample originated relative to the well. Conversely, all particles occupy volume; biofilm biomass and other inorganic particles can reduce effective porosity through obstructing connected flow paths through the subsurface. When the degree of obstruction becomes significant, the estimation of the sample's position using purge volume based-approaches also becomes increasingly inaccurate. Therefore, extrapolations beyond the time of sample collection using the same purging protocol (e.g., similar flow rates, type of pump, etc.) has not been recommended for comparisons of field scale microbial water quality data acquired from purge water sampling (Cullimore, 2007).

4.5.2 ATP measurements can be useful for indicating aquifer-representative microbial water quality

The utility of ATP concentrations as indicators of minimal purging requirements for informing microbial water quality sampling was evaluated. ATP evaluation is useful in describing microbial activity and density (Webster et al., 1985; Jensen, 1989; Metge et al., 1993; Hammes et al., 2010). Over the past decades, advances of the ATP method to address possible matrix interference effects have resulted in greater sensitivity (limit

of detection [$\text{LOD}] \leq 0.1 \text{ pg-ATP/mL}$) than the first ATP assays developed ($\text{LOD} \geq 10 \text{ pg ATP/mL}$). Thus, it should be useful in differentiating between microorganisms suspended in aquifer pore water and the often higher densities of sessile microorganisms on well screens and unconsolidated sedimentary aquifer materials in the vicinity of wells (Harvey et al., 1984; Hazen et al., 1991; Griebler and Lueders, 2009; Sorensen et al., 2013), especially at highly aerobic conditions that should strengthen the correlation between ATP concentrations and microbial densities (McCarthy, 1991; Abelho, 2005; van der Kooij et al., 2017). Its sensitivity relative to other microbial water quality metrics lends itself for its use as a discriminatory indicator.

Recognizing that potentially dynamic aquifer water quality could be confounded with measurable changes in ATP concentrations attributable to purging, the concurrent purging of two adjacent wells situated in the same aquifer (here, 5 meters apart) was necessary. Although the mean differences in ATP concentrations between the wells were never significantly different from zero, their variability decreased by nearly an order of magnitude during the latter stages of purging (Table 4.1). The results from other microbial parameters evaluated (i.e., BART^{TM} results, microbial community analysis through 16S rRNA gene sequencing) were largely consistent with this interpretation. Accordingly, these multiple lines of evidence suggest that purging times on the order of two hours or more are likely necessary to move beyond contributions from biofilms in the immediate vicinity of the well. These observations are in alignment with the recommended sampling times proposed by Cullimore (2007): at 1 hour, samples taken would be from the outer edge of the biofilm biomass and partially reflect microbial loadings from the groundwater; at 1.5 hours, samples should indicate the outer edge of the subsurface biofilm biomass surrounding the well; and at 2 hours, the sample should be from beyond the biomass but may still be subject to lingering impacts of detached biofilms. Although ATP stabilization occurred during some events to suggest that the outer edge of the biomass can be reached prior to 2 hours, a site-specific—arguably even well-/season-specific—understanding of the subsurface biofilm community’s behaviour may still need to be developed prior to attempting to minimize purging time. This work is the first to document the behaviour of ATP concentrations throughout purging in a shallow, unconfined and highly aerobic aquifer and demonstrate that this biochemical measurement generally aligns with the behaviour expected of other microbial water quality metrics.

It is commonly acknowledged that extended purging time at rates consistent with the goals of low-flow purging will likely yield increasingly representative indications of aquifer microbial water quality void of unwanted well-related biofilm artefacts. Purging times of over several hours per monitoring well are seldom pragmatic due to constraints

such as purge water disposal and the need for most microbiological parameters to be evaluated within 24 hours of sample collection. Indeed, extended purging times of at least 24 to 48 hours have been recommended so that the sample obtained is most representative of the “natural flows of [suspended] microorganisms through the well” (i.e., not unduly influenced by biofilm artefacts from pumping) (Cullimore, 2007); some studies have suggested that even longer periods are necessary (Kwon et al., 2008; Roudnew et al., 2014). From this perspective, the additional effort to track ATP concentrations beyond several hours to minimize purging time for representative aquifer water quality sampling is not likely warranted and beyond the scope of the current work. However, in circumstances where extended purging is not practical or possible, the relative (in)stability of ATP concentrations may still be useful for contextualizing the possible influence of subsurface biofilms on microbial results.

4.5.3 Possible insights related to the spatio-temporal scales of ATP patterns and associated subsurface biofilm behaviour

The heterogeneity of subsurface microbiology has been well-documented at a range of spatial and temporal scales (Young et al., 2008). Due to the complexity of these spatial patterns that result from environmental controls at multiple scales and over time (Ettema and Wardle, 2002), much of the existing work has been qualitative in nature (e.g. Young et al. 2008) and does not fully exploit the information inherent to the acquired data (such as that available from sequential measurements). In fact, spatial patterns of microbial biomass have been shown to be more complex than those of other soil/porous media properties based on fractal dimension (Oline and Grant, 2002). In this work, to facilitate the systematic examination of the patterns emanating from purge water ATP measurements, dynamic time warping was applied. This approach overcomes limitations of irregular sampling time intervals for which parametric time series analysis approaches would not be possible. Moreover, the reliance upon a measure of microbial activity to characterize subsurface biofilm expanse requires consideration of its sensitivity to systematic differences between wells (e.g., steady-state flow rate achieved and inherently higher ATP concentrations in WO78). This taken into consideration, the differences—and remarkably, similarities—exhibited by the normalized ATP concentration sequences (i.e., *relative* microbial activity) throughout purging can be used to gain further insights related to subsurface biofilm dynamics. This may have more specific implications for applications such as remediation of biofouled wells and studies of subsurface microbial ecology in well environments.

The recurring ATP patterns characteristic of each well (i.e., double peaks after an

hour of purging in WO77, a single peak less than an hour of purging in WO78) during consecutive events was supported by the systematic evaluation of their (dis)similarities using the approach proposed in this work. Spatial heterogeneity clearly underpins the distinctive patterns exhibited in each well despite exhibiting practically equivalent physical and chemical water quality characteristics. The proximity of WO77 to two monitoring wells installed within 2 meters may exert considerable influence on the double ATP peaks observed in WO77. We suspect the visually similar ATP patterns (i.e., sampling events 5, 8, and 12) within each well that follow consecutive sampling events during which the recurring characteristic ATP patterns were observed may reflect a maturation of subsurface biofilm in the vicinity of the well screen at relatively stable hydrogeological conditions. We believe that the similar ATP patterns *between* the wells on January 30 coinciding with the first substantial increase in groundwater levels observed over the one year period reflect a hydrogeological perturbation to both wells that overwhelms underlying spatial biofilm heterogeneity normally characteristic of each well. Seasonal fluctuations of the water table have been suggested to promote multi-directional flow (i.e., vertical and horizontal flow), which can result in a greater degree of mixing in the aquifer (Smith et al., 2018). Collectively, the ATP patterns observed at this site supports previously documented hydrogeological influences on microbial water quality changes at field scale manifested as changes to seasonal groundwater elevations (Lin et al., 2012), groundwater flow velocities, and/or hydraulic gradients (van Driezum et al., 2018).

4.6 Conclusions

1. The use of ATP concentrations as a sensitive, field-ready tool for indicating microbial water quality stability during well purging activities was evaluated in a shallow, unconfined, and highly aerobic aquifer. ATP observations were supported by the results of other microbial water quality parameters measured concurrently (i.e., microbial densities through flow cytometry, relative abundance/activity of biofilm-related groups of microorganisms) and reflected time-limited (bio)particle transport processes rather than that of dissolved solutes throughout purging. However, the extrapolation of these results to other environments—such as anaerobic subsurface environments where ATP concentrations and microbial densities may not exhibit the same degree of correlation—must be further investigated.
2. Whereas physical and chemical water quality characteristics appeared to stabilize as anticipated within 10 to 15 minutes (approximately 3 to 5 well volumes), fluc-

tuations in ATP concentrations occasionally persisted beyond 2 hours of purging (>30 well volumes). Assessments of microbial quality of drinking water sources originating in the subsurface must therefore be conducted using approaches that are designed to adequately reflect these differences in temporal scale. For example, this may involve relying on pumping tests of several days duration rather than relying on a limited number of well purge volumes for analysis.

3. Extended purging time at rates consistent with the goals of low-flow purging is commonly acknowledged to yield increasingly representative indications of aquifer microbial water quality. However, purging times of over several hours per monitoring well are seldom pragmatic. In circumstances where extended purging is not practical or possible, the (in)stability of ATP concentrations may be used to contextualize the possible influence of subsurface biofilms on microbial results.
4. Patterns based on the relative microbial activity, as captured using normalized ATP fluctuations throughout purging, are suggestive of subsurface biofilm behaviour. Such patterns can be systematically examined using the non-parametric time series analysis approach applied. The recurrence of characteristic patterns in each well during multiple sampling campaigns appears linked to seasonal hydrogeological conditions (specifically, groundwater level fluctuations). Disruptions of these patterns are speculated to be attributable to sufficiently large hydrogeological perturbations, which may overwhelm ATP patterns normally characteristic of each well. These dynamics warrant further investigation and should be considered in the design and interpretation of subsurface microbial water quality investigations.

Chapter 5

Evaluation of groundwater bacterial community composition to inform waterborne pathogen vulnerability assessments

Alex H.S. Chik
Monica B. Emelko
William B. Anderson
Kaitlyn E. O'Sullivan
Domenico Savio
Andreas H. Farnleitner
Alfred P. Blaschke
Jack F. Schijven

Submitted to *Science of the Total Environment*

5.1 Summary

Microbial water quality evaluations are essential for determining the vulnerability of subsurface drinking water sources to fecal pathogen intrusion. Rather than directly monitor waterborne pathogens using culture- or enumeration-based techniques, the potential of assessing bacterial community using 16S rRNA gene amplicon sequencing to support these evaluations was investigated. A framework for analyzing 16S rRNA gene amplicon sequencing results featuring negative-binomial generalized linear models is demonstrated, and applied to bacterial taxa sequences in purge water samples collected from a shallow, highly aerobic, unconfined aquifer. Bacterial taxa relevant as indicators of fecal source and surface connectivity were examined using this approach. Observed sequences of *Escherichia*, a genus that is suggestive of fecal source, were consistently detected but not confirmed by culture-based methods. On the other hand, episodic appearance of anaerobic taxa sequences in this highly aerobic environment, namely *Clostridia* and *Bacteroides*, warrants further investigation as potential indicators of fecal contamination. *Betaproteobacteria* sequences varied significantly on a seasonal basis, and therefore may be linked to understanding surface-water groundwater interactions at this site. However, sequences that are often encountered in surface water bodies (*Cyanobacteria* and *Flavobacteriia*) were notably absent or present at very low levels, suggesting that microbial transport from surface-derived sources may be rather limited. This work demonstrates the utility of 16S rRNA gene amplicon sequencing for contextualizing and complementing conventional microbial techniques, allowing for hypotheses about source and transport processes to be tested and refined.

5.2 Introduction

Updated terms of reference in the province of Ontario, Canada for treatment requirements of drinking water derived from Groundwater Under the-Direct-Influence of Surface Water (GUDI/GWUDI) sources will be promulgated imminently. They focus on establishing lines of evidence in support of human/mammalian fecal pathogen contamination and/or microbial transport from the surface (Ahmed et al., 2013). Although the updated terms of reference will rely heavily on conventional culture- and enumeration-based microbial methods, the potential for emerging methods to support the establishing additional lines of evidence has been acknowledged. Due to inherent challenges of microbial concentration estimation through groundwater sampling, non-detects are not uncommon despite their actual presence in the source (Chik et al.,

2018). When pathogens or fecal indicators are not observed, additional lines of evidence can be sought to gain a better conceptual understanding of microbial source and transport at a given site.

16S rRNA gene amplicon sequencing offers a broad coverage of microbial water quality, and its costs have been reduced substantially over time. Therefore, it is an attractive tool for complementing conventional microbial methods in assessing the vulnerability of a subsurface drinking water source to fecal pathogen contamination (Griebler and Lueders, 2009; Vierheilig et al., 2015; Savio et al., 2018). The highly conserved regions (e.g., V3-V4 regions) of the 16S rRNA gene are present in every bacterium and archaeon, but are adequately unique to enable their detection and differentiation to the genus level of classification (Yarza et al., 2014). While this technique does not inform whether these sequences originate from viable or culturable microorganisms, it provides a rapid means of characterizing bacterial community composition and has become a staple in many ecological microbiome studies (Vierheilig et al., 2015; Savio et al., 2018).

Microbial communities within subsurface environments suitable for potable water use have been increasingly characterized using 16S rRNA gene sequencing and similar bio-molecular techniques in the past decades (e.g., Griebler and Lueders, 2009; Korbelt et al., 2017; Savio et al., 2019). In these contexts, indicator microbial taxa are often sought to formulate hypotheses about relevant processes underpinning microbial water quality, especially those that may indicate source water vulnerability to fecal pathogen intrusion. However, most microbiome studies to date have relied upon proportions (i.e., relative abundance) and rarefied counts (i.e., re-sampling of sequences reads from samples with adequate library sequencing depth) for the identification of indicator taxa (McMurdie and Holmes, 2014). These approaches have been demonstrated to bear a strong risk of bias towards false positive indications of taxa exhibiting differential abundance (McMurdie and Holmes, 2014). Approaches based on relative taxa sequence abundance fail to account for differences in library sequencing depth and amplification biases; rarefied sequence reads can also lead to a loss of statistical power (McMurdie and Holmes, 2014). While statistically rigorous approaches for comparing taxa sequence read abundance across samples have been proposed (McMurdie and Holmes, 2014) and developed (e.g., Robinson et al., 2010; Love et al., 2014), they have not been applied systematically to identify relevant indicator microbial taxa for supporting pathogen vulnerability assessments. Critically, identification of such indicators requires consideration of variability attributable to the sample collection protocol and other key spatial and temporal gradients spanned by the samples collected.

The groundwater bacterial community was monitored over nine sampling campaigns

spanning a year to demonstrate a progression of steps for identifying relevant bacterial indicators. A focused investigation of two wells installed less than 5 meters apart in the same shallow, unconfined aquifer was deployed. The bacterial core community (i.e., taxa that are abundant and prevalent across the samples collected) was inspected for obvious trends prior to evaluating sample (dis)similarity based on taxa sequences and environmental parameters concurrently measured. In the absence of obvious trends exhibited by the bacterial core community at high taxonomic levels, differential sequence read abundance testing was performed to examine the behaviour of known bacterial indicators and screen for potentially relevant indicator taxa across spatio-temporal factors. This demonstration provides proof-of-concept for the use, interpretations, and limitations of 16S rRNA gene sequencing for supporting vulnerability evaluations of a subsurface drinking water source to fecal pathogen contamination.

5.3 Materials and methods

5.3.1 Site location

The Woodstock groundwater study area is situated near drinking water production wells six kilometers south of Woodstock, Ontario, Canada (Chik et al., 2020b). The site comprises of rolling hills and drumlin features. A series of monitoring wells were installed in a shallow unconfined aquifer comprising of mainly sand and gravels interspersed with discontinuous silty till aquitard units (Haslauer, 2005).

A focused investigation of two wells situated 5 meters apart in a direction perpendicular to regional groundwater flow was conducted. Both monitoring wells (50.8 millimeter [mm] diameter, polyvinyl chloride [PVC] construction) were screened approximately 5.91 m to 16.57 m below ground surface (mBGS) and vented to the atmosphere. Additional well installation details are provided in Koch (2009) and Critchley (2010). The groundwater table was consistently above the top of the screened interval throughout the duration of this study.

The aquifer in which the wells are developed is the shallowest of the regional aquifer system and features extremely high groundwater velocities (aquifer hydraulic conductivity in the well screened interval estimated to be 4.8×10^{-4} to 1.9×10^{-2} m/s (Devlin et al., 2012; Critchley, 2010)) and highly aerobic conditions (mean dissolved oxygen during study period = 7.9 mg/L). The highly aerobic conditions are suggestive of a relatively high degree of groundwater-surface connectivity. Regional groundwater

flow is generally in a southeasterly direction. Annual recharge at this location is estimated to be 396 mm/year (Koch, 2009). Local infiltration travel time through the unsaturated zone to the water table is estimated to be on the order of 2.8 to 5.6 years (Sousa, 2013).

5.3.2 Experimental design

Given that the atmosphere is the primary source of dissolved oxygen in groundwater (Rose and Long, 1988), the extremely aerobic nature of this unconfined aquifer suggests a high degree of surface connectivity. Accordingly, bacterial community composition was expected to exhibit seasonality. However, spatial heterogeneity and the extent of purging performed prior to sample collection are also known to confound bacterial community characterization. Accordingly, three primary factors that contribute to the variability of the 16S rRNA gene amplicon sequences (herein referred to as ‘sequence reads’) were investigated: the season during which sampling was performed (three levels), the well sampled (two levels), and the extent of purging (three levels). Seasonality was represented by three sampling campaigns per season as guided by surficial climatic conditions and groundwater level measurements (Appendix D1). At this site, the respective periods are: summer (“baseline” period characterized by relatively stable water table conditions, typically June - September), winter (transitional period typically exhibiting the lowest but variable annual water table levels due to occasional surface melt events, typically October - January), and spring freshet (period exhibiting highest annual water table levels due to snowmelt recharge, typically February - May) (Pasha, 2018). Two wells (WO77, WO78) within the same hydrogeological formation were sampled to characterize spatial heterogeneity attributable to the individual wells. WO77 was sampled for only one event during the summer and phased in such that both wells were consistently monitored in all sampling campaigns for the following seasons. As groundwater samples are expected to be increasingly representative of the suspended microorganisms in the aquifer pore water without artefacts from well-related biofilms with extended purging (Cullimore, 2007; Chik et al., 2020b), a minimum of two groundwater samples were collected in each well during early (<20 minutes), intermediate- (>20 minutes, <1 hours) and late stages of purging (>1 hours, <4 hours) that loosely correspond with major zones of interrogation projections as described by Cullimore (2007). These projections delineate zones for describing the influence of subsurface biofilms in and surrounding wells on abstracted water quality.

A total of 83 samples were collected. All factors were handled as fixed effects as levels associated with each factor are not random; inferences extrapolated based on

these data would not be very precise. A greater number of years of sampling from additional wells would allow for these factors to be treated as random effects and broader extrapolation of this work. However, this level of characterization was beyond the scope of this proof-of-concept demonstration and does not contribute to supporting the key objectives of this work.

5.3.3 Sample collection

Samples were collected in general accordance with purge-water sampling protocols described by Cullimore (2007). Briefly, wells during each sampling campaign were pumped from quiescence using dedicated pumps. The pumps (Flojet model No. 4105 Series diaphragm pump, Irvine, California, USA) and attached PVC tubing were pre-sterilized in the laboratory using a dilute bleach solution (0.6% sodium hypochlorite) followed by a sterile deionized water rinse. Sterile deionized water was also used on site to prime the pumps as necessary. Groundwater samples were collected in sterile WheatonTM bottles throughout well purging. Physical and select chemical water quality parameters (turbidity, pH, dissolved oxygen, electrical conductivity, and temperature) were monitored using a portable turbidimeter (Hach[®], Colorado, USA) and a portable multiparameter meter (YSI Quattro Professional Plus, YSI Inc./Xylem Inc., Ohio, USA). These parameters generally did not fluctuate substantially after the removal of 3-5 well purge volumes (approximately 10 to 15 minutes of purging) (Chik et al., 2020b). All probes and meters were calibrated using standards as per the manufacturers' instructions. Additional chemical water quality parameters, including one biochemical parameter (i.e., adenosine triphosphate), were also monitored in a related study (Chik et al., 2020b). Summary statistics are presented in Appendix D2.

Two fecal indicators were concurrently enumerated in 100 mL aliquots of the groundwater samples. *Escherichia coli* was enumerated by membrane filtration (Standard Method 9222) in samples using Trypton-Bile-X-Glucuronide (TBX) medium (Oxoid, Hampshire, UK) and incubation at $44\pm 0.5^{\circ}\text{C}$ for 44 ± 4 h, and male-specific (F+) coliphage was enumerated in accordance with US EPA Method 1601 (US EPA, 2001). All fecal indicator samples yielded non-detects. Additional parameters relevant for microbial water quality were also evaluated in selected samples. Microbial cell densities estimated using the FACSCaliburTM flow cytometer (BD Biosciences, New Jersey, USA) yielded results below reliable quantification limits of the method (<1 000 particles/mL)(Chik et al., 2020b). A qualitative culture-based tool, Biological Activity Reaction Tests (ALGE-BARTTM, Droycon Bioconcepts Inc., Regina, Canada), was also deployed during selected sampling campaigns to test for the presence of grass-green

algae, blue-green algae, desmids, diatoms and euglenoids. These tests consistently yielded non-detects.

5.3.4 16S rRNA gene amplicon sequencing

750 mL aliquots of the groundwater samples were filtered through EMD Millipore Sterivex polyethersulfone 0.22 μm syringe filters (Millipore, Massachusetts, USA), stored in dedicated Whirl-PakTM (Nasco[©], Fort Atkinson, Wisconsin, USA) and kept in a -80°C freezer until DNA extraction. A commercial kit (DNeasy PowerWater Sterivex Kit, Qiagen, MO BIO Laboratories, USA) was then used to extract the DNA (e.g., (Fiedler et al., 2018)). 16S rRNA gene fragments were amplified and labelled with a unique, sample-specific multiplex-identifier (“barcode”) in a PCR-based one step bar-coding procedure. Pro341F-Pro805R (5'-CCT ACG GGN BGC ASC AG-3', 5'-GAC TAC NVG GGT ATC TAA TCC-3') primers were used to target 16S rRNA V3-V4 regions (Takahashi et al., 2014). Briefly, PCR was set in triplicate for each sample (25 microlitres (μL)). The reaction mixture contained 2.5 μL of $10\times$ standard Taq buffer, 0.5 μL of 10 millimolar (mM) dNTP, 1.0 μL of bovine serum albumin (20 mg/mL), 5.0 μL of 1 micromolar (μM) forward primer, 5.0 μL of 1 μM reverse primer, 1.0 μL DNA, 0.125 μL of Taq DNA polymerase (5u/ μL) and 9.875 μL of PCR water, DNA was denatured at 95°C for 5 min, followed by 30 cycles of 95°C for 30 sec, 50°C for 30 sec and 72°C for 50 sec and followed by a final extension step at 72°C for 10 min. The triplicate PCR products were pooled. Two μL of PCR amplicons were loaded onto a 2% TAE agarose gel to verify the amount and size of PCR products. PCR products with equivalent amounts of correct amplicons were pooled, gel purified, and quantified using the Qubit dsDNA HS assay kit. Library DNA was sequenced with MiSeq Reagent Kit v2 (2x250 cycles); raw sequencing data have been deposited in the NCBI SRA database (accession number PRJNA625549). A laboratory blank comprising of sterile DI water was performed for every 30 samples as negative controls. The microbiome data analyses described below were first conducted with these quality control samples included to evaluate the extent of the reagent contamination. While these results are not central to the goals of the present investigation, they are integral to the broader interpretation of results emanating from the field purge water samples. Accordingly, results of the statistical analyses performed to identify bacterial taxa that require judicious interpretation are summarized and presented in Appendix D3.

5.3.5 Microbiome data analysis

Taxa have routinely been described as Operational Taxonomic Units (OTUs) generated from the *de-novo* clustering of sequences into bins using a threshold of 97% sequence similarity. Clustering-independent approaches for determining amplicon sequence variants (ASVs, also known as sub-OTUs (Knight et al., 2018)) have been more recently advocated to replace OTUs given their ability to distinguish subtle but real biological sequence variants (Callahan et al., 2017). While ASVs undoubtedly contribute to increased differentiation of taxa at the species and sub-species levels (which may be fundamentally important in cladistics-based studies), appreciable differences in bacterial community composition related to changing source water quality conditions appear to be captured at higher taxonomic classifications (e.g., class/order) across many studies (Lin et al., 2012; Flynn et al., 2013; Ben Maamar et al., 2015; Gülay et al., 2016; Graham et al., 2017; Pogoda, 2017; Lee et al., 2018; Fiedler et al., 2018). Accordingly, a phenetics-based OTU approach was applied.

The UPARSE amplicon analysis algorithm (Edgar, 2013) from the USEARCH8 (32-bit) package (Edgar, 2010) implemented in IMNGS (Lagkouvardos et al., 2016) using default quality settings was used to generate OTUs. A minimum abundance cutoff of 0.25% for each OTU in a sample was set for its inclusion in the final OTU feature table. OTU sequences were aligned and classified as taxa using the SILVA reference database (v. 132, Quast et al., 2013; Glöckner et al., 2017); OTUs classified as chloroplast (2 965 sequence reads, 0.1% of sequence reads), mitochondria (602 sequence reads, 0.02% of sequence reads), or archaea (79 582 sequence reads, 3% of sequence reads) were also filtered prior to data analysis.

After sequence quality filtering, a total of 2 541 261 sequence reads across all 83 purge water samples were assignable to 288 bacterial OTUs. The mean relative abundance of unclassified bacterial sequence reads accounted for 26% at the class level, and increased to 55% at the genus level. Twenty bacterial OTUs were unclassified at the phylum rank (mean relative abundance 9.5%) and were taxonomically filtered as these sequences are likely associated with DNA that were unintended targets of the primers used, or are novel chimeras (Haas et al., 2011). The mean sequencing library depth after quality and taxonomic filtering was 27 891 sequence reads. The sequences affiliated with 183 of 259 bacterial OTUs (for which model fitting was possible) were deemed to be either significantly greater or present exclusively in the environmental samples (Appendix D3).

All analyses were performed within the “R” environment using the packages “phyloseq” (McMurdie and Holmes, 2013) “vegan” (Oksanen et al., 2019) and “DESeq2”

(Love et al., 2014) on the raw sequence read data. Alpha diversity was evaluated using the number of observed species and the *Chao1* and *Shannon* indices (Chao, 1984; Shannon, 1948). Non-metric multidimensional scaling (NMDS) was conducted using Bray-Curtis dissimilarities calculated from log-transformed sequence counts (with a pseudo-count of one for non-detects) to facilitate a two-dimensional representation of the sample relationships. This log-transformation down weights the influence of the most abundant OTUs given that most differences between samples are anticipated to be observed with the transient/rare bacterial community. The pseudo-count adjustment was used only for the visualization of samples rather than for statistical testing. Permutational Multivariate Analysis Of Variance (PERMANOVA) using the ordinated distance matrices against key factors was performed using the *adonis* function in the package “vegan” (Oksanen et al., 2019). Constrained Analysis of Principal Coordinates (CAP) was also performed to explore relationships between sample 16S rRNA bacterial composition and physical and chemical parameters measured concurrently.

Key taxa contributing to the observed sample relationships were identified by statistical testing of the original raw sequence counts. Rather than omitting samples due to low library sequence depth, the negative binomial generalized linear model (NB GLM) framework as implemented through DESeq2 (Love et al., 2014) was used to explore taxa differential abundance. A negative binomial model allows for overdispersion in the sequence count distribution. Briefly, NB GLMs were applied to sequence counts n_{ij} for each taxon i in sample j with fitted mean μ_{ij} and a taxon-specific dispersion parameter α_i . The fitted mean is a function of a sample-specific size factor s_j and a parameter q_{ij} proportional to the expected true concentration of that taxon for sample j . The coefficients β_i provide the \log_2 -fold-difference estimates for each taxon. \log_2 -fold-difference estimates were re-expressed as \log_{10} -fold-differences for more intuitive order-of-magnitude interpretations. The default setting for shrinkage estimates was utilized; this allows for imprecise \log_{10} -fold-difference estimates associated with OTUs comprising of low counts to be shrunk and facilitate a better comparison of these estimates across factors (Love et al., 2014). Saturated models comprising of two- and three-way interactions between all factors were first investigated. Overall model significance was evaluated by determining the difference in likelihood values between a fitted model and a reduced model. All models were reduced to simple additive GLMs of the three main factors after a lack of significant ($p > 0.05$) interactions was observed for the majority of taxa. The significance of differential abundance estimated for a taxon between factor levels was tested using a Wald chi-square test (Love et al., 2014). All p -values were adjusted for multiple comparisons controlling for the false discovery rate (Benjamini and Hochberg, 1995).

5.4 Results

5.4.1 The core microbial community: dominant taxonomic groups observed in groundwater samples

Two major discontinuities in the empirical taxa abundance distribution comprising all groundwater samples were observed (Appendix D4) and are used to distinguish the bacterial core (OTUs in greater than 91% of samples), transient (OTUs between 56% and 91% of samples), and rare bacterial community (OTUs in less than 56% of samples). The discontinuities are marked by OTU sequence reads affiliated with the genera *Novosphingobium* (present in 76 of 83 samples, Figure 5.1 dashed-line) and *Brevibacillus* (present in 46 of 83 samples, Figure 5.1 dash-dotted line). Additionally, 107 bacterial OTUs were consistently observed in all groundwater samples collected; their reads collectively comprised 78% of reads within each sample, on average. Sequence reads affiliated with OTUs from the phyla *Proteobacteria* (372 762 reads, 27 OTUs), *Omnitrophia* (646 620 reads, 56 OTUs), and *Patescibacteria* (167 169 reads, 8 OTUs) dominated the core bacterial community. *Rokubacteria* (115 352 sequence reads, 9 OTUs), *Nitrospirae* (38 080 sequence reads, 4 OTUs), *Cyanobacteria* (48 081 sequence reads, 3 OTUs) can also be considered part of this environment's core community. Notably, sequence reads belonging to an OTU of the *Escherichia* genus was the most abundant of the *Proteobacteria* (93 794 reads) and ubiquitous across all 83 purge water samples. *Escherichia* sequence reads were also present in the negative controls, albeit at significantly lower levels than in the purge water samples ($p < 0.001$). *Bacteroidetes*, *Firmicutes* and *Proteobacteria* were the only phyla that had OTU sequences represented as part of the core, transient and rare bacterial communities.

A relative abundance plot of bacterial taxa on phylum level across all purge water samples was generated (Figure 5.2); additional relative abundance plots comprising of only core community taxa, and a focused examination of *Proteobacteria* composition, are presented in Appendix D4. These plots complement the empirical taxa distribution plot and highlight the stability of the core bacterial community and the dominance of members affiliated with the *Proteobacteria*, *Omnitrophia*, and *Patescibacteria* phyla. Sequence reads affiliated with members of the *Firmicutes* phylum are notably excluded from the core community, but were generally present at higher relative abundances during the intermediate stages of purging. Sequence reads of *Proteobacteria* OTUs appear more prominent in the summer than the following seasons. Over the same seasons, the relative proportions of *Betaproteobacteria* to other *Proteobacteria* classes changed substantially.

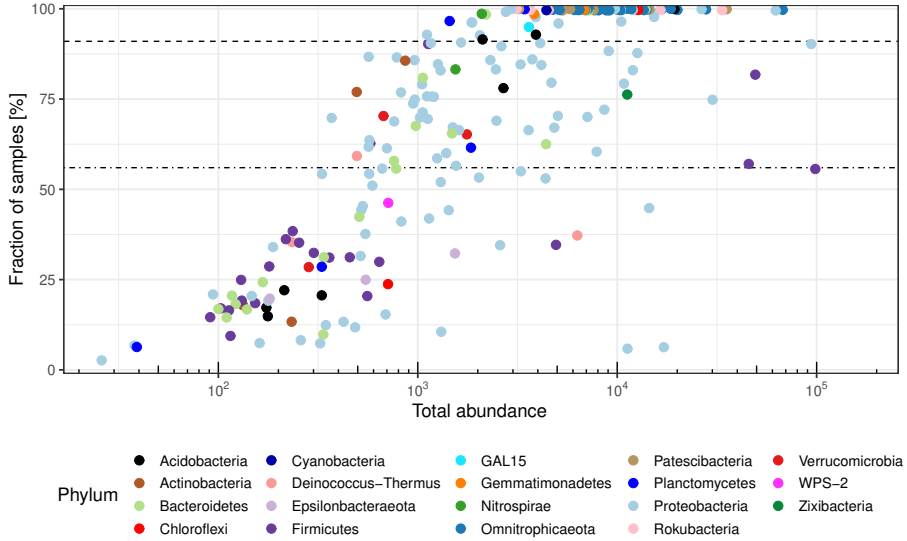


Figure 5.1: Total abundance of OTU sequence reads across all samples plotted against their prevalence (fraction of samples across which an OTU was detected). Dashed (91%) and dot-dashed (56%) lines denote the prevalence thresholds determined based on the empirical taxa abundance distribution of all samples (Appendix D4).

5.4.2 Seasonal/event-based influences and spatial heterogeneity as key drivers of bacterial community diversity

All alpha diversity metrics calculated (observed diversity richness [OTUs], estimated OTU richness [Chao1], and Shannon diversity, Appendix D5) were consistently higher in well WO78 than WO77 as supported by a Wilcoxon rank sum test ($p < 0.001$). Alpha diversity was not linked to the stage of purging ($p > 0.20$). Seasonality appeared to be influential only for some alpha diversity metrics (i.e., number of OTUs and Chao1 indices, $p < 0.10$), which indicated that more diversity was exhibited in samples collected during the summer period.

Bray-Curtis distances were calculated to characterize the beta-diversity between samples; NMDS performed based on these distances (Figure 5.3) revealed a clear separation of samples between the well (PERMANOVA, $R^2 = 0.189$, $p < 0.001$) and the seasons (PERMANOVA, $R^2 = 0.142$, $p < 0.001$) from which the samples were collected. The purging stage appeared least influential of the factors investigated but remained a significant factor within the ordination (PERMANOVA, $R^2 = 0.054$, $p < 0.001$). When PERMANOVA was repeated using individual sampling campaigns as levels in place of

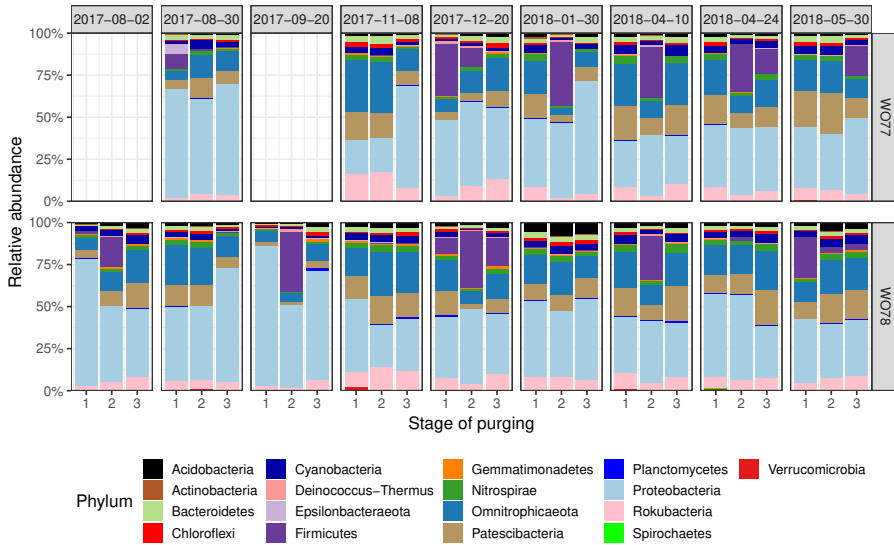


Figure 5.2: Summary of relative abundance of bacterial taxa on phylum level, by well membership (WO77 vs. WO78) organized by sampling occasion.

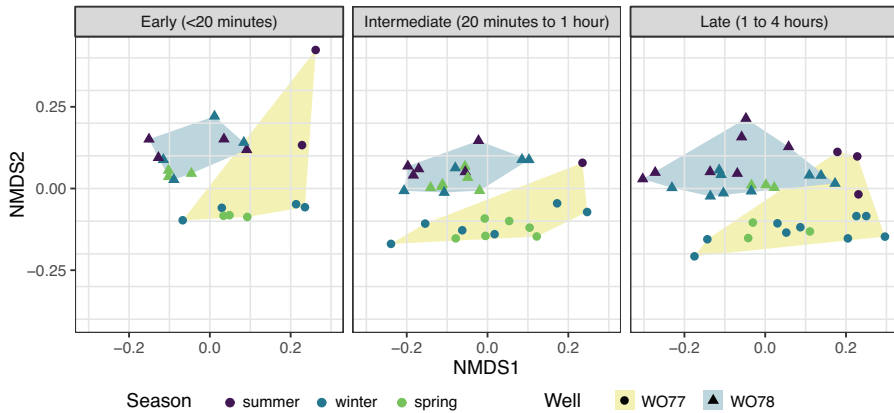


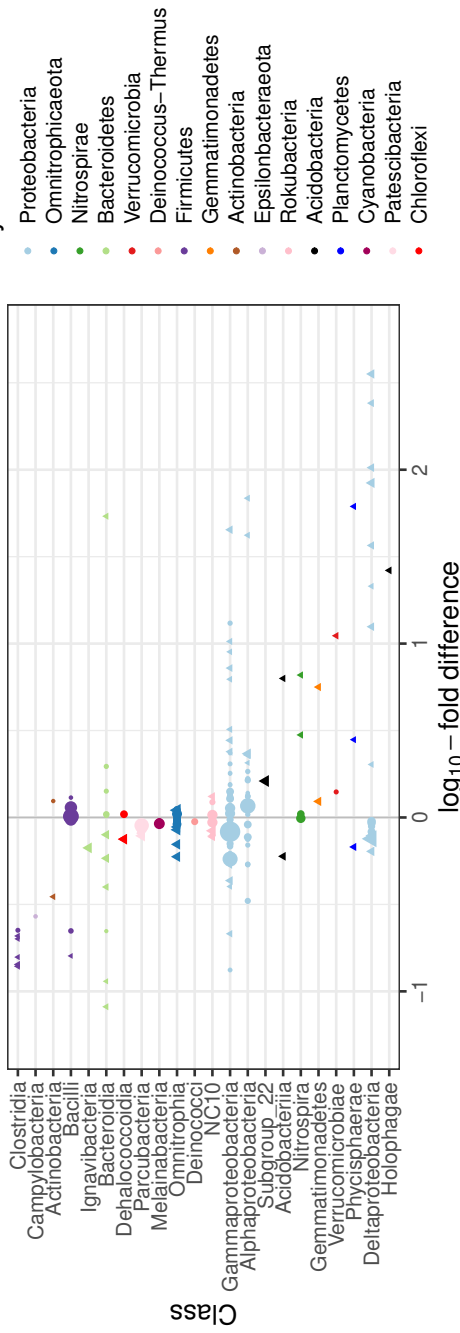
Figure 5.3: Non-metric multidimensional scaling (NMSD) ordination of purge water samples using Bray-Curtis dissimilarities (Stress=0.138). Three facets were used to portray the purging-stage achieved at the time of sample collection. Circles and triangles represent the wells from which the samples were collected (WO77 and WO78, respectively); their colours are indicative of the season during which the samples were collected.

the seasonal factor, substantially more of the variation was explained by sampling campaigns ($R^2 = 0.395$, $p < 0.001$) than the well membership ($R^2 = 0.160$, $p < 0.001$) and purging stage ($R^2 = 0.044$, $p < 0.001$). The implications from this analysis are two-fold. First, spatial heterogeneity is demonstrated to exert considerable influence on bacterial community composition in the samples, even when collected from wells spaced merely 5 m apart in the same aquifer. Second, sampling campaign-based changes explained more of the variability in bacterial community composition than the seasonal factor (i.e., summer, winter, spring as specified) while accounting for spatial heterogeneity and the purging extent. This observation suggests that 16S rRNA gene amplicon sequencing is sufficiently sensitive—and therefore suitable for—denoting episodic microbial water quality changes occurring at event-based scales. However, a full site characterization (including the re-definition of seasonality at this site) is beyond the scope of this proof-of-concept demonstration. Consequently, subsequent analyses of individual OTUs in this study were focused on the factors originally specified. CAP yielded similar inferences as the NMDS ordination, although the variability in bacterial community composition attributable to seasonal/sampling campaign influences was made more prominent (Appendix D6). This result was to be expected as the measured environmental variables (e.g., groundwater electrical conductivity, temperature, pH) differed more between sampling campaigns than between wells or throughout the purging process. As expected, seasonally-influenced parameters of temperature, pH and electrical conductivity were weakly aligned with the first two ordination axes (Appendix D6).

5.4.3 Identification of relevant OTUs through differential abundance testing

Useful indicator taxa were identified by ranking the significance of factor coefficients within the three-factor NB GLM fitted for each OTU (Appendix D7). OTUs comprising of more than an average of 5 sequence reads per sample are represented in Figure 5.4 and summarized by class, in order of increasing \log_{10} -fold differences in comparison to the base level of each factor. Rather than blanket removal of all OTUs detected in negative control samples, the same statistical framework was used to flag OTUs in the purge water samples that were also detected in the negative controls to account for the possibility of background/reagent contamination.

WO77 vs. WO78



(a)

Figure 5.4: Differential abundance of various OTUs by class, in order of increasing log₁₀-fold differences in comparison to the base level of each factor across (a) the well factor, (b) the purging stage factor (early vs. intermediate vs. late stages), and (c) the seasonal factor (baseline summer vs. transitional winter vs. spring freshet periods). The base level is underlined in the heading above each plot. Thus, a positive log₁₀ fold difference indicates elevated OTU sequences relative to the base level. Circles represent OTUs that did not exhibit statistically significant different factor coefficients ($p < 0.05$), triangles represent those that are significant. The size of the symbols is proportional to the normalized mean number of OTU sequences across all samples.

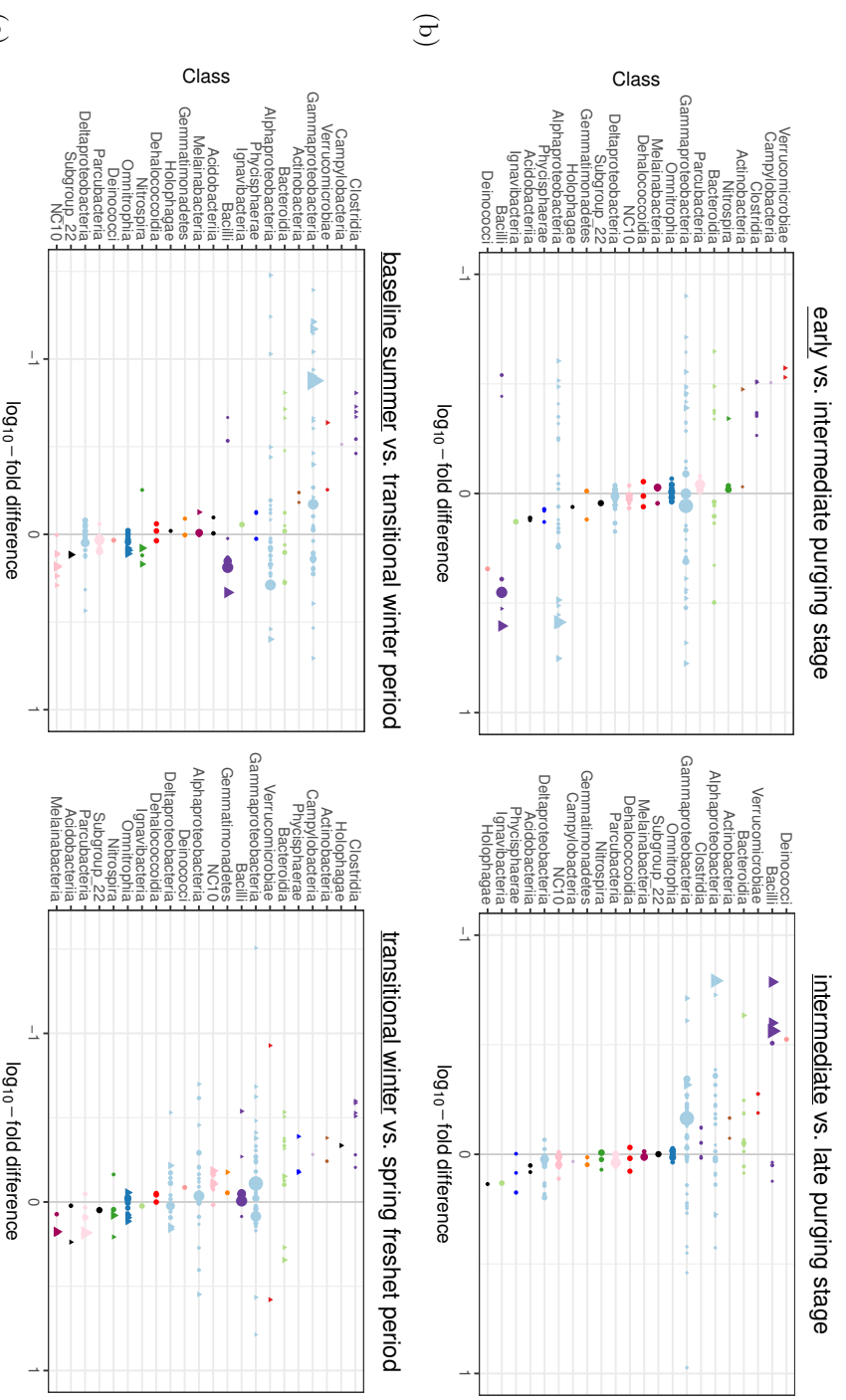


Figure 5.4: (continued)

Despite a generally higher alpha-diversity in WO78 than WO77, the sequences of the vast majority of OTUs comprising the core bacterial community were not significantly different between the two wells. The higher alpha-diversity is suspected to be linked with a higher abundance of bacteria present in WO78; ATP measurements in the same wells indicated higher levels of microbial activity in WO78 that is suggestive of higher microbial densities (Chik et al., 2020b). Many OTUs of the same bacterial class responded to each factor investigated similarly. For the purging stage factor coefficients, several OTUs affiliated with the *Bacilli* and *Alphaproteobacteria* classes were significantly higher during the intermediate purging stage. No OTUs between the intermediate and the late stages of purging exhibited a significant increase in sequence read abundance. The shifts of *Proteobacteria* composition between seasons noted by means of relative abundance was confirmed; sequence reads of OTUs from the bacterial class *Betaproteobacteria* were most significantly different between seasons. Sequence reads affiliated with classes of *Bacilli*, *Clostridia*, and *Bacteroidetes* OTUs were also noted to be significantly different between seasons.

5.5 Discussion

5.5.1 Bacterial diversity and core community analysis as initial indicators for surface-water groundwater interactions

A handful of studies deploying NGS methods to document microbial communities in subsurface environments suitable for potable water supplies have emerged over the past decade (e.g. Vierheilig et al., 2015; Savio et al., 2018). Although there is no evidence that a community endemic to groundwater sources exists or has been described (Griebler and Lueders, 2009), differences in groundwater and surface water bacterial communities may be used to evaluate the vulnerability of subsurface water sources to fecal pathogen intrusion (e.g. Lin et al., 2012; Ben Maamar et al., 2015; Braun et al., 2016; Graham et al., 2017; Pogoda, 2017; Fiedler et al., 2018). Savio *et al.* (2019) observed dynamic response of *Flavobacteriia* proportions in karst aquifer water bacterial core community during recharge events. Where the hydrogeological setting is susceptible to rapid changes to microbial water quality (e.g., due to flashy responses attributable to preferential flow paths in mature karst aquifers with high secondary porosity), major bacterial core community shifts are likely indicative of heightened vulnerability to surface-derived microbial transport. In contrast, the hydrogeological setting investigated herein is an unconsolidated sand-and-gravel aquifer where changes

to groundwater quality can be expected to be gradual in comparison. In this study, relative abundance and prevalence-based approaches at higher taxonomic levels did not reveal substantial shifts to the core bacterial community observed. Accordingly, a bacterial core community analysis of 16S rRNA gene amplicon sequencing data may serve only as an initial litmus test for obvious signs of vulnerability to surface-derived fecal pathogen intrusion.

Failure to observe substantial shifts to the bacterial core community should not be directly interpreted as a lack of surface water influence or source water quality changes. Both alpha- and beta-diversity metrics indicated that spatial heterogeneity and aspects of seasonality at this site were influential to overall bacterial community composition. In taking spatial heterogeneity into consideration, bacterial community composition was revealed to be sensitive to episodic microbial water quality changes occurring at event-based scales. However, their influences on relevant microbial indicator taxa—especially those comprising the transient/rare community—must be further parsed out considering the inherent limitations of 16S rRNA gene amplicon sequencing (e.g., artefacts arising due to unequal sample sequencing depths). This requires dedicated approaches for differential abundance analysis of sequence read count data. The ensuing discussion focuses on how site-specific source and transport hypotheses can be developed based on the taxa identified from the progression of core community analysis to differential abundance testing. These hypotheses can be refined and further tested to inform groundwater vulnerability assessments to fecal pathogen intrusion.

5.5.2 *Escherichia* sequence reads from 16S rRNA gene sequencing require judicious interpretation

Fecal contamination of human/mammalian origin remains the most relevant source of human waterborne diseases globally (Santo Domingo and Ashbolt, 2012). Many members of the genus *Escherichia* are universal gut biota found in humans and warm blooded animals (Madigan et al., 2017); *Escherichia coli* remains the *de facto* universal indicator of fecal contamination (Santo Domingo and Ashbolt, 2012; Odonkor and Ampofo, 2013). While the negative controls did yield some *Escherichia* sequence reads to indicate possible contamination of the DNA extraction kits used (e.g., Corless et al., 2000; Salter et al., 2014; Pollock et al., 2018), the levels observed in the controls were significantly lower than those typically found in the groundwater samples. Their presence and abundance across all groundwater samples (mean 1 165.7 sequence reads per sample) considering the small *effective* volume amplified and sequenced indicate the probable presence of a persistent fecal source likely quantifiable by culture-based

methods (only 3.0 μL of undiluted DNA extract from the 100 μL elution volume representing the original 750 mL filtered sample volume was amplified, and only a portion equivalent to 4 nanomolar (nM) of the pure amplicons was sequenced). However, none of the conventional culture-based enumeration procedures (*Escherichia coli* and male-specific coliphage) performed using 100 mL aliquots yielded detections of the target fecal indicators. This discrepancy underscores the challenges of corroborating results emanating from culture-based and molecular-based methods.

Many factors can ultimately confound the use of *Escherichia* sequences detected using 16S rRNA gene sequencing as fecal indicators. This may include the inherent limitations of measurement error, physiological state of bacteria from which the recovered sequences originate, and the presence of background environmental DNA sequence fragments from deceased microorganisms. Moreover, *Escherichia coli* may be more prevalent in the natural environment than historically anticipated, including those that may not originate from mammalian fecal sources (Scott et al., 2002; van der Wielen and Medema, 2010; Frick et al., 2018). Therefore, extreme caution should be exercised in using these gene sequences directly as replacements of culture-based fecal indicators without enhanced characterization of these sequences. This type of characterization is possible through the design and use of targeted primers and protocols for *Escherichia coli* as in Sabat et al. (2000) and Reischer et al. (2008) or more extensive source monitoring programs designed to track known sources (e.g., septic tanks, agricultural manure, poikilothermic organisms). The use of targeted fecal gene sequences for concentration estimation and comparisons with culture-based results would require evaluation against a standard PCR regression curve (e.g., as applied in quantitative PCR); the method sensitivity limit (MSL, Chik et al., 2018) based on the minimal detectable number of fecal gene sequences per reaction volume (Reischer et al., 2006, 2008) can further facilitate comparisons of method sensitivity using different instruments, protocols, and water matrices.

5.5.3 Anaerobic taxa sequences as potential vulnerability indicators within an aerobic environment

Sequences of *Clostridia* OTUs in this subsurface environment were generally rare (with several OTUs detected in the negative controls). However, the occurrence of sequence reads of these obligate anaerobes coincided with other anaerobic taxa such as *Bacteroides* and *Prevotella* (phylum *Bacteroidetes*) as well as a notable decrease in several aerobic *Bacillales* OTUs during some sampling campaigns. These anaerobic taxa dominate human and mammalian gut microbiota (Embley and Stackebrandt,

1997; Turnbaugh et al., 2009), but are not exclusive to these sources (van der Wielen and Medema, 2010). In similar aerobic freshwater environments, they are expected to be orders of magnitude lower than aerobic bacteria (e.g., Balkwill and Ghiorse, 1985). Coupled with the fact that *Clostridia* are subject to documented biases attributable to DNA/RNA extraction from gram-positive bacteria (which have structures that are difficult to disrupt, as typical of bacteria within the *Firmicutes* phylum (Embley and Stackebrandt, 1997; von Wintzingerode et al., 1997; Frostegård et al., 1999)), their detection from the overwhelmingly aerobic conditions characteristic of this aquifer (mean dissolved oxygen = 7.9 mg/L, historical mean dissolved oxygen = 8.7 mg/L (Critchley, 2010)) was made more remarkable. Accordingly, the divergent behaviour of anaerobic and aerobic genera may be suggestive of episodic microbial source water quality changes and warrants additional investigation as a site-specific vulnerability indicator of fecal contamination.

5.5.4 *Betaproteobacteria* sequences as potential indicators of seasonal subsurface processes

Consistent with the majority of past studies, sequences belonging to the phylum *Proteobacteria* were most dominant in this study (Lin et al., 2012; Flynn et al., 2013; Ben Maamar et al., 2015; Gülay et al., 2016; Graham et al., 2017; Pogoda, 2017; Lee et al., 2018; Fiedler et al., 2018). The conditions of the present study are perhaps most closely aligned with bacterial groups observed from nitrate-impacted groundwaters in Brittany, France (Ben Maamar et al., 2015). The nitrate levels were much higher in the groundwater samples (45-57 mg/L) than in the present study (14.0 mg/L) and sequences from the *Betaproteobacteria* class were identified to be the most abundant. Ben Maamar et al. (2015) observed that owing to their prevalence across samples and their distinct relative abundance profiles, *Betaproteobacteria* were useful to distinguish “recent” (< 25 years) from “older” (> 40 years) groundwaters. OTUs belonging to the *Comamonadaceae*, *Burkholderiaceae*, and *Oxalobacteraceae* bacterial families were observed to dominate over other *Betaproteobacteria* families in “recent” groundwaters compared to “older” groundwaters. In another study of a riverbank filtration site, the increase of *Burkholderiaceae* sequences in wells relative to other *Proteobacteria* has also been suggested to be indicative of allochthonous microbial loads (Fiedler et al., 2018).

In the present study, OTUs from the *Betaproteobacteria* class (classified in SILVA v. 132 as the order *Betaproteobacteriales* within the class of *Gammaproteobacteria*) were identified to exhibit the most significant seasonal responses. Their levels (both in terms of relative and differential abundance) were elevated during the summer when

elevated groundwater electrical conductivity measurements may otherwise suggest an “older” average groundwater age. While this observation apparently contradicts the inferences drawn by Ben Maamar *et al.* for their use as an indicator of groundwater age, we speculate that both observations may be collectively explained by the elevated nitrate levels associated with the “recent” groundwater in Ben Maamar’s study and the historically elevated nitrate concentrations during the growing season at this agricultural site (Haslauer, 2005; Koch, 2009). Closer examination of relative abundance at lower levels of bacterial taxonomy (Appendix D8) revealed the dominance of *Pseudomonadaceae* and *Burkholderia* OTUs. Both families are known to contain denitrifying organisms (and generally regarded as copiotrophs) that could catalyze the full series of denitrification reactions or specialize in the reduction of nitrite produced by bacteria affiliated with the *Enterobacteriales* order (Lycus *et al.*, 2017; Griebmeier and Gescher, 2018). Additional microbial and chemical water quality analyses focused on *Betaproteobacteria* taxa and nitrogenous chemical species in wells within the vicinity may be warranted to investigate their use as indicators of surface water-groundwater interaction (as they may be linked to agricultural activities occurring at this site). Bacterial biomass turnover times through analysis of leucine incorporation as in van Driezum *et al.* (2018) may also be used to discern whether the seasonal response exhibited by *Betaproteobacteria* is a result of autochthonous (i.e., growth of existing bacteria in response to nutrient availability in the groundwater) or allochthonous processes (i.e., introduction of microorganisms from external water source[s]).

5.5.5 *Bacillales* and *Sphingomonadales* sequences as potential indicators of adequate well purging

Subsurface microbial biofilms in and surrounding the well screen are known to contribute microbial water quality artefacts in abstracted well water, especially upon initiation of well purging activities (Cullimore, 2007; Chik *et al.*, 2020b). Consequently, extensive purging is often necessary to obtain samples that are representative of suspended microorganisms in the aquifer pore water void of these artefacts (Harter, 2003). In this regard, the elevated levels of sequence reads affiliated with *Bacillales* and *Sphingomonadales* OTUs during the intermediate purging stage appear to indicate the relative position of the “focal biomass” (Cullimore, 2007) in the subsurface. Both genera of bacteria are ubiquitous in soil and groundwater environments (Brooks *et al.*, 2015). Some of the *Bacillales* (*Paenibacillus*, *Brevibacillus*) and *Sphingomonadales* (*Sphingomonas*, *Novosphingobium*) sequences belong to genera known to produce extracellular polymeric substances (EPS) and are key taxa found in biofilm reactors

mimicking aerobic groundwater conditions (Ross et al., 2001; Rickard et al., 2002). However, as *Bacillales*-affiliated OTUs may be subject to aforementioned biases attributable to DNA/RNA extraction from gram-positive bacteria, sequence reads of *Sphingomonadales* OTUs may be more useful to indicate biofilm artefacts related to purging activities. *Sphingomonas* spp. in particular have been identified to be a “highly co-aggregating genera” that are quantitatively important members of freshwater biofilms due to their ability to provide other community members with a colonization advantage through adhesion mechanisms (Rickard et al., 2002). The significant reduction of sequence reads from both *Bacillales* and *Sphingomonas* OTUs—and the lack of significant increase of all other OTUs during the last purging stage—provides a useful indication of adequate well purging for aquifer-representative microbial water quality samples.

5.5.6 Scarce *Cyanobacteria* and *Flavobacteriia* sequences as potential lines of evidence of limited surface water-groundwater bacterial transport processes

Recently, pigment-bearing microorganisms including those commonly found in the phylum *Cyanobacteria* have been suggested as indicators of recent surface-derived microbial water quality influence (Ahmed et al., 2013). Many of the bacteria within this phylum are obligate photoautotrophs which have been deemed unlikely to persist in the subsurface void of light sources (Minda et al., 2008). Therefore, their increased levels in the subsurface is likely indicative of surface-derived microbial transport. *Flavobacteriia* abundance in the bacterial community has also been suggested to indicate microbial water quality changes associated with surface discharge events in riverbank filtration and karst environments respectively (Fiedler et al., 2018; Savio et al., 2019).

In this study, all *Cyanobacteria* sequences were affiliated with three OTUs belonging to the class *Melainabacteria*. However, this bacterial class has been more recently proposed as a candidate phylum sibling to *Cyanobacteria* as its members can fix nitrogen and does not perform photosynthesis (Di Rienzi et al., 2013). The lack of other *Cyanobacteria* taxa and scarce levels of *Flavobacteriia*-affiliated sequences across all samples—along with consistent non-detects across all culture-based ALGE-BARTs conducted—indicate that the shallow unconfined aquifer in the present study may not be highly susceptible to local/recent surface-derived microbial water quality impacts through transport. This hypothesis is further supported by hydrogeological evidence that average travel times through the unsaturated zone to the water table at this location has been estimated to be on the order of 2.8 to 5.6 years (Sousa, 2013). The

confirmation of the abundance and prevalence in higher soil horizons or nearby surface water features (e.g., ephemeral streams and agricultural ditches) may be a promising line of evidence indicating limited surface-derived microbial transport processes.

5.6 Conclusions

In this work, the utility of bacterial community composition to inform vulnerability assessments of subsurface water sources to fecal pathogen intrusion was investigated. An initial evaluation of the core bacterial community did not yield any obvious indication of a high level of vulnerability. Spatio-temporal factors influencing bacterial community composition and identification of useful indicator taxa was subsequently demonstrated. Spatial heterogeneity exerted appreciable influence on bacterial community composition; its consideration was paramount to parsing out changes attributable to event-specific/seasonal factors. The consideration of these factors are necessary to meaningfully characterize microbial water quality at the aquifer scale. Accordingly, examination of bacterial indicator taxa relevant as indicators of fecal source and surface connectivity was performed using NB GLMs considering these key spatio-temporal factors.

16S rRNA gene sequences linked to fecal contamination (i.e., *Escherichia* genus) were consistently detected in all samples. However, their use in lieu of culture-based fecal indicators is not recommended due to inherent differences between culture-based and culture-independent methods. On the other hand, the divergent behaviour of anaerobic taxa (specifically *Clostridia* and *Bacteroides*) and aerobic taxa (*Bacilli*) sequences during some events was observed. Their episodic appearance in highly aerobic conditions at this site warrants further investigation as possible indicators of fecal contamination. The sequence reads belonging to the order *Burkholderiales* (class: *Betaproteobacteria*) exhibited dynamics that likely reflect seasonal water quality changes and nutrient dynamics at this site. How these changes are linked to surface-water groundwater interactions can further be elucidated through focused evaluation of this taxon, the use of biochemical techniques such as leucine incorporation, and chemical water quality analyses. Additionally, the abundance of *Bacillales* and *Sphingomonadales* sequences may be useful for informing the adequacy of well purging to obtain groundwater samples representative of suspended microorganisms in the aquifer pore water. Finally, the few *Cyanobacteria* and *Flavobacteriia* gene sequences observed—both of which are commonly reported in microbial community studies of surface water sources—provide a promising line of evidence that microbial transport from surface-

derived sources may be rather limited at this site. Additional targeted chemical and microbial sampling informed by this work can be used to refine site-specific hypotheses related to microbial source and transport, thereby informing vulnerability assessments of these subsurface water sources to fecal pathogen intrusion.

Chapter 6

Synthesis of findings & research outlook

6.1 Synthesis

The overall objective of this dissertation was to investigate opportunities to **improve microbial water quality monitoring as it pertains to evaluating vulnerability of a subsurface drinking water source to pathogens**. Whereas surface waters are automatically deemed to be at higher risk of microbial contamination from human and mammalian fecal sources (e.g., US EPA’s Surface Water Treatment Rules), subsurface water sources are typically considered to be less susceptible to contamination. Accordingly, pathogen vulnerability assessments of subsurface drinking water sources have often been conducted to ascertain the level of risk associated with such sources. The onus is on the water purveyor to demonstrate that adequate treatment is provided for the range of source water qualities observed and reasonably anticipated.

Ironically, microbiological water quality of pristine water sources is amongst the most difficult to demonstrate. Source waters of high microbiological water quality for potable use necessarily implies that pathogen concentrations in the source water are low. However, absence of evidence is not evidence of absence (Chapter 2). Imperfect methodologies used in the estimation of pathogen/fecal indicator concentrations, coupled with our inability to monitor all pathogens/indicators comprehensively present in the source, have led to the need for multiple lines of evidence to determine treatment needs. In some cases, the effort involved in demonstrating good microbial water quality can outweigh that associated with providing treatment based on a conservative estimate of source water quality. However, the pervasive application of the precautionary principle throughout risk assessment can inadvertently lead to the adoption of risk management actions that are disproportionate to the actual threat posed.

This was a key consideration behind the Ontario Ministry of Environment’s efforts in updating the soon-to-be superseded *Terms of Reference: Hydrogeological Study to Examine Groundwater Sources Potentially Under Direct Influence of Surface Water*. Many “GUDI” / “GWUDI” studies and assessments of public drinking water supplies across the majority of North American jurisdictions have historically relied upon mandatory evaluations of whether specific hydrogeological environments are more vulnerable to waterborne pathogen contamination. However, **(microbial) water quality, both current and anticipated, underpins drinking water treatment design and operational decisions**. Until uncertainties associated with models of microbial source and transport through the subsurface can be substantially reduced, this is the core principle enshrined in the updated Terms of Reference. This dissertation therefore serves to inform a prudent level of microbial water quality monitoring, and the potential (and limitations) of emerging microbial tools to support pathogen vulner-

ability assessments. The contributions of this work can be contextualized as answers to the progression of research questions previously formulated in Chapter 1:

- RQ1. How have microbial non-detects been reported, interpreted, and analyzed, and do some approaches lead to bias?
- RQ2. What level of precision in the estimation of mean microbial concentrations is attainable and which aspects of microbial monitoring program design are most influential to improving or compromising precision?
- RQ3. What additional considerations of the subsurface environment are necessary to indicate that representative microbial water quality sampling of aquifer water quality is achieved, and what tools are available to support these considerations?
- RQ4. How can bacterial community analysis using emerging biomolecular tools such as 16S rRNA gene amplicon sequencing be used to inform vulnerability assessments?

6.2 Key contributions

6.2.1 Handling microbial non-detects without incurring bias

Non-detects of microbial pathogens and fecal indicators alike are an important line of evidence that would be suggestive of a high quality water source. However, existing conventions to report and analyze them as a measured concentration less than one microorganism within the analyzed volume can lead to a conservative bias (i.e., overestimation) when used to estimate microbial concentrations. This reported value is widely regarded as the method detection limit. However, as revealed through a critical review of the state-of-the-scientific-practice, this convention is misleading. The method detection limit is not intended to be used for, and therefore has been misapplied in, detection and enumeration-based methods that count discrete microorganisms. Whereas the method detection limit facilitates comparisons of method sensitivity to the target analyte in chemistry, non-detects in microbiology are not strictly a reflection of the lack of method sensitivity (i.e., the probability of detection when the target microorganisms are actually present) but also the limitations associated with successfully capturing and enumerating discrete organisms within a finite sample volume to estimate concentration. Both sample volume and analytical recovery can and were demonstrated to influence the probability of observing a non-detect. Furthermore,

uncertainty in concentration estimates precludes an MDL-based interpretation of results; a non-detect can arise from a non-zero source concentration just as a non-zero observation can arise from a source concentration that is known to be less than the purported method detection limit. The concepts described herein also apply to increasingly common biochemical molecular methods (e.g., qPCR, 16S rRNA gene amplicon sequencing). However, additional assumptions and complexities would be necessary in the inference of concentration estimates using these approaches.

The reporting of microbial non-detects as censored values relative to the purported MDL is therefore misleading, and has led to the misuse of censored data statistical approaches. Proper statistical analysis approaches of raw microbial data (i.e., raw count and analyzed sample volume) that account for measurement error are critically underscored. Censored data approaches should be reserved for data correctly interpreted as being censored, such as TNTC plate counts where the actual count is known only to exceed a specified threshold. Bias attributable to interpretation and analysis of microbial non-detects as censored concentrations is regarded as conservative and “precautionary” because risk is overstated. However, this safety factor progressively fails to be incorporated at higher pathogen concentrations that might pose a health risk. Accurate risk assessments with known, quantifiable safety factors are preferable over relying on “rounding errors” to provide peace of mind in delivering microbiologically safe drinking water.

6.2.2 Source water monitoring program design to ensure adequate treatment of protozoan pathogens

Quantitative microbial risk assessment (QMRA) has been increasingly applied as a tool to set health-based targets for quantifying the disease burden attributable to waterborne pathogens. Commonly-used health-based targets used in drinking water include an acceptable infection risk level of 1 person per 10 000 consumers per year (10^{-4}) (The Netherlands, US EPA), or 10^{-6} disability-adjusted life years (DALYs) per person per year (World Health Organization, Health Canada, Australia) for each pathogen. When such a target is set as an acceptable level of mean annual risk, a reverse QMRA can be performed to determine the minimum treatment required to attain the target given the mean pathogen concentration in the source water. Therefore, accurate mean pathogen concentrations are critical for prescribing treatment that would be adequately, but not excessively, protective of public health. Excessive treatment is not always desirable as this could lead to other implications such as the formation of

hazardous by-products when using chlorine disinfection, or costly infrastructure for the treatment of specific pathogens such as *Cryptosporidium* and *Giardia* (oo)cysts.

The level of precision attainable as a function of sample sizes (for native protozoan enumeration and for quantification of analytical recovery) has not been demonstrated or quantified in the peer-reviewed literature. Accordingly, this work provides an approach that explicitly links minimum pathogen sample size requirements to the attainable level of precision. This is particularly useful for systems that serve small populations or are far from laboratories offering these analyses where minimization of sampling effort without substantially compromising precision is a key consideration for monitoring program design. Given the relevance of the two aforementioned protozoan pathogens for determining treatment needs, an extensive protozoan monitoring dataset from the City of Calgary was used for demonstrating Bayesian retrospective power analysis as a tool to evaluate the level of precision attainable through changes in the design of the monitoring program. Specifically, the loss of precision in the mean concentration estimate attributable to small sample sizes was explicitly quantified. Even with as few as four samples for protozoan enumeration and limited quantification of analytical recovery, mean concentration estimates within 0.5-log_{10} of the true mean can usually be attained across the scenarios investigated. However, extrapolation of the results emanating from this analysis requires that key caveats of stationarity, sampling program duration, assessment based on routine (non-temporally auto-correlated) monitoring results, and consistency with gamma-distributed concentrations must be met. Such an assessment may be a useful first step to inform future monitoring efforts.

The results emanating from this work further underscored that analytical recovery—if accounted for in the estimation of concentration—should not be handled independently of the enumerated count within the analyzed sample volume. Although in many cases analytical recovery may appear uncorrelated with the observed count, the two are necessarily linked. When the expected number of recovered microorganisms is low and/or method recovery is variable, the relationship between these necessarily dependent variables may be obscured. Treating the two variables as being independent will inevitably bias concentration estimates as well as the level of precision attained. The results of this work suggests that the collection of additional samples for pathogen enumeration is likely to improve precision in mean concentration estimation more than devoting additional effort to quantify analytical recovery, especially when the number of pathogens enumerated are expected to be low.

In this work, two goals of precision of the mean concentration estimate were evaluated. One was evaluated with respect to a difference of less than 0.5-log_{10} between the posterior mode and the true mean concentration used to simulate the data; the

other was a 1.0-log_{10} interval corresponding to the 95% of the most probable values of the posterior mean concentrations (i.e., the Bayesian highest density interval [HDI] width). This work underscores that although both goals of precision are related to a span of 1.0-log_{10} , the latter is a much more stringent goal of precision and therefore required on the order of 24 to 52 samples to attain this goal of precision consistently for the scenarios investigated. The HDI was demonstrated to be a useful and perhaps a more intuitively interpretable metric pertaining to the precision of the mean concentration estimate; however, it was noted that computational artefacts can lead to its inflation and hence would require judicious interpretation, especially when non-detects are expected to be prevalent in the modelled (or real) data sets.

6.2.3 Microbial water quality observed during purge water abstraction from monitoring wells are subject to time-limited (bio)colloid mobilization and transport mechanisms

Well water purging is widely regarded as common practice to obtain samples void of artefacts attributable to contrasting hydrogeochemical conditions between the well and the aquifer. Although both chemical and microbial water quality are well-recognized to be subject to influences of the well environment, and that their transport phenomena in the subsurface are known to be mechanistically different, many studies continue to presume that physico-chemical water quality stability achieved through purging of 3 to 5 well volumes is indicative of the stability of target microbial analytes. Therefore, the utility of adenosine triphosphate (ATP) as a sentinel for subsurface microbial water quality change was investigated. Specifically, the fluctuations of this biochemical were monitored throughout well purging activities, with the expectations that i) ATP concentrations will progressively decrease with extended purging, but independently of other physico-chemical water quality parameters, ii) ATP concentrations between two “identical” wells (with similar hydrogeological characteristics) developed in close proximity (<5 m apart, cross-gradient) to each other will become progressively similar, and iii) ATP patterns observed in each well during each sampling occasion may provide insights to subsurface biofilm dynamics.

In Chapter 4 the results of the investigation were presented. Observations generally confirmed these initial hypotheses. Physical and chemical water quality parameters typically stabilized after 3 to 5 well volumes (usually within 10 to 15 minutes of purging), as expected. ATP fluctuations sometimes persisted for the entire duration of purging activities conducted; however, the difference between ATP concentrations observed in both wells generally decreased with extended purging. Although

the confirmation of the first two hypotheses was rather unsurprising given the documented/expected phenomena for other microbial parameters, the emergent, recurring similarities in ATP patterns that were suggestive of hysteretic biofilm behaviour was least expected. (Dis)similarities in ATP patterns over the course of a year are believed to reflect the growth and maturation of subsurface biofilms at relatively stable hydrogeological conditions characteristic of each well. An apparent interruption to the pattern appeared to coincide with a shift in the hydrogeological conditions that may be indicative of surface water groundwater interaction dynamics at the site.

This work reaffirmed the critical importance of dedicated considerations for microbial water quality monitoring in environments that cannot be assumed to be at a pseudo-steady state (e.g., continuously pumping wells). Moreover, this work—amongst a limited handful of published works—explicitly documents the influence of purging on ATP concentrations in monitoring wells and highlights the dedicated need for time-based approaches for microbial water quality sampling from subsurface environments. Although only continuous purging for up to 4 hours was performed in piezometers during the various sampling campaigns, this work provides proof-of-concept for an analogous investigation in other subsurface environments. In the updated Terms of Reference, a continuous, 72 hour pumping test is proposed for new production well installations. While this was originally intended to support the evaluation of well performance, capacity, zone of influence and aquifer characteristics (e.g., storativity, transmissivity, anisotropy), a pumping test of this length likely ensures that microbial water quality samples collected prior to termination of the pumping should be adequately representative of the aquifer.

From an analytical perspective, this work also features the use of dynamic time warping (DTW), a time series analysis approach for which conventional parametric time series analysis approaches would be deemed unsuitable. Where samples are collected at irregular intervals in time or space (often due to minimization of resources to capture anticipated phenomena), features such as peaks and troughs within the time series can be systematically compared to that of another using DTW algorithms. Features are matched given user-specified constraints, and pairwise normalized distances can be generated as a measure of (dis)similarity between time series. (Dis)similarity measures can be subsequently used in a variety of exploratory ordination and clustering approaches that would further allow for the generation and testing of hypotheses. This type of analysis that originated from applications in speech recognition has seldom been used to analyze patterns arising in the natural sciences; this work is the first documented use of DTW for elucidating subsurface biofilm dynamics with ATP patterns.

6.2.4 The use and limitations of “next-generation sequencing” biomolecular tools for supporting vulnerability assessments of subsurface water sources to pathogen intrusion

In the foreseeable future, enumeration-based tests for protozoan pathogens *Cryptosporidium* and *Giardia* and culture-based tests for *Escherichia coli* will continue to be relied upon in microbial monitoring programs to evaluate source water vulnerability to pathogen/fecal contamination, respectively. In light of advances in the past decades, analogous bio-molecular methods for detection and estimation of fecal indicators and pathogen concentrations have become equally, or even more feasible, compared to their enumeration- or culture-based counterparts. They have been touted for their increased sensitivity (probability of detection given that a single sequence were present in the analyzed sample), specificity (detection of specific strains of interest), and quick turn-around-time. However, several fundamental limitations preclude them from unequivocally replacing widely-used enumeration- and culture-based methods:

1. Perfect sensitivity in a small aliquot of the processed sample volume does not enable estimation of low target analyte concentrations
2. PCR chemistry has been deemed to be more influenced by inhibitors than culture-based methods in environmental matrices
3. Viability and infectivity of target pathogens typically require a cell-culture step (e.g., integrated cell-culture reverse transcription polymerase chain reaction)

Steps necessary for biomolecular evaluations such as sample DNA concentration, aliquoting of eluted DNA concentrate, amplification, and sequencing collectively contribute to additional biases that are often not practically quantifiable. Across various forms of quantitative PCR, standard calibration curves are used to estimate concentrations in water samples indirectly. Accordingly, the utility of bacterial community analysis by means of 16S rRNA gene amplicon sequencing was investigated to provide broader context for understanding microbial water quality changes and to identify potentially relevant indicators for a water source’s vulnerability to pathogen intrusion. Negative binomial generalized linear models—which allow for the consideration of overdispersed sequence read counts amongst different sequencing depths achieved for each sample—facilitate systematic comparisons of bacterial taxa sequence read counts across the major spatial and temporal gradients spanned by the study. Despite the quantitative nature of this approach, it is emphasized that sequence read counts

emanating from 16S rRNA gene amplicon sequencing facilitate relative comparisons between samples but are generally not suitable for concentration estimation.

Proof-of-concept of the progressive approach described was presented in Chapter 5, through examining the sequence reads of the bacterial communities present at the same field site used for the ATP investigation (Chapter 4). From this analysis, it was revealed that the bacterial core community in this environment was rather stable at high levels of taxonomic classification; however, differences in bacterial taxa sequence abundance at sub-class taxonomic levels can be linked to the major spatial and temporal gradients spanned in the study. A focused investigation of sequence reads related to known fecal-indicators and sequence reads exhibiting the most dynamic response to each of the key factors identified was used to contextualize subsurface microbial water quality at this site. *Escherichia* sequences were abundant and prevalent in the bacterial core community; however, this observation was not supported by any of the culture-based fecal indicators performed (which consistently yielded non-detects). This observation underscores the disparity between culture- and biomolecular-based methods. Some rare sequence reads affiliated with known anaerobic taxa (*Clostridia*, *Bacteroides*) common to human/mammalian gut microbiome were detected, which may suggest their potential use as site-specific fecal indicator should their relationship to a fecal source be ascertained. Sequence reads affiliated with *Bacillales* and *Sphingomonadales* were significantly elevated in samples collected during the intermediate (between 20 minutes and 1 hour) stage of low-flow purging; these taxa may be useful indicators of the previously-described foci of biomass associated with subsurface biofilms in the vicinity of the well installations. Sequences of several *Betaproteobacteria* taxa were amongst those exhibiting the greatest seasonal responses and may reflect shifting nutrient conditions in the subsurface. Finally, the limited number of *Cyanobacteria* (all of which were *Melainabacteria* and not capable of photosynthesis) and *Flavobacteriia* sequence reads observed may be useful as lines of evidence that recent surface water intrusion may not play a substantial role at this site.

6.3 General conclusion and research outlook: Recommendations for further research

As biological sciences and technology continue to advance, how microbial water quality evaluations are performed will continue to evolve. In this age of “big data”, it can be easy to lose sight of the underlying purpose of a study/tool/technology, or rationale for a particular policy or regulation. More data can be generated than can be

meaningfully interpreted whilst data generators and data analysts take on increasingly specialized roles (e.g., bioinformaticians for some tools). This dissertation serves to inform a prudent level of microbial water quality monitoring, and the potential and limitations of emerging microbial tools to support pathogen vulnerability assessments. Chapters 2 and 4 provide two key examples where concepts and ideas originating from chemistry (i.e., method detection limits, physical and chemical water quality indicators of adequate well purging) have likely propagated—and been indiscriminately (mis)applied—to the realm of microbial water quality evaluations. Chapter 3 highlights that in some circumstances, “more” data (i.e., more samples) will not always yield an appreciable improvement in precision of the mean concentration estimate and associated microbial risk estimates. Finally, Chapter 5 demonstrates the use of an increasingly popular biomolecular method in 16S rRNA gene amplicon sequencing and how it should be interpreted alongside existing culture-based tools rather than outright replacing them. Although microbial data, tools, and approaches will inevitably evolve, this dissertation emphasizes the persistent need to critically evaluate their use and “fit-for-purpose” ability to support the underlying goal of public health protection through the provision of microbiologically safe drinking water.

Based on the conclusions drawn from this work, research in the following areas may contribute to further advancements in the provision of safe drinking water from subsurface water sources:

- Examining the concept of method sensitivity limits as it applies to compare microbial analytical methods, especially PCR-based methods that rely on standard curve calibrations for estimation of concentration
- Tailoring of a hierarchical model to support the use of native microbial indicators (e.g., somatic coliphages) as an estimate of analytical recovery of microbial pathogens (e.g., enterovirus)
- Investigating the generalizability of sample sizes required to make adequately precise mean concentration estimates through further factorial design of simulated scenarios (e.g., as part of a *prospective* power analysis)
- Investigating the impact of substantially higher flow-rates during well-water purging and other subsurface environments, including those which are anaerobic, on the use of ATP as a sentinel for microbial water quality change
- Exploring the use of non-parametric alternatives (such as DTW) to facilitate comparisons of time series data between studies deployed at offset time intervals (e.g., to elucidate mechanisms controlling natural or engineered (bio-)filtration)

- Applying DTW to other domains of public health research to elucidate trends and mechanisms (e.g., positive test cases of coronavirus disease (COVID-19) with hospitalization/mortality rates that are monitored/reported at different time intervals across jurisdictions)
- Characterizing spatio-temporal heterogeneity of bacterial community composition in wells and surface water features in the vicinity of the site investigated (along with “standard” culture-based tools), especially to further elucidate the role(s) of the potential microbial vulnerability indicators identified
- Examining the source of *Escherichia* sequence reads observed and what they imply about source water quality
- Investigating the role of *Betaproteobacteria* in the macro- and micro- nutrient dynamics at this site, especially in context of the biological removal of nitrates through denitrification

References

- Abelho, M. (2005). Extraction and Quantification of ATP as a Measure of Microbial Biomass. In *Methods to Study Litter Decomposition*, pages 223–229. Springer.
- Adam, E. A., Yoder, J. S., Gould, L. H., Hlavsa, M. C., and Gargano, J. W. (2016). Giardiasis Outbreaks in the United States, 1971–2011. *Epidemiol. Infect.*, 144(13):2790–2801.
- Ahmed, A., Emelko, M. B., Conant, B., and Chik, A. H. S. (2013). Development of the New MOE Guidance Document to Determine Minimum Treatment Requirements for Municipal Residential Groundwater Systems. In *OWWA/OMWA Joint Annual Conference*, Ottawa, ON.
- Alberta Environment and Sustainable Resource Development (2012). *Standards and Guidelines for Municipal Waterworks, Wastewater and Storm Drainage Systems*. Alberta Queen’s Printer, Edmonton.
- Allen, M. J., Clancy, J. L., Rice, E. W., and others (2000). The Plain, Hard Truth about Pathogen Monitoring. *Journal AWWA*, 92(9):64–76.
- Analytical Methods Committee (1987). Recommendations for the Definition, Estimation and Use of the Detection Limit. *Analyst*, 112(2):199–204.
- APHA, AWWA, and WEF (2017). *Standard Methods For the Examination of Water & Wastewater*. American Public Health Assn, Washington, DC, 23rd edition.
- Armbruster, D. A. and Pry, T. (2008). Limit of Blank, Limit of Detection and Limit of Quantitation. *The Clinical Biochemist Reviews*, 29(Suppl 1):S49–S52.
- ASTM (2016). *ASTM D5465-16 - Standard Practices for Determining Microbial Colony Counts from Waters Analyzed by Plating Methods*. ASTM.

- Bagnall, A., Lines, J., Bostrom, A., Large, J., and Keogh, E. (2017). The Great Time Series Classification Bake off: A Review and Experimental Evaluation of Recent Algorithmic Advances. *Data Mining and Knowledge Discovery*, 31(3):606–660.
- Balkwill, D. L. and Ghiorse, W. C. (1985). Characterization of Subsurface Bacteria Associated with Two Shallow Aquifers in Oklahoma. *Applied and Environmental Microbiology*, 50(3):580–588.
- Barcelona, M., Varljen, M., Puls, R., and Kaminski, D. (2005). Ground Water Purging and Sampling Methods: History vs. Hysteria. *Groundwater Monitoring & Remediation*, 25(1):52–62.
- Ben Maamar, S., Aquilina, L., Quaiser, A., Pauwels, H., Michon-Coudouel, S., Vergnaud-Ayraud, V., Labasque, T., Roques, C., Abbott, B. W., and Dufresne, A. (2015). Groundwater Isolation Governs Chemistry and Microbial Community Structure along Hydrologic Flowpaths. *Frontiers in Microbiology*, 6.
- Benidickson, J. (2017). The Evolution of Canadian Water Law and Policy: Securing Safe and Sustainable Abundance. *McGill J. Sust. Dev. L.*, 13:59.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.
- Bichai, F. and Smeets, P. W. M. H. (2013). Using QMRA-Based Regulation as a Water Quality Management Tool in the Water Security Challenge: Experience from the Netherlands and Australia. *Water Research*, 47(20):7315–7326.
- Bradford, S., Schijven, J., and Harter, T. (2015). Microbial Transport and Fate in the Subsurface: An Introduction to the Special Collection. *Journal of Environmental Quality*, 10.
- Bradford, S. A. and Torkzaban, S. (2008). Colloid transport and retention in unsaturated porous media: A review of interface-, collector-, and pore-scale processes and models. *Vadose Zone Journal*, 7(2):667–681.
- Braun, B., Schröder, J., Knecht, H., and Szewzyk, U. (2016). Unraveling the Microbial Community of a Cold Groundwater Catchment System. *Water Research*, 107:113–126.
- Brook, J. M. (2012). *Evaluating Innovative Nutrient Management Options and Seasonal Groundwater Recharge Dynamics in an Agricultural Source Water Protection Area*. Master’s Thesis, University of Waterloo, Waterloo, Ontario, Canada.

- Brooks, G., Carroll, K. C., Butel, J., Morse, S. A., and Mietzner, T. A. (2015). Spore-Forming Gram-Positive *Bacilli*: *Bacillus* and *Clostridium* Species. In *Jawetz, Melnick, & Adelberg's Medical Microbiology*. McGraw-Hill Education, New York, NY, 27th edition.
- Busschaert, P., Geeraerd, A. H., Uyttendaele, M., and Van Impe, J. F. (2010). Estimating Distributions out of Qualitative and (Semi) Quantitative Microbiological Contamination Data for Use in Risk Assessment. *Int. J. Food Microbiol.*, 138(3):260–269.
- Buttigieg, P. L. and Ramette, A. (2014). A Guide to Statistical Analysis in Microbial Ecology: A Community-Focused, Living Review of Multivariate Data Analyses. *FEMS microbiology ecology*, 90(3):543–550.
- Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017). Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis. *The ISME Journal*, 11(12):2639–2643.
- Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, 11(4):265–270.
- Chik, A. H. S., Emelko, M. B., Anderson, W. B., O'Sullivan, K. E., Savio, D., Farnleitner, A. H., Blaschke, A. P., and Schijven, J. F. (2020a). Evaluation of Groundwater Microbial Community Composition to Inform Waterborne Pathogen Vulnerability Assessments. *Submitted to Science of the Total Environment*.
- Chik, A. H. S., Emelko, M. B., Blaschke, A. P., and Schijven, J. F. (2020b). Illuminating Subsurface Microbial Water Quality Patterns Using Adenosine Triphosphate and Dynamic Time Warping Approaches. *Ground Water Monitoring & Remediation*.
- Chik, A. H. S., Schmidt, P. J., and Emelko, M. B. (2018). Learning Something From Nothing: The Critical Importance of Rethinking Microbial Non-detects. *Frontiers in Microbiology*, 9.
- Christie, M., Rudolph, D. L., Payment, P., and Locas, A. (2009). Monitoring the Occurrence of Microbial Contaminants within the Wellhead Protection Area of a Municipal Well Field in an Agricultural Setting. In *1st International Conference on Assessing Pathogen Fate, Transport and Risk in Natural and Engineered Water Treatment*, Niagara-on-the-Lake, Ontario, Canada.
- Commeau, N., Parent, E., Delignette-Muller, M.-L., and Cornu, M. (2012). Fitting a Lognormal Distribution to Enumeration and Absence/Presence Data. *Int. J. Food Microbiol.*, 155(3):146–152.

- Corless, C. E., Guiver, M., Borrow, R., Edwards-Jones, V., Kaczmarek, E. B., and Fox, A. J. (2000). Contamination and Sensitivity Issues with a Real-Time Universal 16S rRNA PCR. *Journal of Clinical Microbiology*, 38(5):1747–1752.
- Cotton, C. A., Owen, D. M., Cline, G. C., and Brodeur, T. P. (2001). UV Disinfection Costs for Inactivating *Cryptosporidium*. *J. American Water Works Association*, 93(6):82–94.
- Cousineau, D. and Helie, S. (2013). Improving Maximum Likelihood Estimation Using Prior Probabilities: A Tutorial on Maximum a Posteriori Estimation and an Examination of the Weibull Distribution. *Tutorials in Quantitative Methods for Psychology*, 9(2):61–71.
- Critchley, C. (2010). Stimulating In Situ Denitrification in an Aerobic, Highly Conductive Municipal Drinking Water Aquifer. Master’s thesis, University of Waterloo, Waterloo, Ontario, Canada.
- Critchley, C., Rudolph, D. L., Devlin, J. F., and Schillig, P. C. (2014). Stimulating in Situ Denitrification in an Aerobic, Highly Permeable Municipal Drinking Water Aquifer. *Journal of Contaminant Hydrology*, 171:66–80.
- Cullimore, D. R. (2007). *Practical Manual of Groundwater Microbiology, Second Edition*. CRC Press.
- Currie, L. A. (1999). Detection and Quantification Limits: Origins and Historical Overview. *Anal. Chim. Acta*, 391(2):127–134.
- Davis, C. (2014). Enumeration of Probiotic Strains: Review of Culture-Dependent and Alternative Techniques to Quantify Viable Bacteria. *J. Microbiol. Methods*, 103:9–17.
- Delignette-Muller, M.-L., Dutang, C., Pouillot, R., Denis, J.-B., and Siberchicot, A. (2017). Fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data. *R package*.
- Denwood, M. J. (2016). Runjags: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS. *Journal of Statistical Software*, 71(9):1–25.
- Devlin, J. F. and Schillig, P. C. (2017). HydrogeoEstimatorXL: An Excel-Based Tool for Estimating Hydraulic Gradient Magnitude and Direction. *Hydrogeology Journal*, 25(3):867–875.

- Devlin, J. F., Schillig, P. C., Bowen, I., Critchley, C. E., Rudolph, D. L., Thomson, N. R., Tsoflias, G. P., and Roberts, J. A. (2012). Applications and Implications of Direct Groundwater Velocity Measurement at the Centimetre Scale. *Journal of Contaminant Hydrology*, 127(1):3–14.
- Di Rienzi, S. C., Sharon, I., Wrighton, K. C., Koren, O., Hug, L. A., Thomas, B. C., Goodrich, J. K., Bell, J. T., Spector, T. D., Banfield, J. F., and Ley, R. E. (2013). The Human Gut and Groundwater Harbor Non-Photosynthetic Bacteria Belonging to a New Candidate Phylum Sibling to *Cyanobacteria*. *eLife*, 2:e01102.
- Dickey, R. P., Pyrzak, R., Lu, P. Y., Taylor, S. N., and Rye, P. H. (1999). Comparison of the Sperm Quality Necessary for Successful Intrauterine Insemination with World Health Organization Threshold Values for Normal Sperm. *Fertility Sterility*, 71(4):684–689.
- Duarte, A. S. R., Stockmarr, A., and Nauta, M. J. (2015). Fitting a Distribution to Microbial Counts: Making Sense of Zeroes. *Int. J. Food Microbiol.*, 196:40–50.
- Dupas, R., Tavenard, R., Fovet, O., Gilliet, N., Grimaldi, C., and Gascuel-Oudou, C. (2015). Identifying Seasonal Patterns of Phosphorus Storm Dynamics with Dynamic Time Warping. *Water Resources Research*, 51(11):8868–8882.
- Dürrenmatt, D. J., Del Giudice, D., and Rieckermann, J. (2013). Dynamic Time Warping Improves Sewer Flow Monitoring. *Water Research*, 47(11):3803–3816.
- Edgar, R. C. (2010). Search and Clustering Orders of Magnitude Faster than BLAST. *Bioinformatics*, 26(19):2460–2461.
- Edgar, R. C. (2013). UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads. *Nature Methods*, 10(10):996–998.
- Efstratiou, A., Ongerth, J. E., and Karanis, P. (2017). Waterborne Transmission of Protozoan Parasites: Review of Worldwide Outbreaks - An Update 2011-2016. *Water Research*, 114:14–22.
- Embley, T. M. and Stackebrandt, E. (1997). Species in Practice: Exploring Uncultured Prokaryote Diversity in Natural Samples. *Systematics Association Special Volume*, 54:61–82.
- Emelko, M. B., Schmidt, P. J., and Reilly, P. M. (2010). Particle and Microorganism Enumeration Data: Enabling Quantitative Rigor and Judicious Interpretation. *Environ. Sci. Technol.*, 44(5):1720–1727.

- Emelko, M. B., Schmidt, P. J., and Roberson, J. A. (2008). Quantification of Uncertainty in Microbial Data—reporting and Regulatory Implications. *Journal AWWA*, pages 94–104.
- Emelko, M. B. and Tufenkji, N. (2010). Transport and Fate of Colloids and Microbes in Granular Aqueous Environments. *Water Research*, 44(4):1027.
- Ettema, C. H. and Wardle, D. A. (2002). Spatial Soil Ecology. *Trends in Ecology & Evolution*, 17(4):177–183.
- Fiedler, C. J., Schönher, C., Proksch, P., Kerschbaumer, D. J., Mayr, E., Zunabovic-Pichler, M., Domig, K. J., and Perfler, R. (2018). Assessment of Microbial Community Dynamics in River Bank Filtrate Using High-Throughput Sequencing and Flow Cytometry. *Frontiers in Microbiology*, 9:2887.
- Flynn, T. M., Sanford, R. A., Ryu, H., Bethke, C. M., Levine, A. D., Ashbolt, N. J., and Santo Domingo, J. W. (2013). Functional Microbial Diversity Explains Groundwater Chemistry in a Pristine Aquifer. *BMC Microbiology*, 13:146. 00047.
- Forum on Environmental Measurements (FEM) Microbiology Action Team (2016). Method Validation of U.S. EPA Microbiological Methods of Analysis. Technical Report 2009-01, US EPA.
- Frick, C., Vierheilg, J., Linke, R., Savio, D., Zornig, H., Antensteiner, R., Baumgartner, C., Bucher, C., Blaschke, A. P., Derx, J., Kirschner, A. K. T., Ryzinska-Paier, G., Mayer, R., Seidl, D., Nadiotis-Tsaka, T., Sommer, R., and Farnleitner, A. H. (2018). Poikilothermic Animals as a Previously Unrecognized Source of Fecal Indicator Bacteria in a Backwater Ecosystem of a Large River. *Applied and Environmental Microbiology*, 84(16).
- Frostegård, Å., Courtois, S., Ramišse, V., Clerc, S., Bernillon, D., Gall, F. L., Jeanin, P., Nesme, X., and Simonet, P. (1999). Quantification of Bias Related to the Extraction of DNA Directly from Soils. *Applied and Environmental Microbiology*, 65(12):5409–5420.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*, volume 2. CRC Press Boca Raton, FL.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevelhierarchical Models*, volume 1. Cambridge University Press Cambridge.
- Gerba, C. P. and Rose, J. B. (2003). International Guidelines for Water Recycling: Microbiological Considerations. *Water Sci. Technol.: Water Supply*, 3(4):311–316.

- Ginn, T. R., Camesano, T., Scheibe, T. D., Nelson, K. E., Clement, T. P., and Wood, B. D. (2005). Microbial transport in the subsurface. *Encyclopedia of Hydrological Sciences*.
- Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R : The **dtw** Package. *Journal of Statistical Software*, 31(7).
- Glaser, J. A., Foerst, D. L., McKee, G. D., Quave, S. A., and Budde, W. L. (1981). Trace Analyses for Wastewaters. *Environ. Sci. Technol.*, 15(12):1426–1435.
- Gleit, A. (1985). Estimation for Small Normal Data Sets with Detection Limits. *Environ. Sci. Technol.*, 19(12):1201–1206.
- Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., and Ludwig, W. (2017). 25 Years of Serving the Community with Ribosomal RNA Gene Reference Databases and Tools. *Journal of Biotechnology*, 261:169–176.
- Gonzales-Barron, U. and Butler, F. (2011). A Comparison Between the Discrete Poisson-gamma and Poisson-lognormal Distributions to Characterise Microbial Counts in Foods. *Food Control*, 22(8):1279–1286.
- Gracias, K. S. and McKillip, J. L. (2004). A Review of Conventional Detection and Enumeration Methods for Pathogenic Bacteria in Food. *Can. J. Microbiol.*, 50(11):883–890.
- Graham, E. B., Crump, A. R., Resch, C. T., Fansler, S., Arntzen, E., Kennedy, D. W., Fredrickson, J. K., and Stegen, J. C. (2017). Deterministic Influences Exceed Dispersal Effects on Hydrologically-Connected Microbiomes. *Environmental Microbiology*, 19(4):1552–1567.
- Griebler, C. and Lueders, T. (2009). Microbial Biodiversity in Groundwater Ecosystems. *Freshwater Biology*, 54(4):649–677.
- Grißmeier, V. and Gescher, J. (2018). Influence of the Potential Carbon Sources for Field Denitrification Beds on Their Microbial Diversity and the Fate of Carbon and Nitrate. *Frontiers in Microbiology*, 9.
- Gronewold, A. D., Borsuk, M. E., Wolpert, R. L., and Reckhow, K. H. (2008). An Assessment of Fecal Indicator Bacteria-Based Water Quality Standards. *Environ. Sci. Technol.*, 42(13):4676–4682.

- Grøtan, V. and Engen, S. (2008). Poilog: Poisson Lognormal and Bivariate Poisson Lognormal Distribution. *R package version 0.4*.
- Gülay, A., Musovic, S., Albrechtsen, H.-J., Al-Soud, W. A., Sørensen, S. J., and Smets, B. F. (2016). Ecological Patterns, Diversity and Core Taxa of Microbial Communities in Groundwater-Fed Rapid Gravity Filters. *The ISME Journal*, 10(9):2209–2222.
- Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., and Sodergren, E. (2011). Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons. *Genome Research*, 21(3):494–504.
- Hale, D. (2013). Dynamic Warping of Seismic Images. *Geophysics*, 78(2):S105–S115.
- Hammes, F. and Egli, T. (2010). Cytometric Methods for Measuring Bacteria in Water: Advantages, Pitfalls and Applications. *Analytical and Bioanalytical Chemistry*, 397(3):1083–1095.
- Hammes, F., Goldschmidt, F., Vital, M., Wang, Y., and Egli, T. (2010). Measurement and Interpretation of Microbial Adenosine Tri-Phosphate (ATP) in Aquatic Environments. *Water Research*, 44(13):3915–3923.
- Harter, T. (2003). Publication 8085 - reference: Groundwater Sampling and Monitoring. *Farm Water Quality Planning*.
- Harter, T., Watanabe, N., Li, X., Atwill, E. R., and Samuels, W. (2014). Microbial Groundwater Sampling Protocol for Fecal-Rich Environments. *Groundwater*, 52(S1):126–136.
- Harvey, R. W., Smith, R. L., and George, L. (1984). Effect of organic contamination upon microbial distributions and heterotrophic uptake in a Cape Cod, Mass., aquifer. *Applied and Environmental Microbiology*, 48(6):1197–1202.
- Haslauer, C. P. (2005). Hydrogeologic Analysis of a Complex Aquifer System and Impacts of Changes in Agricultural Practices on Nitrate Concentrations in a Municipal Well Field: Woodstock, Ontario. Master’s thesis, University of Waterloo, Waterloo, Ontario, Canada.
- Havelaar, A., Blumenthal, U. J., Strauss, M., Kay, D., and Bartram, J. (2001). Guidelines: The Current Position. *Water Quality: Guidelines, Standards and Health*, pages 17–42.

- Hazen, T. C., Jiménez, L., de Victoria, G. L., and Fliermans, C. B. (1991). Comparison of bacteria from deep subsurface sediment and adjacent groundwater. *Microbial Ecology*, 22(1):293–304.
- Health Canada (2019). Enteric Protozoa: *Giardia* and *Cryptosporidium*. Technical report, Federal-Provincial-Territorial Committee on Drinking Water of the Federal-Provincial-Territorial Committee on Health and the Environment.
- Helsel, D. R. (2005). *Nondetects and Data Analysis*. John Wiley and Sons, New York.
- Helsel, D. R. (2006). Fabricating Data: How Substituting Values for Nondetects Can Ruin Results, and What Can Be Done about It. *Chemosphere*, 65(11):2434–2439.
- Helsel, D. R. and Cohn, T. A. (1988). Estimation of Descriptive Statistics for Multiply Censored Water Quality Data. *Water Resources Research*, 24(12):1997–2004.
- Hipel, K. W. and McLeod, A. I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*, volume 45. Elsevier.
- Howsam, P. (1988). Biofouling in Wells and Aquifers. *Water and Environment Journal*, 2(2):209–215.
- ISO (2005). ISO 8199:2005 General Guidance on the Enumeration of Micro-Organisms by Culture. Technical Report ISO 8199:2005, International Organization for Standardization, Geneva, Switzerland.
- ISO (2013). Microbiology of Food and Animal Feeding Stuffs – General Requirements and Guidance for Microbiological Examinations. Technical Report ISO 7218:2007, International Organization for Standardization, Geneva, Switzerland.
- Jensen, B. K. (1989). ATP-Related Specific Heterotrophic Activity in Petroleum Contaminated and Uncontaminated Groundwaters. *Canadian Journal of Microbiology*, 35(8):814–818.
- John, A., Onno, W., Larry, J., and others (2007). *Reference Leukocyte (WBC) Differential Count (Proportional) and Evaluation of Instrumental Methods; Approved Standard*. Clinical and Laboratory Standards Institute (CLSI) Document H2O-A2.
- Keer, J. T. and Birch, L. (2003). Molecular Methods for the Assessment of Bacterial Viability. *J. Microbiol. Methods*, 53(2):175–183.
- Kenkel, N. C. and Orloci, L. (1986). Applying Metric and Nonmetric Multidimensional Scaling to Ecological Studies: Some New Results. *Ecology*, 67(4):919–928.

- Keogh, E. and Lin, J. (2005). Clustering of Time-Series Subsequences Is Meaningless: Implications for Previous and Future Research. *Knowledge and Information Systems*, 8(2):154–177.
- Knezev, A. and van der Kooij, D. (2004). Optimisation and significance of ATP analysis for measuring active biomass in granular activated carbon filters used in water treatment. *Water Research*, 38:3971–9.
- Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolk, T., McCall, L.-I., McDonald, D., Melnik, A. V., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., Zaneveld, J. R., Zhu, Q., Caporaso, J. G., and Dorrestein, P. C. (2018). Best Practices for Analysing Microbiomes. *Nature Reviews Microbiology*, 16(7):410.
- Koch, J. (2009). Evaluating Regional Aquifer Vulnerability and BMP Performance in an Agricultural Environment Using a Multi-Scale Data Integration Approach. Master’s thesis, University of Waterloo, Waterloo, Ontario, Canada.
- Korbel, K., Chariton, A., Stephenson, S., Greenfield, P., and Hose, G. C. (2017). Wells Provide a Distorted View of Life in the Aquifer: Implications for Sampling, Monitoring and Assessment of Groundwater Ecosystems. *Scientific Reports*, 7:srep40702.
- Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.
- Kwon, M. J., Sanford, R. A., Park, J., Kirk, M. F., and Bethke, C. M. (2008). Microbiological Response to Well Pumping. *Groundwater*, 46(2):286–294.
- Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., and Clavel, T. (2016). IMNGS: A Comprehensive Open Resource of Processed 16S rRNA Microbial Profiles for Ecology and Diversity Studies. *Scientific Reports*, 6:33721.
- Lambert, D., Peterson, B., and Terpenning, I. (1991). Nondetects, Detection Limits, and the Probability of Detection. *J. American Statistical Association*, 86(414):266–277.
- Lee, J.-H., Lee, B.-J., and Unno, T. (2018). Bacterial Communities in Ground-and Surface Water Mixing Zone Induced by Seasonal Heavy Extraction of Groundwater. *Geomicrobiology Journal*, 35(9):768–774.
- Lin, X., McKinley, J., Resch, C. T., Kaluzny, R., Lauber, C. L., Fredrickson, J., Knight, R., and Konopka, A. (2012). Spatial and Temporal Dynamics of the Microbial Community in the Hanford Unconfined Aquifer. *The ISME Journal*, 6(9):1665–1676.

- Liu, N., Skauge, T., Landa-Marbán, D., Hovland, B., Thorbjørnsen, B., Radu, F. A., Vik, B. F., Baumann, T., and Bødtker, G. (2018). Microfluidic Study of Effects of Flow Velocity and Nutrient Concentration on Biofilm Accumulation and Adhesive Strength in Microchannels. *bioRxiv*, page 375451.
- Lorimer, M. F. and Kiermeier, A. (2007). Analysing Microbiological Data: Tobit or Not Tobit? *Int. J. Food Microbiol.*, 116(3):313–318.
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2. *Genome Biology*, 15:550.
- Lund, B., Baird-Parker, T. C., and Gould, G. W. (2000). *Microbiological Safety and Quality of Food*, volume 1. Springer Science & Business Media.
- Lundkvist, M., Gangelhof, U., Lunding, J., and Flindt, M. R. (2007). Production and Fate of Extracellular Polymeric Substances Produced by Benthic Diatoms and Bacteria: A Laboratory Study. *Estuarine, Coastal and Shelf Science*, 75(3):337–346.
- Lycus, P., Lovise Bøthun, K., Bergaust, L., Peele Shapleigh, J., Reier Bakken, L., and Frostegård, Å. (2017). Phenotypic and Genotypic Richness of Denitrifiers Revealed by a Novel Isolation Strategy. *The ISME Journal*, 11(10):2219–2232.
- Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M., and Stahl, D. A. (2017). *Brock Biology of Microorganisms*. Pearson, New York, 15th edition.
- McCarthy, B. J. (1991). The Use of ATP Measurements in Biodeterioration Studies. In Nelson, W. H., editor, *Physical Methods for Microorganisms Detection*, pages 129–138. CRC Press, Boca Raton FL, USA.
- McElroy, W. D. (1947). The Energy Source for Bioluminescence in an Isolated System. *Proceedings of the National Academy of Sciences*, 33(11):342–345.
- McMurdie, P. J. and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8(4):e61217.
- McMurdie, P. J. and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data is Inadmissible. *PLOS Computational Biology*, 10(4).
- Metge, D. W., Brooks, M. H., Smith, R. L., and Harvey, R. W. (1993). Effect of treated-sewage contamination upon bacterial energy charge, adenine nucleotides, and DNA content in a sandy aquifer on Cape Cod. *Applied and Environmental Microbiology*, 59(7):2304–2310.

- Millard, S. P., Neerchal, N. K., and Dixon, P. (2012). *Environmental Statistics with R and S-Plus Second Edition*. Taylor & Francis Group.
- Minda, R., Joshi, V. P., and Bhattacharjee, S. K. (2008). The Evolutionary Significance of 'obligate' Photoautotrophy of *Cyanobacteria*. *Current Science*, 94(7):850–852.
- Molnar, I. L., Johnson, W. P., Gerhard, J. I., Willson, C. S., and O'Carroll, D. M. (2015). Predicting Colloid Transport through Saturated Porous Media: A Critical Review. *Water Resources Research*, 51(9):6804–6845.
- Mueen, A. and Keogh, E. (2016). Extracting Optimal Performance from Dynamic Time Warping. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2129–2130. ACM.
- Nahrstedt, W. and Gimbel, R. (1996). A Statistical Method for Determining the Reliability of the Analytical Results in the Detection of *Cryptosporidium* and *Giardia*. *Aqua*, 45(3):101–111.
- National Research Council (US) Subcommittee on Microbiological Criteria (1985). *Definitions, Purposes, and Needs for Microbiological Criteria*. National Academies Press (US).
- Nieminski, E. C., Schaefer, F. W., and Ongerth, J. E. (1995). Comparison of Two Methods for Detection of *Giardia* Cysts and *Cryptosporidium* Oocysts in Water. *Appl. Environ. Microbiol.*, 61(5):1714–1719.
- O'Connor, D. (2002). Part One Report of the Walkerton Commission of Inquiry - Ministry of the Attorney General. Technical report, Government of Ontario.
- Odonkor, S. T. and Ampofo, J. K. (2013). *Escherichia coli* as an Indicator of Bacteriological Quality of Water: An Overview. *Microbiology Research*, 4(1):e2–e2.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. (2019). *vegan: Community Ecology Package*. R package version 2.5-6.
- Oline, D. K. and Grant, M. C. (2002). Scaling Patterns of Biomass and Soil Properties: An Empirical Analysis. *Landscape Ecology*, 17(1):13–26.
- Ongerth, J. (2017). *Cryptosporidium* and *Giardia* in Water: Reassessment of Occurrence and Significance. *Journal of Environmental Engineering*, 143(3):04016084.

- Ontario Ministry of Environment, Conservation and Parks (2019). Procedure for Disinfection of Drinking Water in Ontario. Technical Report Ontario Regulation 170/03, “Drinking-Water Systems”, Ministry of Environment, Conservation and Parks, Ontario.
- Palmer, R. J. and White, D. C. (1997). Developmental Biology of Biofilms: Implications for Treatment and Control. *Trends in Microbiology*, 5(11):435–440.
- Parkhurst, D. F. and Stern, D. A. (1998). Determining Average Concentrations of *Cryptosporidium* and Other Pathogens in Water. *Environ. Sci. Technol.*, 32(21):3424–3429.
- Pasha, E. (2018). *Quantifying Groundwater Recharge During Dynamic Seasonality in Cold Climates*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada.
- Petterson, S. R., Signor, R. S., and Ashbolt, N. J. (2007). Incorporating Method Recovery Uncertainties in Stochastic Estimates of Raw Water Protozoan Concentrations for QMRA. *Journal of Water and Health*, 5(S1):51–65.
- Plummer, M. (2016). JAGS: Just another Gibbs sampler.
- Pogoda, K. (2017). Targeted 16S rRNA Metagenomic Analysis of *E. coli* Contaminated Private Drinking Well Waters in Southern Ontario. Master’s thesis, Queen’s University, Kingston, Ontario, Canada.
- Pollock, J., Glendinning, L., Wisedchanwet, T., and Watson, M. (2018). The Madness of Microbiome: Attempting to Find Consensus “Best Practice” for 16S Microbiome Studies. *Applied and Environmental Microbiology*, 84(7):e02627–17.
- Pouillot, R., Hoelzer, K., Chen, Y., and Dennis, S. (2013). Estimating Probability Distributions of Bacterial Concentrations in Food Based on Data Generated Using the Most Probable Number (MPN) Method for Use in Risk Assessment. *Food Control*, 29(2):350–357.
- Puls, R. W. and Barcelona, M. J. (1996). Low-Flow (Minimal Drawdown) Groundwater Sampling Procedures. Technical Report EPA/540/S-95/504, US EPA.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. (2013). The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Research*, 41(D1):D590–D596.

- Quintero-Betancourt, W., Peele, E. R., and Rose, J. B. (2002). *Cryptosporidium Parvum* and *Cyclospora Cayetanensis*: A Review of Laboratory Methods for Detection of These Waterborne Parasites. *Journal of Microbiological Methods*, 49(3):209–224.
- Regli, S., Rose, J. B., Haas, C. N., and Gerba, C. P. (1991). Modeling the Risk from *Giardia* and Viruses in Drinking Water. *Journal AWWA*, pages 76–84.
- Reischer, G. H., Haider, J. M., Sommer, R., Stadler, H., Keiblinger, K. M., Hornek, R., Zerobin, W., Mach, R. L., and Farnleitner, A. H. (2008). Quantitative Microbial Faecal Source Tracking with Sampling Guided by Hydrological Catchment Dynamics. *Environmental Microbiology*, 10(10):2598–2608.
- Reischer, G. H., Kasper, D. C., Steinborn, R., Mach, R. L., and Farnleitner, A. H. (2006). Quantitative PCR Method for Sensitive Detection of Ruminant Fecal Pollution in Freshwater and Evaluation of This Method in Alpine Karstic Regions. *Applied and Environmental Microbiology*, 72(8):5610–5614.
- Reynolds, D. T., Slade, R. B., Sykes, N. J., Jonas, A., and Fricker, C. R. (1999). Detection of *Cryptosporidium* Oocysts in Water: Techniques for Generating Precise Recovery Data. *Journal of Applied Microbiology*, 87(6):804–813.
- Rickard, A. H., Leach, S. A., Hall, L. S., Buswell, C. M., High, N. J., and Handley, P. S. (2002). Phylogenetic Relationships and Coaggregation Ability of Freshwater Biofilm Bacteria. *Applied and Environmental Microbiology*, 68(7):3644–3650.
- Rittman, B. E. (1982). The Effect of Shear Stress on Biofilm Loss Rate. *Biotechnology and Bioengineering*, 24(2):501–506.
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics*, 26(1):139–140.
- Rochelle, P. A., De Leon, R., Johnson, A., Stewart, M. H., and Wolfe, R. L. (1999). Evaluation of Immunomagnetic Separation for Recovery of Infectious *Cryptosporidium parvum* Oocysts from Environmental Samples. *Appl. Environ. Microbiol.*, 65(2):841–845.
- Rose, S. and Long, A. (1988). Monitoring Dissolved Oxygen in Ground Water: Some Basic Considerations. *Ground Water Monitoring & Remediation*, 8(1):93–97.

- Roser, D. J. and Ashbolt, N. J. (2005). *Source Water Quality Assessment and the Management of Pathogens in Surface Catchments and Aquifers*. CRC for Water Quality and Treatment.
- Ross, N., Villemur, R., Marcandella, E., and Deschenes, L. (2001). Assessment of Changes in Biodiversity When a Community of Ultramicrobacteria Isolated from Groundwater Is Stimulated to Form a Biofilm. *Microbial Ecology*, 42(1):56–68.
- Roudnew, B., Lavery, T. J., Seymour, J. R., Jeffries, T. C., and Mitchell, J. G. (2014). Variability in Bacteria and Virus-Like Particle Abundances During Purging of Unconfined Aquifers. *Groundwater*, 52(1):118–124.
- Ruecker, N. J., Braithwaite, S. L., Topp, E., Edge, T., Lapen, D. R., Wilkes, G., Robertson, W., Medeiros, D., Sensen, C. W., and Neumann, N. F. (2007). Tracking Host Sources of *Cryptosporidium spp.* in Raw Water for Improved Health Risk Assessment. *Applied and Environmental Microbiology*, 73(12):3945–3957.
- Sabat, G., Rose, P., Hickey, W. J., and Harkin, J. M. (2000). Selective and Sensitive Method for PCR Amplification of *Escherichia coli* 16S rRNA Genes in Soil. *Applied and Environmental Microbiology*, 66(2):844–849.
- Sakoe, H. and Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49.
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., and Walker, A. W. (2014). Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses. *BMC Biology*, 12(1):87.
- Santo Domingo, J. and Ashbolt, N. J. (2012). Fecal Pollution of Water. In *Encyclopedia of Earth*. National Council for Science and the Environment, Washington, D.C., USA.
- Savio, D., Stadler, P., Reischer, G. H., Demeter, K., Linke, R. B., Blaschke, A. P., Mach, R. L., Kirschner, A. K. T., Stadler, H., and Farnleitner, A. H. (2019). Spring Water of an Alpine Karst Aquifer is Dominated by a Taxonomically Stable but Discharge-Responsive Bacterial Community. *Frontiers in Microbiology*, 10.
- Savio, D., Stadler, P., Reischer, G. H., Kirschner, A. K. T., Demeter, K., Linke, R., Blaschke, A. P., Sommer, R., Szewzyk, U., Wilhartitz, I. C., Mach, R. L., Stadler,

- H., and Farnleitner, A. H. (2018). Opening the Black Box of Spring Water Microbiology from Alpine Karst Aquifers to Support Proactive Drinking Water Resource Management. *WIREs. Water*, 5(3):e1282.
- Schijven, J. F. and de Roda Husman, A. M. (2011). Applications of Quantitative Microbial Source Tracking and Quantitative Microbial Risk Assessment. In Hagedorn, C., Blanch, A. R., and Harwood, V. J., editors, *Microbial Source Tracking: Methods, Applications & Case Studies*, pages 559–583. Springer New York.
- Schijven, J. F. and Hassanizadeh, S. M. (2000). Removal of Viruses by Soil Passage: Overview of Modeling, Processes, and Parameters. *Critical Reviews in Environmental Science and Technology*, 30(1):49–127.
- Schijven, J. F., Teunis, P. F., Rutjes, S. A., Bouwknecht, M., and de Roda Husman, A. M. (2011). QMRASpot: A Tool for Quantitative Microbial Risk Assessment from Surface Water to Potable Water. *Water Research*, 45(17):5564–5576.
- Schmidt, P. J. and Emelko, M. B. (2011). QMRA and Decision-Making: Are We Handling Measurement Errors Associated with Pathogen Concentration Data Correctly? *Water Research*, 45(2):427–438.
- Schmidt, P. J., Emelko, M. B., and Reilly, P. M. (2010). Quantification of Analytical Recovery in Particle and Microorganism Enumeration Methods. *Environ. Sci. Technol.*, 44(5):1705–1712.
- Schmidt, P. J., Emelko, M. B., and Thompson, M. E. (2013). Analytical Recovery of Protozoan Enumeration Methods: Have Drinking Water QMRA Models Corrected or Created Bias? *Water Research*, 47(7):2399–2408.
- Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R., and Lukasik, J. (2002). Microbial Source Tracking: Current Methodology and Future Directions. *Applied and Environmental Microbiology*, 68(12):5796–5803.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423.
- Smeets, P. W. M. H., Rietveld, L. C., van Dijk, J. C., and Medema, G. J. (2010). Practical Applications of Quantitative Microbial Risk Assessment (QMRA) for Water Safety Plans. *Water Science and Technology: A Journal of the International Association on Water Pollution Research*, 61(6):1561–1568.

- Smith, H. J., Zelaya, A. J., De León, K. B., Chakraborty, R., Elias, D. A., Hazen, T. C., Arkin, A. P., Cunningham, A. B., and Fields, M. W. (2018). Impact of Hydrologic Boundaries on Microbial Planktonic and Biofilm Communities in Shallow Terrestrial Subsurface Environments. *FEMS Microbiology Ecology*, 94(12).
- Smith, S. A. and Comeskey, A. E. (2009). *Sustainable Wells: Maintenance, Problem Prevention, and Rehabilitation*. CRC Press.
- Sorensen, J. P. R., Maurice, L., Edwards, F. K., Lapworth, D. J., Read, D. S., Allen, D., Butcher, A. S., Newbold, L. K., Townsend, B. R., and Williams, P. J. (2013). Using Boreholes as Windows into Groundwater Ecosystems. *PLOS ONE*, 8(7):e70264.
- Sousa, M. (2013). *Using Numerical Models for Managing Water Quality in Public Supply Wells*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2007). OpenBUGS, Version 3.0. 2. User Manual. *MRC and Imperial College of Science, Technology and Medicine*. Available at <http://mathstat.helsinki.fi/openbugs/>[Verified August 2009].
- Strigrow, B. (2017). Groundwater Sampling. Technical Report SESDPROC-301-R4, US Environmental Protection Agency, Science and Ecosystem Support Division, Athens, Georgia.
- Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., and Nishijima, M. (2014). Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing. *PLOS ONE*, 9(8):e105592.
- Teunis, P. F. M., Medema, G. J., Kruidenier, L., and Havelaar, A. H. (1997). Assessment of the Risk of Infection by *Cryptosporidium* or *Giardia* in Drinking Water from a Surface Water Source. *Water Research*, 31(6):1333–1346.
- Tolhurst, T. J., Consalvey, M., and Paterson, D. M. (2008). Changes in Cohesive Sediment Properties Associated with the Growth of a Diatom Biofilm. *Hydrobiologia*, 596(1):225–239.
- Tufenkji, N. (2007). Modeling Microbial Transport in Porous Media: Traditional Approaches and Recent Developments. *Advances in Water Resources*, 30(6):1455–1469.
- Turnbaugh, P. J., Hamady, M., Yatsunencko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., and Gordon, J. I. (2009). A Core Gut Microbiome in Obese and Lean Twins. *Nature*, 457(7228):480–484.

- United States Pharmacopeial Convention (2014). <61> Microbiological Examination of Nonsterile Products: Microbial Enumeration Tests. Technical Report USP 37, United States Pharmacopeial Convention.
- US EPA (2001). Method 1601: Male-Specific (F+) and Somatic Coliphage in Water by Two-Step Enrichment Procedure. Technical Report EPA 821-R-01-030, Office of Water, Washington, D.C.
- US EPA (2005). Method 1623: *Cryptosporidium* and *Giardia* in Water by Filtration/IMS/FA. Technical Report EPA 815-R-05-002, US EPA.
- US EPA (2006). National Primary Drinking Water Regulations: Long Term 2 Enhanced Surface Water Treatment Rule.
- US EPA (2012). Method 1623.1: *Cryptosporidium* and *Giardia* in Water by Filtration/IMS/FA. Technical Report EPA 816-R-12-001, US EPA.
- US EPA (2016). Definition and Procedure for the Determination of the Method Detection Limit, Revision 2. Technical Report EPA821-R-16-0 06, US EPA Office of Water, Washington, D.C.
- US FDA (1998/2017). Bacteriological Analytical Manual. Technical Report Edition 8 Revision A, 1998, US Food and Drug Administration.
- van der Kooij, D., Bakker, G. L., Italiaander, R., Veenendaal, H. R., and Wullings, B. A. (2017). Biofilm Composition and Threshold Concentration for Growth of *Legionella Pneumophila* on Surfaces Exposed to Flowing Warm Tap Water without Disinfectant. *Applied and Environmental Microbiology*, 83(5):e02737–16.
- van der Wielen, P. W. J. J. and Medema, G. (2010). Unsuitability of Quantitative Bacteroidales 16S rRNA Gene Assays for Discerning Fecal Contamination of Drinking Water. *Applied and Environmental Microbiology*, 76(14):4876–4881.
- van Driezum, I. H., Chik, A. H. S., Jakwerth, S., Lindner, G., Farnleitner, A. H., Sommer, R., Blaschke, A. P., and Kirschner, A. K. T. (2018). Spatiotemporal Analysis of Bacterial Biomass and Activity to Understand Surface and Groundwater Interactions in a Highly Dynamic Riverbank Filtration System. *Science of the Total Environment*, 627:450–461.
- van Driezum, I. H., Derox, J., Saracevic, E., Kirschner, A. K., Sommer, R., Farnleitner, A. H., and Blaschke, A. P. (2017). Does Pumping Volume Affect the Concentration of Micropollutants in Groundwater Samples? *Groundwater Monitoring & Remediation*, 37(4):82–88.

- Vierheilig, J., Savio, D., Ley, R. E., Mach, R. L., Farnleitner, A. H., and Reischer, G. H. (2015). Potential Applications of Next Generation DNA Sequencing of 16S rRNA Gene Amplicons in Microbial Water Quality Monitoring. *Water Science and Technology*, 72(11):1962–1972.
- von Wintzingerode, F., Göbel, U. B., and Stackebrandt, E. (1997). Determination of Microbial Diversity in Environmental Samples: Pitfalls of PCR-Based rRNA Analysis. *FEMS Microbiology Reviews*, 21(3):213–229.
- Webster, J. J., Hampton, G. J., Wilson, J. T., Ghiorse, W. C., and Leach, F. R. (1985). Determination of Microbial Cell Numbers in Subsurface Samples. *Groundwater*, 23(1):17–25.
- WHO (2017). Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum. Technical report, World Health Organization.
- Williams, M. S. and Ebel, E. D. (2012). Methods for Fitting the Poisson-Lognormal Distribution to Microbial Testing Data. *Food Control*, 27(1):73–80.
- Worley-Parsons (2015). Roberts Bank Terminal 2: Biofilm Regeneration Study. Technical Data Report 307071-00790-01-EN-REP-0401, Worley Parsons Canada.
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby, J., Amann, R., and Rosselló-Móra, R. (2014). Uniting the Classification of Cultured and Uncultured Bacteria and Archaea Using 16S rRNA Gene Sequences. *Nature Reviews Microbiology*, 12(9):635–645.
- Yates, M. V., Yates, S. R., and Gerba, C. P. (1988). Modeling Microbial Fate in the Subsurface Environment. *Critical Reviews in Environmental Control*, 17(4):307–344.
- Young, I. M., Crawford, J. W., Nunan, N., Otten, W., and Spiers, A. (2008). Microbial Distribution in Soils: Physics and Scaling. *Advances in Agronomy*, 100:81–121.

Appendices

Appendix A

Supporting information for
*Learning something from nothing:
The critical importance of rethinking
microbial non-detects*

A1 Calculation of the probability of a non-detect observation

In this work, the number of microorganisms present within a sample volume V collected from a presumably homogeneous source with a true concentration of c can be represented as the Poisson random variable X :

$$P(X = x) = \frac{e^{-cV} (cV)^x}{x!} \quad (\text{A.1})$$

The probability of a non-detect observation assuming all microorganisms that are present are successfully recovered and enumerated is simply:

$$P(X = 0) = \frac{e^{-cV} (cV)^0}{0!} = e^{-cV} \quad (\text{A.2})$$

The product of the concentration and the volume provides the number of microorganisms that can be expected to be present. Therefore, the expected number of microorganisms for a recovery of r would be multiplied by this factor, as reflected in the exponent:

$$P(X = 0|r) = e^{-cVr} \quad (\text{A.3})$$

Recovery can also be a random variable R represented by the standard beta distribution with parameters a , b and the following characteristics:

$$E[R] = \frac{a}{a+b} \quad (\text{A.4})$$

$$Var[R] = \frac{ab}{(a+b)^2(a+b+1)} \quad (\text{A.5})$$

$$f_R(r) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} r^{a-1} (1-r)^{b-1} \quad (\text{A.6})$$

The probability of a non-detect considering the recovery profile as defined by the beta distribution is obtained by integrating over the range of possible recovery values, thereby obtaining the confluent hypergeometric function ${}_1F_1(a, b, c, V)$ (*hyperg1F1* in R):

$$P(X = 0; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 e^{-cVr} r^{a-1} (1-r)^{b-1} dr \quad (\text{A.7})$$

A2 Evaluation of the posterior probability density function of the microbial concentration based on a single observation

In this work, Bayesian techniques were used to evaluate the uncertainty in the true concentration. A relatively uninformative semi-infinite uniform prior ($f(c) = 1, c > 0$) on the concentration was updated with the additional information provided by the count observation (incorporated through its likelihood function) to obtain a posterior distribution. Based on Bayes theorem:

$$\text{posterior} \propto \text{prior} \times \text{likelihood} \tag{A.8}$$

The likelihood function of the concentration is given by:

$$\mathcal{L}(c) = \frac{e^{-cV}(cV)^x}{x!} \tag{A.9}$$

Therefore the posterior becomes:

$$P(c|x) \propto \frac{e^{-cV}(cV)^x}{x!} \tag{A.10}$$

$$\propto (cV)^x e^{-cV} \tag{A.11}$$

This posterior for concentration has the form of a gamma distribution with shape parameter $\rho = x + 1$ and scale parameter $\lambda = 1/V$.

A3 Fitting a Poisson log-normal distribution to *Giardia* cyst raw data assuming 100% analytical recovery

In this work, maximum likelihood estimates (MLEs) of arithmetic mean and standard deviation of the log-normal distribution within the Poisson log-normal model for count data were evaluated based on raw *Giardia* data (including both counts and sample volumes) and assuming 100% analytical recovery of the method (i.e., perfect sensitivity). Analytical recovery data were available for this dataset based on ColorSeed™ internal seeding standards; however, such data are not widely available in practice for all drinking water treatment systems and therefore have not been widely mandated in the current interpretation and application of water regulations such as those specified in Alberta ((Alberta Environment and Sustainable Resource Development, 2012)). Variable analytical recovery and other factors influencing *Giardia* concentration estimates can be accounted for within hierarchical Bayesian frameworks, but is not considered here for simplicity and has been discussed elsewhere (e.g., Schmidt et al., 2013). The concentration c can be estimated if n samples with volumes ($V_i, i = 1, 2, \dots, n$) are assumed to be independent and collectively representative of the log-normally distributed concentration. If the cysts in the source are randomly dispersed and the sample is well mixed, then the number of cysts observed in each sample (X_i) is Poisson-distributed with mean cV_i , assuming all present cysts are successfully enumerated. By accounting for random sampling error, non-detects are integrated seamlessly with non-zero counts (and their respective analytical volumes).

$$\ln(c) \sim N(\mu, \sigma^2) \tag{A.12}$$

$$X_i \sim \text{Poisson}(\lambda_i) \tag{A.13}$$

$$\text{where } \lambda_i = cV_i$$

$$\text{and } \ln(\lambda_i) = \ln(c) + \ln(V_i)$$

By mathematical expectation,

$$E[\ln(c)] = \mu \tag{A.14}$$

$$\text{Var}[\ln(c)] = \sigma^2 \tag{A.15}$$

The expectation of the natural logarithm of λ yields:

$$\begin{aligned} E[\ln(\lambda_i)] &= E[\ln(c)] + E[\ln(V_i)] \\ &= \mu + \ln(V_i) \end{aligned} \tag{A.16}$$

$$\begin{aligned} \text{Var}[\ln(\lambda_i)] &= \text{Var}[\ln(c)] + \text{Var}[\ln(V_i)] \\ &= \sigma^2 + 0 = \sigma^2 \end{aligned} \tag{A.17}$$

Therefore, the probability of the number of cysts observed in each sample provided the assumptions inherent in Equations A.12 and A.13 is Poisson log-normally distributed with mean $\mu + \ln(V_i)$ and standard deviation σ . The probability density function can be solved numerically using the *poilog* package (v. 0.4) ((Grøtan and Engen, 2008)) in *R*, which was originally developed for the analysis of species abundance data in ecology. The *dpoilog* function within the package does not explicitly allow for the specification of “sampling intensity” (or in this case, the sample volume) as the sampling intensity is seldom known for species abundance data in many ecological applications. However, given mathematical expectation properties (Equations A.16 and A.17), we can re-express the probability of the number of cysts observed in each sample in terms of μ , σ and the measured volume V_i :

$$P(X_i|\mu; \sigma; V_i) = \text{dpoilog}(X_i, \mu + \ln(V_i), \sigma) \tag{A.18}$$

Clearly, the unbiased estimates of μ and σ accounting for measured (i.e., known) sample volumes can be easily determined (V. Grøtan, *personal communication*, January 9th, 2018). The joint probability of all X_i observations gives the likelihood function that can be expressed as the product of the probability of each observation.

$$\mathcal{L} = \prod_{i=1}^n \text{dpoilog}(X_i, \mu + \ln(V_i), \sigma) \tag{A.19}$$

The resulting likelihood function can be maximized by testing values of the MLE parameters μ and σ iteratively.

Appendix B

Supporting information for

*An end justified by means: Ensuring
adequate treatment of protozoa by precise
mean concentration estimation*

B1 Hierarchical model description

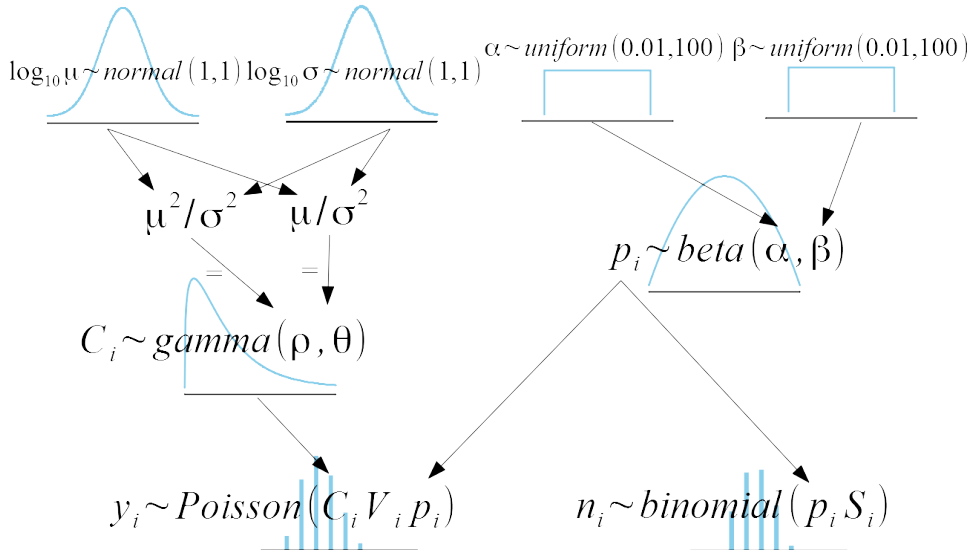


Figure B1: Hierarchical model used to estimate posterior distribution of mean protozoan concentrations (μ)

The probabilistic model and the priors used in the Bayesian analysis are represented as follows:

$\log_{10}(\mu) \sim \text{Normal}(1,1)$	(prior distribution)
$\log_{10}(\sigma) \sim \text{Normal}(1,1)$	(prior distribution)
$\rho = \mu^2/\sigma^2$	(shape parameter of the gamma distribution)
$\theta = \mu/\sigma^2$	(rate parameter of the gamma distribution)
$C_i \sim \text{gamma}(\rho, \theta)$	(in units of microorganisms/100 L)
$y_i \sim \text{Poisson}(p_i C_i V_i)$	(no. of enumerated microorganisms in processed sample volume)
$\alpha \sim \text{uniform}(0.01,100)$	(prior distribution)
$\beta \sim \text{uniform}(0.01,100)$	(prior distribution)
$p_i \sim \text{beta}(\alpha, \beta)$	(stochastic, analytical recovery)
$n_i \sim \text{binomial}(p_i S_i)$	(no. of microorganisms recovered from matrix-spike)

Appendix C

Supporting information for
*Illuminating subsurface microbial water
quality patterns using adenosine triphosphate
and dynamic time warping approaches*

C1 Groundwater levels

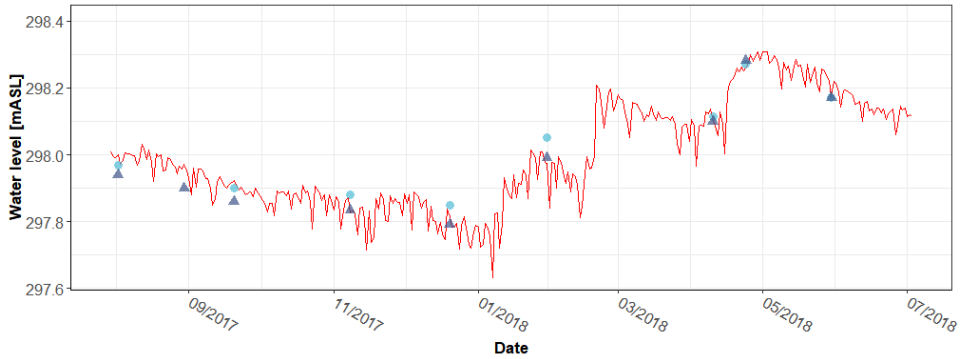


Figure C1: Groundwater levels (in meters above sea level) were measured manually in wells WO77 and WO78 (dark blue triangles) during each of the sampling events and consistently differed by less than 1 cm between the two wells. Groundwater levels were continuously monitored at nearby well WO75 (red) and calibrated using manual water levels (light blue circles).

C2 Illustrative example of dynamic time warping (DTW) distance determination

Although model-based approaches have been applied to many time series datasets ((Bagnall et al., 2017)), and extensively in various environmental applications ((Hipel and McLeod, 1994)), the relatively short time series (n per time series ≤ 20) encountered in this study are unique in that they do not typically contain sufficient information for estimating model parameters. While model-based approaches may be applied for long time series with unequal lengths ((Bagnall et al., 2017)), they are not commonly proposed for investigating time series with irregularly-spaced intervals. We underscore that time series analysis as presented in this work exploits the dynamics exhibited by the *whole* time series rather than intervals or subsequences of the time series ((Keogh and Lin, 2005))—all features exhibited by the data sequence will be incorporated into the estimation of the pairwise, normalized DTW distance and therefore are implicitly considered in analyses related to the ATP pattern (dis)similarities. Therefore, the approach described herein provides an organizational framework to further explore potential mechanisms and drivers of the subsurface biofilm patterns observed.

Briefly, pairs of normalized sequences (say, T1 and T2) of length m and n respectively are aligned using DTW by constructing an m by n matrix, where each element comprises of a local distance measure between the i th datum of T1 and j th datum of T2. A path through the cost matrix that minimizes the total cumulative distance between T1 and T2, subject to imposed constraints (i.e., boundary conditions, monotonic condition, continuity condition, adjustment window condition), is designated as the warping path for optimal alignment of the time series. In this work, the optimal warping path through the distance matrix was constrained using a slanted band window of size 3 (i.e., allowing for a measurement to be aligned with another taken within three sequential measurements) and a symmetric step pattern (i.e., equally weighing a diagonal step with the sum of equivalent Manhattan steps) with an open-ended alignment (to accommodate sampling occasions that had to be terminated before the four hour mark). This temporal realignment allowed for ATP time series of differing lengths and irregular sampling intervals to be compared, while providing some tolerance for out-of-phase (i.e., shifts in the position of) peak ATP concentrations.

Consider the two time series from WO77 in April 2018 presented in Table C1. ATP concentrations were re-scaled between 1 and 0 with respect to the maximum and minimum values of each respective time series. A cost matrix comprising the Euclidean distance between each pair of normalized values was calculated, subject to a "slanted

Table C1: Adenosine triphosphate concentrations measured in piezometer WO77, April 2018

Time since start of purging [minutes]	5	10	15	20	30	40
WO77 - April 10, 2018 [pg/mL]	0.1925	0.1866	0.2491	0.2661	0.1472	0.1699
Normalized ATP [-]	0.2286	0.2000	0.5144	0.6001	0	0.1143
WO77 - April 24, 2018 [pg/mL]	0.2917	0.1379	0.0796	0.1220	0.2228	0.0689
Normalized ATP [-]	0.6688	0.2070	0.0318	0.1592	0.4618	0
Time since start of purging [minutes]	50	60	70	80	90	100
WO77 - April 10, 2018 [pg/mL]	0.1982	0.1925	0.3220	0.3220	0.1756	0.1990
Normalized ATP [-]	0.2572	0.2286	0.8818	0.8818	0.1433	0.2614
WO77 - April 24, 2018 [pg/mL]	0.1485	0.1962	0.3406	0.3797	0.0345	0.2289
Normalized ATP [-]	0.2388	0.3822	0.8156	0.9329	0.4971	0.4803
Time since start of purging [minutes]	120	140	160	180		
WO77 - April 10, 2018 [pg/mL]	0.2517	0.3454	0.2985	0.1873		
Normalized ATP [-]	0.5273	1	0.7637	0.2024		
WO77 - April 24, 2018 [pg/mL]	0.3406	0.4020	0.2289	0.1675		
Normalized ATP [-]	0.8156	1	0.4803	0.2959		

band” window that allows for the distance to be calculated for values within three sequential measurements (values outside this window are not compared) (Table C2).

A continuous path through the cost matrix subject to a symmetric2 step pattern (i.e., equally weighted diagonal steps with the sum of equivalent Manhattan steps) that minimizes the cumulative distance between the two time series is designated as the optimal warping path (Figure C2). This path allows for features (such as peaks and troughs) to be matched, or “re-aligned” from one time series to the other (Figure C3). A smaller “normalized dynamic time warping (DTW) distance” (i.e., the cumulative distance of the optimal path normalized by the number of steps taken) reflects time series that are closely aligned (more similar). A normalized DTW distance is calculated for each pair of time series compared to generate a triangular similarity distance matrix for subsequent hierarchical clustering.

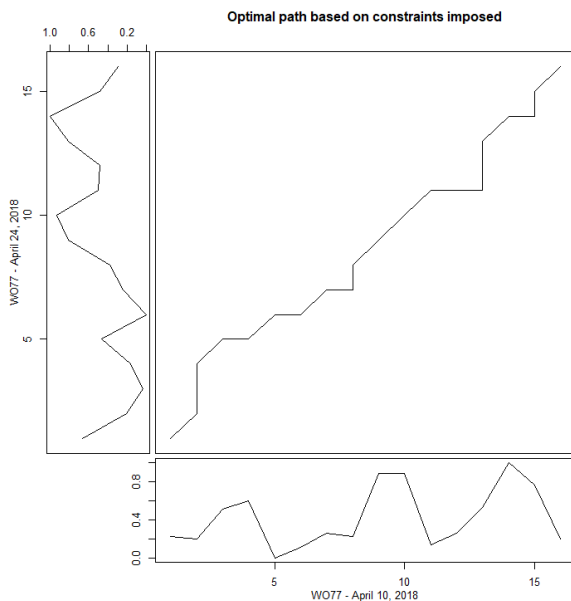


Figure C2: Optimal dynamic time warping path based on constraints imposed (i.e., slanted band, window size of 3, open-ended alignment, symmetric2 step pattern).

Calculated DTW distances can be quite sensitive to the constraints that are imposed. The constraints imposed must be tailored to the application and the nature of the time series data available. In this study, an open-ended DTW alignment with a slanted band of size 3 and a symmetric2 step pattern was used to accommodate time

Table C2: Matrix of Euclidean distances between each pair of normalized ATP concentrations (Vertical: April 10, Horizontal: April 24)

	0.6688	0.2070	0.0318	0.1592	0.4618	...	0.2959
0.2286	0.4402	0.0216	0.1968	0.0694	-	-	-
0.2000	0.4687	0.0070	0.1682	0.0408	0.2617	-	-
0.5144	0.1544	0.3074	0.4825	0.3552	0.0526	...	-
0.6001	0.0687	0.3931	0.5683	0.4409	0.1383	...	-
0	-	0.2070	0.0318	0.1592	0.4618
...	-	-	0.4677
0.2024	-	-	-	0.2780	0.0936

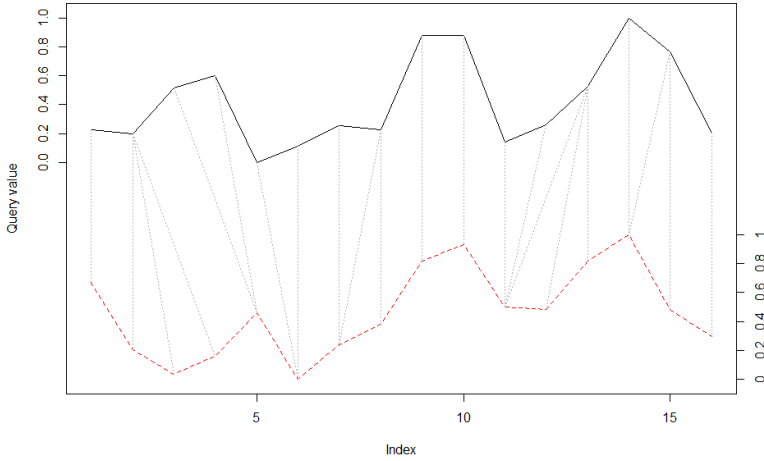


Figure C3: Time series aligned by dynamic time warping. Black: April 10, Red: April 24.

series of differing lengths, with the potential for a sample of position n in one time series to be aligned to a sample of position $n \pm 3$ in another time series, and a continuity constraint that equally weights a diagonal step with the sum of the equivalent Manhattan steps within the distance matrix, respectively. While the results of the current analysis were less sensitive to minor changes to these constraints, application- and data-specific adjustments to the stringency of constraints imposed may be required (e.g., adjustment window condition may be relaxed for a time series with relatively small but frequent time intervals). Although more advanced, complex algorithms for DTW distance calculation may also be adapted to the approach described herein (e.g., weighted DTW, derivative DTW, or combinations thereof)((Bagnall et al., 2017)), the “basic” algorithm applied in this work offers an adaptable, parsimonious approach to support exploratory data analysis and the generation of hypotheses when working with time series with irregularly-spaced sampling intervals.

C3 Background physical and chemical water quality

Table C3: Summary statistics of background physical and chemical water quality characteristics

Parameter	WO77			WO78		
	Mean	SD	n	Mean	SD	n
Temperature (°C)	8.78	0.92	126	9.57	0.74	190
Electrical conductivity ($\mu\text{S}/\text{cm}$)	521.9	27.9	126	563.2	38.6	190
pH	7.18	0.15	126	7.23	0.13	190
Turbidity (NTU)	8.78	0.92	126	9.57	0.74	190
Dissolved oxygen(mg/L)	8.20	1.20	126	8.16	1.47	190
Nitrate (mg/L)	13.8	0.55	10	13.7	0.38	10
Nitrite (mg/L)	<MDL	-	10	<MDL	-	10
Sulfate (mg/L)	33.2	1.69	7	33.3	0.79	10
Bromide (mg/L)	<MDL	-	7	<MDL	-	10
Manganese (mg/L)	<MDL	-	7	<MDL	-	10
Iron (mg/L)	<MDL	-	7	<MDL	-	10
ATP (pg/mL)	0.53	1.14	126	1.04	4.66	190

MDL - method detection limit

C4 Supporting microbial water quality parameters evaluated

BARTTM reactions were observed for 10 days as per the manufacturers' instructions. The day on which a specific reaction occurs is inversely proportional to estimated abundance of related bacteria responsible for their expression in the original sample collected. Various reactions may be observed for the same test; each reaction may also be expressed to a different extent (Figures C4 and C5), which may not necessarily be captured by the number of days until the *first* reaction was observed as summarized in the Table C4 below. Flow cytometry performed to estimate cell densities for the corresponding samples were generally deemed below reliable quantification limits of the method.

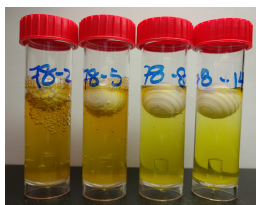


Figure C4: Example of IRB-BARTTM observed from sampling occasion conducted April 24, 2018 (WO78), arranged by increasing extent of purging. Reactions for iron-related bacteria (foaming, brown coloration) are visibly more prominent in the early stages of purging than the latter.

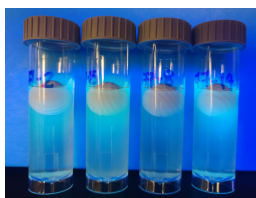


Figure C5: Example of SLYM-BARTTM observed from sampling occasion conducted April 24, 2018 (WO77), arranged by increasing extent of purging. Reactions for slime-forming bacteria (fluorescence, cloudiness) are visibly more prominent in the first two samples than the latter two, even though reactions were recorded to have occurred on the same day.

Table C4: Summary statistics of flow cytometry and BARTTM results

Date	Time ^e	WO77				WO78					
		FCM ^b	IRB ^c	SRB ^c	SLYM ^c	ALGE ^c	FCM ^e	IRB ^c	SRB ^c	SLYM ^c	ALGE ^c
Dec. 20, 2017	10	0.11	3	>10	2	>10	0.06	3	>10	2	>10
Dec. 20, 2017	60	0.14	4	>10	2	>10	0	6	>10	2	>10
Dec. 20, 2017	100	0.01	4	>10	2	>10	0.01	6	>10	2	>10
Dec. 20, 2017	140	1.33	4	>10	2	>10	0.01	4	>10	2	>10
Jan. 30, 2018	10	0.38	4	>10	3	>10	20.1	5	>10	3	>10
Jan. 30, 2018	30	0.10	5	>10	3	>10	0.14	4	>10	3	>10
Jan. 30, 2018	60	0.05	4	>10	3	>10	n/a	4	>10	3	>10
Jan. 30, 2018	140	0.03	5	>10	3	>10	0.04	4	>10	3	>10
Apr. 10, 2018	10	0.14	4	>10	3	>10	0.26	6	10	3	>10
Apr. 10, 2018	30	0.10	6	>10	3	>10	0.07	4	>10	3	>10
Apr. 10, 2018	60	0.06	4	>10	3	>10	0.01	3	8	3	>10
Apr. 10, 2018	140	0.03	6	>10	3	>10	0.01	6	>10	3	>10
Apr. 24, 2018	10	0.06	4	>10	3	>10	0.26	4	>10	3	>10
Apr. 24, 2018	30	0.06	6	10	3	>10	0.07	4	>10	3	>10
Apr. 24, 2018	60	0	4	>10	3	>10	0.04	6	10	3	>10
Apr. 24, 2018	140	0.06	9	>10	3	>10	0.04	6	>10	3	>10
May 30, 2018	10	0.15	4	>10	3	>10	0.12	4	>10	3	>10
May 30, 2018	30	0.10	4	>10	3	>10	0.06	4	>10	3	>10
May 30, 2018	60	0.12	5	7	3	>10	0.04	4	>10	3	>10
May 30, 2018	140	0.17	5	>10	3	>10	0.04	4	>10	3	>10

^atime since start of purging [min]

^b flow cytometry (FCM) sample concentration estimate [count/ μ L]

^cIRB - iron-related bacteria, SRB - sulfate reducing bacteria, SLYM - slime forming bacteria, ALGE - algae; shown are the number of days until *first* clear reaction.

Higher values indicate less aggressive/abundant microorganisms in the original sample.

Appendix D

Supporting information for
*Evaluation of groundwater bacterial
community composition to inform waterborne
pathogen vulnerability assessments*

D1 Groundwater levels

Groundwater levels during the study period are presented in Appendix C1.

D2 Water quality parameters evaluated along with 16S rRNA gene amplicon sequencing samples

Table D1: Summary statistics of physical and chemical water quality characteristics (subset of parameters from Chik *et al.*, [in review])

Parameter	WO77			WO78		
	Mean	SD	n ^a	Mean	SD	n ^a
Temperature (°C)	8.91	0.92	37	9.41	0.57	46
Electrical conductivity ($\mu\text{S}/\text{cm}$)	538	20.2	37	551	30.4	45
pH	7.16	0.18	37	7.24	0.15	46
Turbidity (NTU)	0.29	0.28	37	0.33	0.54	46
Dissolved oxygen (mg/L)	8.05	1.43	37	7.92	1.65	46
Nitrate (mg/L)	13.8	0.55	10	13.7	0.38	10
Nitrite (mg/L)	<0.01 ^c	-	10	<0.01	-	10
Sulfate (mg/L)	33.2	1.69	10	33.3	0.79	10
Bromide (mg/L)	<0.10	-	10	<0.10	-	10
Manganese (mg/L)	<0.0005	-	10	<0.0005	-	10
Iron (mg/L)	<0.05	-	10	<0.05	-	10
ATP (pg/mL)	0.52	1.13	130	0.99	4.54	200

^aChemical parameters were monitored intensively during a sampling event prior to this study and ascertained to stabilize for the duration of the event after three well purge volumes (generally achieved within <10 to 15 minutes of well purging); select chemical water quality parameters were henceforth evaluated at least once per event at the end of the last purging stage for each well.

D3 Quality control samples

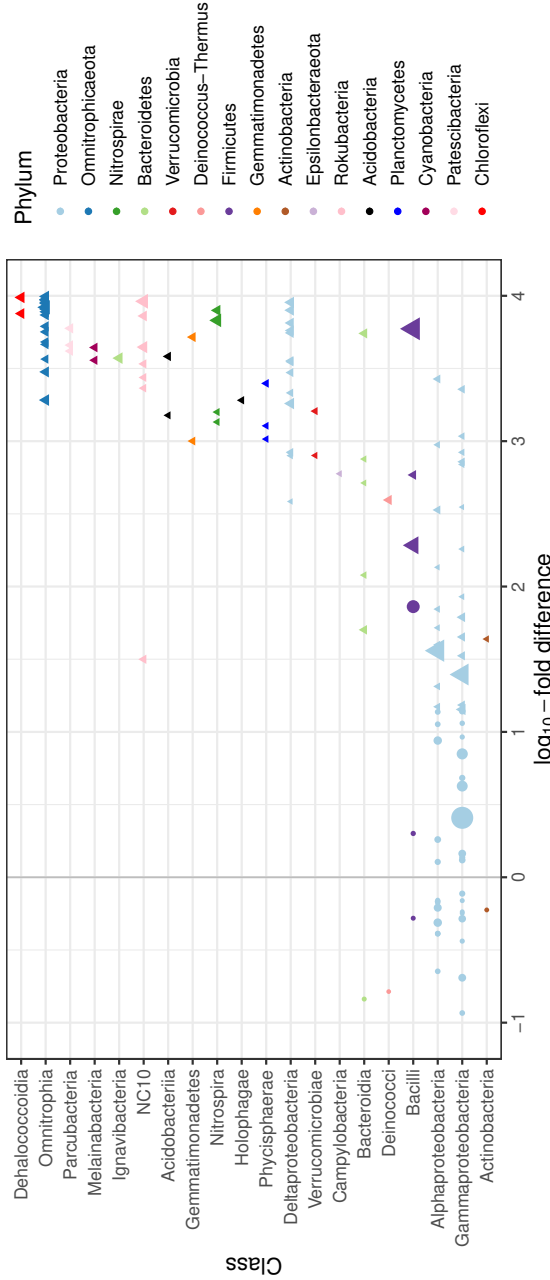


Figure D1: Differential abundance of various OTUs by class, in order of decreasing log₁₀-fold differences in comparison to the negative controls as the base level. Circles represent OTUs that did not exhibit statistically significant different factor coefficients ($p < 0.05$), triangles represent those that are significant. The size of the symbols is proportional to the normalized mean number of sequences of the OTU across all samples.

183 of 259 OTUs that allowed for the sequence counts to converge within the NB GLM framework were identified to be significantly different in abundance in purge water samples compared to negative controls, of which 12 were found exclusively in the purge water samples (Wald-test, $p < 0.05$). The majority of OTU sequences were expressed higher in the purge water samples than in the negative controls. All taxa with sequences at least 2.0-log_{10} -fold higher in the purge water samples than in the negative controls were statistically significant. However, there are some notable exceptions (e.g., sequences affiliated with the genus *Escherichia/Shigella* were only 1.395-log_{10} greater than those exhibited in the negative controls but deemed statistically significant [$p = 0.00096$]).

D4 Bacterial core community analysis

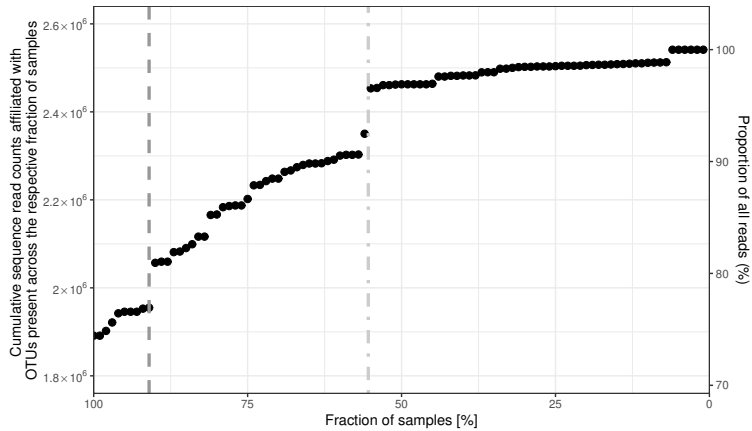


Figure D2: Empirical taxa abundance distribution to determine the core community across all purge water samples. Dashed and dash-dotted lines denote key discontinuities observed in at least 91% and 56% of purge water samples, respectively.

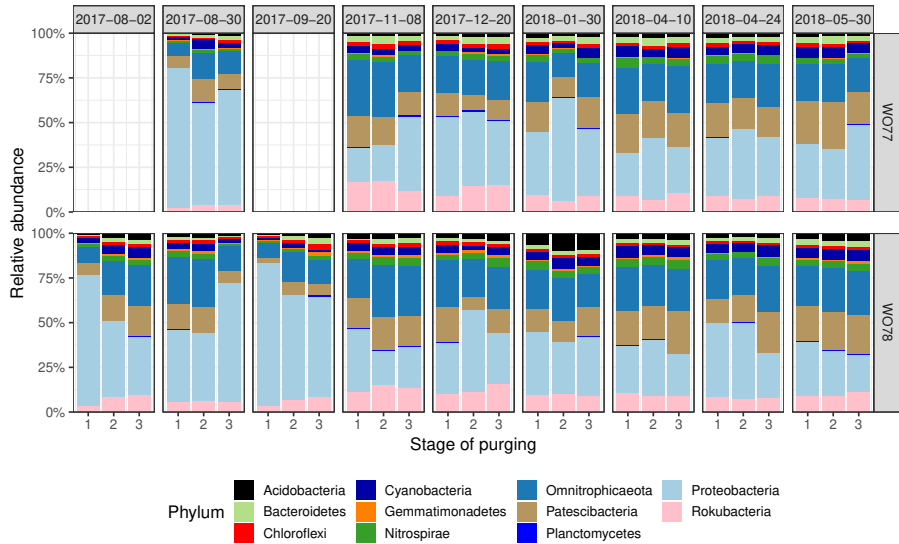


Figure D3: Summary of relative abundance of bacterial taxa on phylum level, core community only (present in at least 91% of samples).

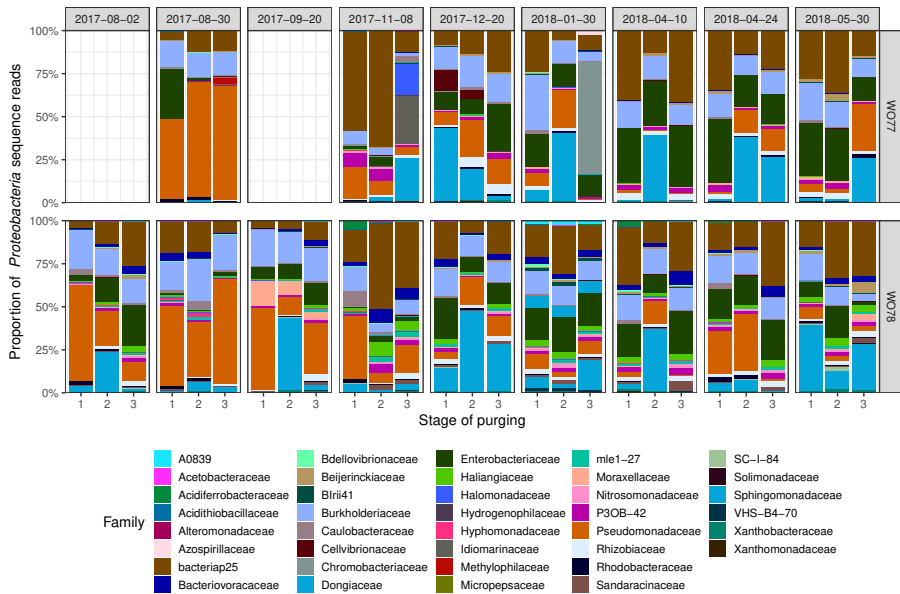


Figure D4: Summary of *Proteobacteria* taxa on family level.

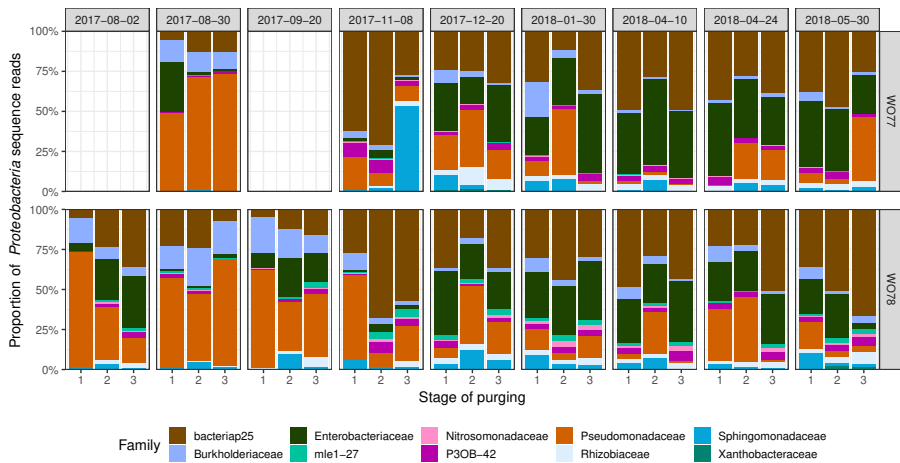


Figure D5: Summary of *Proteobacteria* taxa on family level, core community only (present in at least 91% of samples).

D5 Alpha-diversity measures associated with 16S rRNA gene amplicon sequencing samples

Table D2: Summary statistics for α -diversity measures

α -diversity measures	WO77			WO78		
	Mean	Median	Range	Mean	Median	Range
Number of OTUs	184	185	146-206	200	202	172-222
Chao1	194	194	159-220	206	207	183-237
Shannon	4.0	4.3	2.0-4.5	4.3	4.5	2.9-4.7

D6 Constrained analysis of principal coordinates

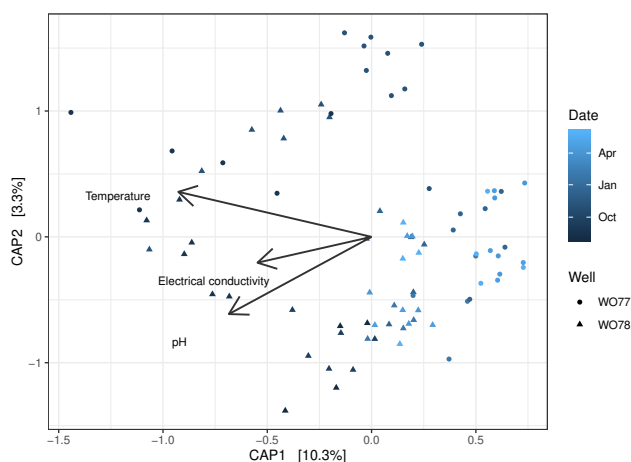


Figure D6: Constrained analysis of principal coordinates (CAP) ordination of purge water samples using Bray-Curtis dissimilarities. Circles and triangles represent the wells from which the samples were collected (WO77 and WO78, respectively). CAP axes 1 and 2 explain 10.4% and 3.3% of the microbial composition, respectively. The influence of environmental variables (temperature, pH, electrical conductivity) are indicated by arrows. Additional variables (i.e., dissolved oxygen, turbidity, ATP) did not yield improved estimates of microbial composition and hence have been omitted from the ordination. The influence of seasonal/sampling occasion to the variability of the samples is more prominent than within the original NMDS ordination.

D7 Identification of potentially relevant indicator taxa

Table D3: Top 20 OTUs ranked in order of decreasing significance of the well membership factor coefficient. Statistics presented in red indicate that the sequence reads affiliated with that OTU are not significantly different ($p>0.05$) from the levels present in the negative controls and therefore warrant cautious interpretation.

Phylum; Class; Order; Family; Genus	Base mean	log ₁₀ -fold difference	p-value
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>Bdellovibrionales</i> ; <i>Bacteriovoraceae</i> ; -	113.9	1.93	p<0.001
<i>Zizibacteria</i> ; -; -; -	120.8	2.38	p<0.001
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>NB1-j</i> ; -; -	79.4	2.61	p<0.001
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>Mycococcales</i> ; <i>Haliangiaceae</i> ; <i>Haliangium</i>	51.1	2.04	p<0.001
<i>Omnitrophicaeota</i> ; -; -; -	87.2	0.73	p<0.001
<i>Proteobacteria</i> ; -; -; -	5.5	-8.97	p<0.001
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>Mycococcales</i> ; <i>Sandamacinaceae</i> ; -	42.3	1.57	p<0.001
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>Mycococcales</i> ; <i>Haliangiaceae</i> ; <i>Haliangium</i>	32.1	2.46	p<0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>PLTA13</i> ; -; -	46.3	1.67	p<0.001
<i>Acidobacteria</i> ; <i>Holophagae</i> ; <i>Subgroup_7</i> ; -; -	27.3	1.43	p<0.001
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>Mycococcales</i> ; <i>mle1-27</i> ; -	50.2	1.10	p<0.001
<i>Bacteroidetes</i> ; <i>Bacteroidia</i> ; -; -; -	15.1	1.77	p<0.001
<i>Planctomycetes</i> ; <i>Phycisphaerae</i> ; <i>CCM11a</i> ; -; -	18.1	1.83	p<0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>PLTA13</i> ; -; -	28.3	0.85	p<0.001
<i>Gemmatimonadetes</i> ; <i>Gemmatimonadetes</i> ; <i>Gemmatimonadales</i> ; <i>Gemmatimonadaceae</i> ; -	38.7	0.74	p<0.001
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; -; -; -	14.7	1.93	p<0.001
<i>Omnitrophicaeota</i> ; -; -; -	41.4	-0.59	p<0.001
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Rhizobiales</i> ; <i>A0839</i> ; -	13.3	1.69	p<0.001
<i>Acidobacteria</i> ; <i>Acidobacteria</i> ; <i>Subgroup_2</i> ; -; -	21.5	0.80	p<0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; -; -; -	10.2	0.96	p<0.001

Table D4: Top 20 OTUs ranked in order of decreasing significance of the first purging stage factor coefficient (early, <20 minutes vs. intermediate, 20 minutes to 1 hour). Statistics presented in red indicate that the sequence reads affiliated with that OTU are not significantly different ($p>0.05$) from the levels present in the negative controls and therefore warrant cautious interpretation.

Phylum; Class; Order; Family; Genus	Base mean	log ₁₀ -fold difference	p-value
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>Acidiferrrobacteriales</i> ; <i>Acidiferrrobacteraceae</i> ; -	16.6	-0.95	p<0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Polaromonas</i>	15.6	-0.75	p<0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>Pseudomonadales</i> ; <i>Pseudomonadaceae</i> ;	76.6	0.87	p<0.001
<i>Pseudomonas</i>			
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Sphingomonadales</i> ; <i>Sphingomonadaceae</i> ;	127.7	0.83	p<0.001
<i>Sphingomonas</i>			
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Rhodoferrax</i>	52.3	-0.57	p<0.001
WPS-2; -; -; -	11.6	-0.68	p<0.001
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Caulobacterales</i> ; <i>Caulobacteraceae</i> ;	38.0	-0.63	p=0.001
<i>Phenglobacterium</i>			
<i>Actinobacteria</i> ; <i>Actinobacteria</i> ; <i>Corynebacteriales</i> ; <i>Mycobacteriaceae</i> ; <i>Mycobacterium</i>	11.7	-0.49	p=0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>Pseudomonadales</i> ; <i>Pseudomonadaceae</i> ; <i>Azotobacter</i>	23.5	0.78	p=0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Massilia</i>	126.7	-0.40	p=0.001
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Caulobacterales</i> ; <i>Hyphomonadaceae</i> ; <i>Hirschia</i>	16.0	-0.54	p=0.002
<i>Verrucomicrobia</i> ; <i>Verrucomicrobiae</i> ; <i>Verrucomicrobiales</i> ; <i>Verrucomicrobiaceae</i> ;	23.2	-0.61	p=0.007
<i>Prostheobacter</i>			
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ;	5.4	-0.69	p=0.008
<i>Herrniaimonas</i>			
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Rhizobiales</i> ; <i>Beijerinckiaceae</i> ; <i>Methylobacterium</i>	11.5	0.53	p=0.009
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Rhodobacteriales</i> ; <i>Rhodobacteraceae</i> ; <i>Rhodobacter</i>	6.1	0.69	p=0.012
<i>Verrucomicrobia</i> ; <i>Verrucomicrobiae</i> ; <i>Pedospherales</i> ; <i>Pedospheraceae</i> ; -	11.7	-0.55	p=0.014
<i>Firmicutes</i> ; <i>Bacilli</i> ; <i>Bacillales</i> ; <i>Bacillaceae</i> ; <i>Anoxybacillus</i>	660.7	0.67	p=0.015
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Methylobium</i>	59.0	-0.47	p=0.019
<i>Bacteroidetes</i> ; <i>Bacteroidia</i> ; <i>Cytophagales</i> ; <i>Spirosomaceae</i> ; <i>Arcicella</i>	12.1	-0.70	p=0.019
<i>Firmicutes</i> ; <i>Bacilli</i> ; <i>Bacillales</i> ; <i>Bacillaceae</i> ; <i>Bacillus</i>	5.9	0.57	p=0.019

Table D5: Top 20 OTUs ranked in order of decreasing significance of the second purging stage factor coefficient (intermediate, 20 minutes to 1 hour vs. late, 1 to 4 hours). Statistics presented in red indicate that the sequence reads affiliated with that OTU are not significantly different ($p>0.05$) from the levels present in the negative controls and therefore warrant cautious interpretation.

Phylum; Class; Order; Family; Genus	Base mean	log ₁₀ -fold difference	p-value
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Sphingomonadales</i> ; <i>Sphingomonadaceae</i> ;	1240.9	-0.86	p<0.001
<i>Novosphingobium</i>			
<i>Firmicutes</i> ; <i>Bacilli</i> ; <i>Bacillales</i> ; <i>Bacillaceae</i> ; <i>Anoxybacillus</i>	660.7	-0.87	p<0.001
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ;	28.8	-0.63	p=0.002
<i>Aquabacterium</i>			
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Ralstonia</i>	65.5	-0.77	p=0.004
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Rhizobiales</i> ; <i>Xanthobacteraceae</i> ; <i>Afpia</i>	16.7	-0.79	p=0.005
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Sphingomonadales</i> ; <i>Sphingomonadaceae</i> ;	41.0	0.28	p=0.007
<i>Sphingomonas</i>			
<i>Bacteroidetes</i> ; <i>Bacteroidia</i> ; <i>Chitinophagales</i> ; <i>Chitinophagaceae</i> ; <i>Asinibacterium</i>	57.8	-0.69	p=0.011
<i>Firmicutes</i> ; <i>Bacilli</i> ; <i>Bacillales</i> ; <i>Bacillaceae</i> ; <i>Geobacillus</i>	600.3	-0.69	p=0.014
<i>Firmicutes</i> ; <i>Bacilli</i> ; <i>Bacillales</i> ; <i>Paenibacillaceae</i> ; <i>Brevibacillus</i>	1326.4	-0.65	p=0.014
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Roseateles</i>	365.3	-0.32	p=0.037
<i>Firmicutes</i> ; <i>Bacilli</i> ; <i>Bacillales</i> ; <i>Bacillaceae</i> ; <i>Bacillus</i>	65.5	-0.60	p=0.046
<i>Omnitrophicaeota</i> ; -; -; -; -	72.5	0.10	p=0.049
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>Xanthomonadales</i> ; <i>Xanthomonadaceae</i> ;	9.9	-0.08	p=0.049
<i>Stenotrophomonas</i>			
<i>Deinococcus-Thermus</i> ; <i>Deinococci</i> ; <i>Deinococcales</i> ; <i>Deinococcaceae</i> ; <i>Deinococcus</i>	80.9	-0.62	p=0.049
<i>Planctomycetes</i> ; <i>Pla4-lineage</i> ; -; -; -	3.5	0.81	p=0.056
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Rhizobiales</i> ; <i>Beijerinckiaceae</i> ; <i>Methylobacterium</i>	11.5	-0.39	p=0.060
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>Mycrococcales</i> ; <i>bacteriap25</i> ; -	140.5	0.20	p=0.062
<i>Planctomycetes</i> ; <i>Phycisphaerae</i> ; <i>CCM11a</i> ; -;	36.2	0.18	p=0.073
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Duganella</i>	5.8	0.58	p=0.092
<i>Proteobacteria</i> ; δ - <i>proteobacteria</i> ; <i>RCP2-54</i> ; -; -	31.1	0.14	p=0.094

Table D6: Top 20 OTUs ranked in order of decreasing significance of the first seasonal factor coefficient (baseline summer vs. transitional winter period). Statistics presented in red indicate that the sequence reads affiliated with that OTU are not significantly different ($p > 0.05$) from the levels present in the negative controls and therefore warrant cautious interpretation.

Phylum; Class; Order; Family; Genus	Base mean	log ₁₀ -fold difference	<i>p</i> -value
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Massilia</i>	126.7	-1.21	$p < 0.001$
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Roseateles</i>	5.5	0.57	$p < 0.001$
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Roseateles</i>	365.3	-1.18	$p < 0.001$
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Polaromonas</i>	9.7	-1.40	$p < 0.001$
<i>WPS-2</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Polaromonas</i>	11.6	-1.50	$p < 0.001$
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Methylobium</i>	59.0	-1.23	$p < 0.001$
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Rhodobacteriales</i> ; <i>Rhodobacteraceae</i> ; <i>Pseudorhodobacter</i>	24.1	-1.49	$p < 0.001$
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Caulobacteriales</i> ; <i>Hyphomonadaceae</i> ; <i>Hirschia</i>	16.0	-1.04	$p < 0.001$
<i>Patescibacteria</i> ; <i>Gracilbacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Noviherbaspirillum</i>	71.9	0.38	$p < 0.001$
<i>GALI5</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Noviherbaspirillum</i>	42.7	0.69	$p < 0.001$
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Undibacterium</i>	7.2	-1.16	$p < 0.001$
<i>Omnitrophicaeota</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Undibacterium</i>	20.0	-1.05	$p < 0.001$
<i>Omnitrophicaeota</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Burkholderiaceae</i> ; <i>Undibacterium</i>	70.3	0.15	$p < 0.001$
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>Pseudomonadales</i> ; <i>Pseudomonadaceae</i> ; <i>Pseudomonas</i>	2568.8	-0.88	$p < 0.001$
<i>Proteobacteria</i> ; γ - <i>proteobacteria</i> ; <i>Pseudomonadales</i> ; <i>Moraxellaceae</i> ; <i>Acinetobacter</i>	131.1	-0.91	$p < 0.001$
<i>Rokubacteria</i> ; <i>NC10</i> ; <i>Rokubacteriales</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Undibacterium</i>	175.4	0.19	$p < 0.001$
<i>Rokubacteria</i> ; <i>NC10</i> ; <i>Rokubacteriales</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Undibacterium</i>	360.6	0.19	$p < 0.001$
<i>Rokubacteria</i> ; <i>NC10</i> ; <i>Rokubacteriales</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Undibacterium</i>	49.0	0.30	$p < 0.001$
<i>Rokubacteria</i> ; <i>NC10</i> ; <i>Rokubacteriales</i> ; γ - <i>proteobacteria</i> ; β - <i>proteobacteriales</i> ; <i>Undibacterium</i>	361.9	0.19	$p < 0.001$
<i>Proteobacteria</i> ; α - <i>proteobacteria</i> ; <i>Sphingomonadales</i> ; <i>Sphingomonadaceae</i> ; <i>Novosphingobium</i>	7.1	-1.26	$p < 0.001$

Table D7: Top 20 OTUs ranked in order of decreasing significance of the second seasonal factor coefficient (transitional winter vs. spring freshet period). Statistics presented in red indicate that the sequence reads affiliated with that OTU are not significantly different ($p>0.05$) from the levels present in the negative controls and therefore warrant cautious interpretation.

Phylum; Class; Order; Family; Genus	Base mean	log ₁₀ -fold difference	p-value
<i>Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; -</i>	5.9	-0.57	p<0.001
<i>Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella_7</i>	3.5	-0.38	p<0.001
<i>Firmicutes; Bacilli; Bacillales; Alicyclobacillaceae; Alicyclobacillus</i>	5.1	-0.32	p<0.001
<i>Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides</i>	5.1	-0.38	p<0.001
<i>Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides</i>	3.9	-0.30	p<0.001
<i>Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; -</i>	4.0	-0.31	p<0.001
<i>Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Anaerostipes</i>	3.4	-0.23	p<0.001
<i>Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; -</i>	3.3	-0.23	p<0.001
<i>Lachnospiraceae_NK4A136-group</i>	3.8	-0.19	p<0.001
<i>Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides</i>	13.2	1.82	p<0.001
<i>Proteobacteria; -; -; -</i>	80.0	0.34	p<0.001
<i>Bacteroidetes; Bacteroidia; Sphingobacteriales; -</i>	3.8	-0.04	p<0.001
<i>Firmicutes; Clostridia; Clostridiales; Family_XVII; Sulfobacillus</i>	20.5	-0.63	p<0.001
<i>Proteobacteria; γ-proteobacteria; -; -; -</i>	89.8	0.14	p<0.001
<i>Omnitrophicaota; -; -; -; -</i>	129.4	-0.15	p<0.001
<i>Omnitrophicaota; -; -; -; -</i>	11.7	-0.95	p<0.001
<i>Verrucomicrobia; Verrucomicrobiae; Pedosphaerales; Pedosphaeraceae; -</i>	361.9	-0.19	p<0.001
<i>Rokubacteria; NC10; Rokubacteriales; -; -</i>	8.5	-1.56	p<0.001
<i>Proteobacteria; γ-proteobacteria; β-proteobacteriales; Chromobacteriaceae; Paludibacterium</i>	15.1	-0.54	p<0.001
<i>Bacteroidetes; Bacteroidia; -; -; -</i>	12.1	-0.66	p<0.001
<i>Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Lachnospiraceae-AC2044-group</i>			

List of Publications

PEER REVIEWED JOURNAL ARTICLES

Chik, A.H.S., Schmidt, P.J., and Emelko, M.B. (2018) Learning Something From Nothing: The Critical Importance of Rethinking Microbial Non-detects. *Front. Microbiol.* 9:2304. doi: 10.3389/fmicb.2018.02304

Chik, A.H.S., Emelko, M.B., Blaschke, A.P., and Schijven, J.F. (2020) Illuminating Subsurface Microbial Water Quality Patterns Using Adenosine Triphosphate and Dynamic Time Warping Approaches. *Groundwater Monitoring and Remediation* doi: 10.1111/gwmmr.12397

(AS CO-AUTHOR)

van Driezum, I.H., Chik, A.H.S., Jakwerth, S., Lindner, G., Farnleitner, A.H., Sommer, R., Blaschke, A.P., and Kirschner, A.K.T. (2018) Spatiotemporal Analysis of Bacterial Biomass and Activity to Understand Surface and Groundwater Interactions in a Highly Dynamic Riverbank Filtration System. *Science of the Total Environment*, 627:450-461 doi: 10.1016/j.scitotenv.2018.01.226

JOURNAL ARTICLES IN REVIEW

Chik, A.H.S., Emelko, M.B., Anderson, W.B., O'Sullivan, K.E., Savio, D., Farnleitner, A.H., Blaschke, A.P., and Schijven, J.F. (2020) Evaluation of Groundwater Microbial Community Composition to Inform Waterborne Pathogen Vulnerability Assessments. *Science of the Total Environment*

JOURNAL ARTICLES IN PREPARATION

Chik, A.H.S., Schmidt, P.J., Emelko, M.B., Schijven, J.F., and Blaschke, A.P. (2020) An End Justified by Means: Ensuring Adequate Treatment of Protozoa by Precise Mean Concentration Estimation. *Journal of Water and Health.*

ABSTRACT-REVIEWED PRESENTATIONS & PROCEEDINGS

Chik, A.H.S., Anderson, W.B., Blaschke, A.P., Emelko, M.B., Schijven, J.F., and Schmidt, P.J. (2021) Perspectives, Insights, and Outlook Concerning Microbial Water Quality Evaluations for the Water Industry. *Ontario's Water Conference & Trade Show 2021*, University Research Session, Niagara Falls, ON, Canada, TBD.

Chik, A.H.S., Schmidt, P.J., Emelko, M.B., Ruecker, N.J., Schijven, J.F., and Blaschke A.P. (2020) Enough Already: Determining Number of Samples Needed to Precisely Evaluate Mean Protozoan Pathogen Concentrations. International Society for Subsurface Microbiology Conference, Utrecht, The Netherlands, TBD.

Chik, A.H.S., Savio, D., Farnleitner, A.H., Blaschke, A.P., Schijven, J.F., Anderson, W.B., and Emelko, M.B. (2019) Informing Groundwater Pathogen Risks Using 16S rRNA Microbial Community Analysis. AWWA's Water Quality Technology Conference & Exposition (WQTC), Dallas, Texas, USA, November 3 [Student poster session].

Chik, A.H.S., Schmidt, P.J., and Emelko, M.B. (2019) Zero, Zilch, Nada: Unadulterated Microbial Non-detects Prevent Bias. IWA's 20th Symposium on Health-related Water Microbiology (HRWM), Vienna, Austria, September 15-20 [poster].

Chik, A.H.S., Schmidt, P.J., and Emelko, M.B. (2018) Microbial Non-detects: How They Should Be Reported and Why It Matters. 2018 National Water and Wastewater Conference, Nov 4-7, 2018, Montreal, Quebec, Canada.

Chik, A.H.S., Schmidt, P.J., and Emelko, M.B. (2018) Demystifying Microbial Non-detects: Raw and Uncensored. Water Quality and Technology Conference Conference 2018, Nov 11-15, 2018, Toronto, Ontario, Canada.

Chik, A.H.S., Beelen, B., O'Sullivan, K.E., Wong, C., Anderson, W.B., Mesquita, M.F., and Emelko, M.B. (2018) ATP as a Rapid Sentinel of Microbial Water Quality Change in Groundwater Supplies. Water Quality and Technology Conference Conference 2018, Poster session, Toronto, Nov 11-15, 2018 , Ontario, Canada.

(AS CO-AUTHOR)

Emelko, M.B., Chik, A.H.S., Anderson, W.B., Spanjers, M.G., and Thompson, J. (2020) Highlighting the Value of Integrated Monitoring Strategies for Process Evaluation and Optimization: Case Studies From Classical and Subsurface Biofiltration. AWWA's 2020 International Symposium on Biological Treatment, Atlanta, Georgia, USA, February 12-13.

Schmidt, P.J., Ruecker, N.J., Kundert, K.L., Chik, A.H.S., and Emelko, M.B. (2019) Who Needs a Framework for Ensuring Adequate Treatment of *Giardia* Anyways? Calgary's Approach to Addressing Enteric Protozoa. AWWA's Water Quality Technology Conference & Exposition (WQTC), Dallas, Texas, USA, November 6, WED03-03 [Pathogens: Sources, Characterization, and Risk session].

Schmidt, P.J., Ruecker, N.J., Kundert, K.L., Chik, A.H.S., and Emelko, M.B. (2019) What About *Giardia*? Exploring Monitoring and Regulatory Frameworks When a City's Raw Water Source has More *Giardia* than *Cryptosporidium*. IWA's 20th Symposium on Health-related Water Microbiology (HRWM), Vienna, Austria, September 15-20 [poster].

Beelen, B., Chik, A.H.S., Wong, C., Mesquita, M.F., Anderson, W.B., and Emelko, M.B. (2018) Field Evaluation of Adenosine Triphosphate as an Indicator of Microbiological Water Quality Change in Subsurface Supplies. *Ontario's Water Conference & Trade Show 2018*, University Research Session, April 29-May 2, 2018, Niagara Falls, ON, Canada.

INVITED TALKS

Chik, A.H.S., Schmidt, P.J., and Emelko, M.B. (2018) Avoid bias: A Zero by Any Other Name is Still a Zero. Invitation by Ontario Ministry of Environment, Conservation and Parks, Toronto, Ontario, Canada, Nov. 27.