

ANALYZING FACTORS WHICH AFFECT *LEGIONELLA* OCCURRENCE IN A FULL-
SCALE GREEN BUILDING PREMISE PLUMBING SYSTEM

By

Ryan Julien

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biosystems Engineering – Doctor of Philosophy
Environmental Science and Policy – Dual Major

2021

ABSTRACT

ANALYZING FACTORS WHICH AFFECT *LEGIONELLA* OCCURRENCE IN A FULL-SCALE GREEN BUILDING PREMISE PLUMBING SYSTEM

By

Ryan Julien

Water consumption in the United States has decreased in recent decades. However, plumbing design guidance has not been updated to reflect this change, resulting in increased hydraulic retention time, disinfectant decay, and the proliferation of opportunistic premise plumbing pathogens (OPPPs) such as *Legionella pneumophila*. Time spent in premise plumbing systems has been shown to impact water quality through such mechanisms as the loss of residual disinfectant, leaching of pipe materials, biofilm formation, and increased concentrations of opportunistic pathogens such as *Legionella* spp.

Quantitative Microbial Risk Assessment (QMRA) is a tool used to evaluate human health risks, and has been used to assess risks associated with *Legionella*. However, these assessments require data regarding the concentration of *Legionella* in water. Due to the ubiquity of *Legionella* in plumbing systems, their growth in biofilms, and the sporadic nature of biofilm detachment, *Legionella* concentrations are poorly understood, thus limiting the utility of QMRA in this instance. Factors which influence the prevalence of *Legionella* have been studied at the bench scale, but never in a full-scale building water system. The work presented herein takes a risk factor approach in exploring how to better monitor or predict concentrations of *Legionella* spp.

This dissertation presents research to help better understand factors which best predict *Legionella* spp. Research objectives of this work were to: (1) identify variables which most effectively predict *Legionella* spp. concentrations, (2) determine the time water spends stored in building plumbing using a novel model, and (3) determine whether compliance with common

temperature guidelines to limit *Legionella* proliferation have a significant impact on concentrations. This research employs a rich data set from a full-scale home, equipped with flowmeters and temperature sensors to assess water conditions. Analytical samples were also collected to determine common water quality variables, as well as enumeration of *Legionella* spp. Multiple statistical analyses were used to investigate variable relationships and to evaluate the value of model results in predicting *Legionella* spp. concentrations.

Principal component analysis suggests that water age and biofilm detachment are the primary drivers of changes observed in water quality, accounting for 53% of the total variance in the data. General linear modeling revealed that heterotrophic plate count, total organic carbon, total cell count, maxTSL and meanTSL, and modeled water age were significant predictors of *Legionella* spp. concentrations. Bayesian variable selection indicated that the 95th percentile of water age and maxTSL were most predictive of *Legionella* spp. concentrations. Results from the water age model were evaluated, indicating that modeled water age is a statistically significant predictor of *Legionella* spp. Compliance with temperature guidelines was found to be significantly correlated to *Legionella* spp.

Results of this research indicate that water quality and use have significant implications to *Legionella* occurrence. Results also provide a framework to investigate *Legionella* spp. using variables which are more commonly and cheaply measured than direct measurement, potentially leading to more widespread monitoring for *Legionella* and reducing cases. These results show that water age is a critical factor in determining *Legionella* spp. prevalence. This knowledge should be applied to plumbing design and maintenance to limit water age and thereby *Legionella* spp. concentrations.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude for the many people who supported me in achieving this degree. First, I'd like to thank Dr. Jade Mitchell for her guidance, encouragement, and compassion. The example she has set has been motivating and grounding. I would also like to thank my committee; Dr. Pouyan Nejadhashemi, Dr. Joan Rose, and Dr. Andy Whelton; for their expertise, guidance, and support in developing this research.

I would also like to thank my lab mate Kara, as well as the other Biosystems Engineering graduate students, for all their supportive conversations, occasional commiseration, and assistance navigating graduate school requirements.

I also want to thank my family and friends, especially my parents, Rob and Kathleen, for fostering my curiosity, help in shaping my personal values, and their loving support. Finally, I'd like to thank my spouse Ashley and my daughter Elliot for their support, affection, understanding, and unwavering belief in me.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
1 Chapter 1 – Introduction.....	1
1.1 Opportunistic Premise Plumbing Pathogens and <i>Legionella</i> spp.	1
1.2 Premise Plumbing Factors.....	5
1.3 Contemporaneous Factors	7
1.4 Modeling Water Age.....	10
1.5 Assessing Risks with QMRA.....	11
2 Chapter 2 – Objectives	14
3 Chapter 3 – Introduction to Methodological Approach and Exploratory Data Analysis	18
3.1 Primary Data Source	18
3.1.1 Analytical Data	20
3.1.2 Electronically Recorded Data	23
3.1.3 Development of water use metrics.....	23
3.2 Statistical Analyses Conducted	26
3.2.1 Spearman’s Rank Correlation	27
3.2.2 Principal Component Analysis	27
3.2.3 Generalized Linear Modeling and Generalized Linear Mixed Modeling.....	28
3.2.4 Bayesian Generalized Linear Regression	30
4 Chapter 4 – Variable Selection.....	31
4.1 Introduction	31
4.2 Methods.....	31
4.3 Results	32
4.3.1 Correlation Coefficients.....	32
4.3.2 Principal Component Analysis	34
4.3.3 General Linear Model	35
4.3.4 Bayesian Variable Selection Method.....	37
4.4 Discussion	38
4.4.1 Flushing, Stagnation, and Water Age	39
4.4.2 Fate of Residual Disinfectant.....	40
4.4.3 Microbial Contaminants.....	42
4.4.4 Research Objectives.....	44
4.4.5 Limitations of this Study.....	45
5 Chapter 5 – Water Age Modeling.....	47
5.1 Introduction	47
5.2 Methods.....	48
5.2.1 Water Age Model Development Process	49

5.2.2	Comparison with EPANET.....	51
5.2.3	Variable Selection.....	52
5.3	Results.....	53
5.3.1	Water Age.....	53
5.3.2	Variable Selection Including Water Age.....	57
5.3.3	Comparison with EPANET.....	63
5.4	Discussion.....	64
5.4.1	Water Age Results.....	64
5.4.2	Comparison with EPANET.....	65
5.4.3	Variable Selection.....	66
5.4.4	Conclusions.....	68
5.4.5	Limitations.....	68
6	Chapter 6 – Assessing Compliance with Thermal Guidance.....	70
6.1	Introduction.....	70
6.2	Methods.....	71
6.3	Results.....	73
6.4	Discussion.....	75
7	Chapter 7 – Conclusions.....	77
7.1	Implications for Quantitative Microbial Risk Analysis.....	77
7.2	Limitations of These Studies.....	78
8	Chapter 8 – Future Research.....	81
	APPENDIX.....	84
	REFERENCES.....	90

LIST OF TABLES

Table 1 - Sample Collection Locations.....	21
Table 2 - Monitored Variables	22
Table 3 - Spearman's rank correlation coefficients.....	33
Table 4 - Principal Component Analysis Results	35
Table 5 - Comparison of Generalized Linear Mixed Models m.top and m.comp	37
Table 6 - BLGR Variable Selection Results (BayesC priors).....	38
Table 7 - PCA results for PC1 through PC3	59
Table 8 - GLMM Results for m.top	62
Table 9 - BGLR Results.....	63

LIST OF FIGURES

Figure 1 - ReNEWW house piping and instrumentation diagram. Pipe diameter is depicted with line width, flowmeters with squares, sample locations in shaded rectangles, and the approximate location of thermistors shown in numbered circles.	20
Figure 2 - Absolute value of Spearman correlation coefficients between water quality variables and water use metrics, calculated using varying time periods from 1 to 120 days prior to sample collection. The red vertical line represents the time period selected for analysis, 14 days.	25
Figure 3 - Boxplot of water age results.....	54
Figure 4 - Scatterplot comparison of the water age metric age.mean with the water use metric meanTSL. The color of each point indicates the location from which each sample was collected, as shown in the legend.	55
Figure 5 - Boxplots comparing the water age metric age.mean (left) with the use metric meanTSL (right).....	56
Figure 6 - Histogram of water age results by fixture	57
Figure 7 - Boxplot comparison of water age results for both models.....	64
Figure 8 - ReNEWW P&ID with temperature zone information. Each pipe is labeled with a number indicating the temperature zone, with each number corresponding to the thermistors depicted in Figure 1.	72
Figure 9 – Mean temperature compliance by fixture from 9/1/2017 through 11/1/2018	74
Figure 10 - Scatterplot of mean temperature compliance during two-week period preceding each sample and measured Legionella spp. concentration. The color of each point represents the location from which that sample was collected.	75
Figure 11 - Raw instantaneous flowrate at hot.kitchen.isl.....	85
Figure 12 - Raw instantaneous flowrate at hot.bath2.sink.....	86
Figure 13 - Example data showing the impact of flowrate on temperature on 5/4/2017.....	87
Figure 14 - Example data showing flowmeter noise with no influence on water temperature on 9/5/2016. The design flowrate of hot.kitchen.isl is shown as a horizontal red line.....	88

1 Chapter 1 – Introduction

1.1 Opportunistic Premise Plumbing Pathogens and *Legionella* spp.

Opportunistic Premise Plumbing Pathogens (OPPPs) are a group of waterborne microorganisms, including *Legionella* species (spp.), *Mycobacterium avium* and other non-tuberculosis Mycobacteria, and *Pseudomonas aeruginosa*. OPPPs are commonly found in water, air, and soil around the world. OPPPs are commonly identified in both WDSs and PPSs, and are not correlated to indicator organisms such as *E. coli*. These differ from classical waterborne pathogens in that they are naturally-occurring and are selected for within plumbing environments, thus increasing their numbers and virulence.¹⁻⁵ Further, these pathogens target especially susceptible hosts, such as the immunocompromised and elderly.^{1,6,7} Due to their ubiquitous presence, OPPPs are considered native to PPS environments and eliminating OPPPs from plumbing is not generally feasible. OPPP abundance in drinking water is related to conditions of the pipe environment rather than an indication of contamination.⁸

OPPPs are specifically adapted to survival in drinking water systems. A review of current literature reveals key adaptations that allow OPPPs to flourish and gain a selective advantage within premise plumbing.^{8,9} Several adaptations were identified, some of which are broadly applicable while others are specific to only a single OPPP. Key adaptations that have been identified in multiple OPPP species are summarized in the following paragraphs to give a broad sense of the complex interactions between premise plumbing features and impacts to OPPP growth.

Resistance to disinfectant: Disinfection is effective in inactivating many microorganisms, but OPPPs share a relative resistance to disinfectants commonly used in drinking water, thus selecting for them in the pipe environment.^{1,6} For example, *Legionella*

pneumophila, *Mycobacterium avium*, and *Pseudomonas aeruginosa* are 83, 567, and 21 times more resistant to chlorine than *E. coli*, respectively.⁸

Proliferation with limited oxygen and carbon: OPPPs are relatively slow growing compared to many other microbial inhabitants of premise plumbing. This slow growth rate allows OPPPs to survive in lower concentrations of carbon and oxygen that would not support other common microbial inhabitants of plumbing.⁶ *Mycobacteria avium* has been show to proliferate in concentrations of assimilable organic carbon as low as 50 µg/L,¹⁰ whereas *E. coli* appears to be growth-limited in concentrations an order of magnitude higher.¹¹

Persistence through phagocytosis: Several OPPPs are resistant to phagocytic killing by free-living amoeba. Phagocytosis, the process by which amoeba engulf materials, is used by amoeba to consume bacteria and other nutrients for use in the cell. However, OPPPs such as *M. avium* and *Legionella* spp. are resistant to killing in this way and have been observed multiplying within amoeba instead. Living within amoeba following phagocytosis has been demonstrated to protect *Legionella* spp. from elevated temperatures and concentrations of disinfectant⁷ as well as increase virulence in *L. pneumophila* and *M. avium*.^{1,6,12}

Growth in biofilms: Microorganisms tend to adhere to inner surfaces of water systems to form biofilm.^{13,14} Wingender and Flemming¹⁵ reported that roughly 95% of microbial growth in plumbing is on surface of pipe with only 5% contained in the bulk-phase water. As such, biofilms are often viewed as a reservoir of OPPPs and a source of contamination for downstream plumbing. Organisms in biofilm secrete extracellular polymeric substances (EPS) that assist in aggregating cells and preventing washout from the plumbing environment.¹⁵ EPS in biofilm has also been shown to shield microbes from environmental hazards such as disinfection compounds,^{16,17} nutrient deficiencies,¹⁸ and thermal shock.¹⁹ Biofilm environments are also

thought to promote horizontal gene transfer of antibiotic resistance and pathogen virulence.^{7,20} Biofilms serve as a source of food and nutrients for OPPPs. *Legionella pneumophila* has been shown to thrive in the presence of dead cells commonly found in biofilm.²¹ Growth in biofilm has been shown to promote phagocytosis by amoeba, which enhances OPPP growth rates as well as virulence.^{1,7} Biofilm becomes detached from pipe walls, leading to suspension into the bulk water and washout. Literature suggests that biofilm detachment occurs for two primary reasons; cellular erosion driven by increased shear stress of moving water in the pipes, and from large-scale sloughing, driven by structural failure of biofilm.^{22–24} Sloughing often occurs after cell die off in lower layers of biofilm leading to structural failure, which may be driven by changes in pH or dissolved oxygen.^{23–25}

Multiple exposure routes: OPPPs can initiate infections via multiple exposure routes, including inhalation of aerosols, as well as ocular and dermal exposures. However, enteric pathogens, which are generally removed or inactivated during drinking water treatment, almost exclusively initiate infection via ingestion. This means that OPPPs may present hazards during activities like bathing, due to the inhalation of aerosols, where enteric pathogens would not.^{7,26}

Legionella spp. are the most common OPPPs found in potable water, and are now the leading cause of waterborne infection in the United States.^{1,27} The scope of this dissertation is focused primarily on the role of *Legionella* spp. in PPSs and resulting health implications, though other OPPPs will be discussed for context and comparison throughout. *Legionella* is a genus of rod-shaped, gram-negative bacteria with over 50 individual species, and approximately 70 different serogroups.^{1,28} While *L. pneumophila* is the *Legionella* species most commonly associated with human disease,¹ roughly 25 other species such as *L. longbeachae* and *L. micdadei* are pathogenic and have been identified in potable water systems.²⁹ *Legionella*

pneumophila was the first species to be discovered, and was identified following the an outbreak at the 1976 American Legion convention in Philadelphia, PA. Of the roughly 2,000 attendees to the convention, 182 are known to have developed a *Legionella* infection, also referred to as Legionellosis, with 147 cases requiring hospitalization, and 29 cases resulting in death.³⁰ While other species and serogroups have been shown to cause disease, *Legionella pneumophila* serogroup 1 remains the primary etiological agent of Legionellosis.^{1,27} Multiple species of *Legionella* are not typically analyzed from the same samples, leaving a knowledge gap regarding the typical distribution of other pathogenic species.³¹ Risk characterization of *Legionella* spp. using only *L. pneumophila* is likely to underestimate risk, as the presence of other, potentially pathogenic, *Legionella* species are not considered. More appropriate, and likely conservative, risk estimates may be produced by conducting analysis for *Legionella* spp. and then treating those results as *L. pneumophila*, the most pathogenic species of *Legionella*.

Legionellae can cause infections, especially in those with immunodeficiencies or other risk factors such as advanced age.⁶ *Legionella* spp. typically infect humans after exposure to *Legionella*-containing aerosols. Common domestic water uses such as showering, toilet flushing, humidifiers, and hot tubs can produce these aerosols. When these aerosols are delivered to the lungs, *Legionellae* have been known to survive phagocytosis by pulmonary macrophages, allowing them to replicate in the lungs and causing to human disease.

Legionellosis can be subdivided into two primary diseases; Legionnaires' disease and Pontiac fever. Legionnaires' disease is a type of pneumonia, which, like other respiratory infections, is commonly associated with cough, shortness of breath, and fever. However, unlike common respiratory infections, Legionnaires' disease is also associated with gastrointestinal and neurological dysfunction, and mortality rates, estimated between 2.9 and 33 percent.¹ Pontiac

fever is a more mild illness which does not cause pneumonia and is non-fatal. Legionellosis has become the most common reportable waterborne disease in the United States, with 6,079 confirmed cases of Legionnaires' disease in 2015.³² However, the exposure source for greater than 95% of infections is never identified,¹ suggesting that the number of legionellosis cases is vastly undercounted.

1.2 Premise Plumbing Factors

Drinking water treatment in the United States has historically focused on eliminating contamination, whether biological or chemical. Tests for fecal bacteria, such as *Escherichia coli*, are used to indicate biological contamination. Most of these monitored pathogens need a mammalian host to reproduce,⁸ so it is often implicitly assumed that concentrations do not increase with time in piping. Likewise, concentrations of chemical contaminants do not increase in plumbing without a leak or source of contamination. As a result, relatively little attention has been focused on HRT, especially in PPSs, as it relates to water quality.

Except for *Legionella*, which has a maximum contaminant goal (MCLG) of zero, federal statutes do not regulate OPP concentrations in drinking water.³³ Further, *Legionella*'s MCLG is non-enforceable and does not require routine monitoring to ensure compliance. Community water systems, which supply water for more than 95% of people in the US,³⁴ may therefore serve as a reservoir of *Legionella* spp. and other such pathogens with potential to contaminate downstream building plumbing. As such, it is not practical to eliminate these organisms from plumbing environments. Instead, focus must be placed on managing their populations and limiting human exposure.⁷ Policy in the United States dictates monitoring standards to ensure safe drinking water within WDSs, but these standards do not apply to PPSs which are downstream of the property line and/or within buildings. A notable exception to this is the Lead

and Copper Rule which contains standards to limit leaching of lead materials within premise plumbing by limiting the corrosivity of water in WDSs.⁴ Existing bulk-water monitoring requirements do not adequately address risks related to opportunistic pathogens, such as *Legionella* spp., which do not correspond to fecal indicator tests.³ Further, there are no commonly agreed-upon guidelines to limit *Legionella* spp. exposure in the United States.³⁵

It is well established that the water quality varies as travels from the water distribution system point of entry to building faucets. These changes occur via a variety of mechanisms such as decay of residual disinfectant, plumbing material leaching, and interactions between bulk water, biofilms, and scales.^{1,4,5} These processes are primarily time-dependent meaning that water age, or the time water spends in contact with plumbing, is a key determinant of water quality degradation.^{4,14} While water conservation efforts have been effective in reducing water consumption, existing plumbing and design guidelines have not been updated to reflect changing demand.³⁶ This has led to increased water age^{1,37} which has been linked to degraded water quality.^{1,4,5,14}

Degradation of water quality is accelerated in building plumbing relative to service lines due to key differences between the structure and use of the plumbing:

Pipe diameter and relative surface area: Building plumbing is generally constructed of smaller-diameter pipes than WDSs, which leads to an increased surface area to volume ratio.⁴ It has been estimated that PPSs contain ten times more surface area per unit volume than WDSs.⁴ Pipe material leaching and area available for biofilm development are functions of wetted surface area in pipes.¹

Elevated temperature: Buried pipes, especially in areas like the northern United States, are typically well-insulated by soil and maintain a relatively constant cool temperature. As water

is delivered to building plumbing it is heated by the ambient temperature of the building.⁴ Increasing the water temperature is understood to boost chemical and biological rates of reaction, which can cause adverse effects such as accelerated loss of residual disinfectant, additional pathogen growth, and accelerated leaching rates by affecting the solubility of plumbing materials.^{5,14,38} Water heaters have the potential to destroy pathogens such as *Legionella* in heated water when temperatures are consistently maintained above 60°C.^{1,39} However, when lower setpoints are used, or if the water temperature in the heated portion of the plumbing frequently cools to below 55°C, this added heat can encourage additional pathogen growth.^{1,9,40}

Intermittent water use patterns: Intermittent water use in buildings can cause stagnation in parts or all of the building plumbing between uses, causing both a greater mean HRT in general leading to more biofilm growth, as well as intermittent high-velocity events that encourage the erosion or detachment of biofilm into bulk water.^{4,41–43} Many OPPPs are particularly well-suited to initiate biofilm formation. For example, *Mycobacteria* spp. are exceptionally hydrophobic enabling them to better adhere to pipe walls have thus been considered biofilm “pioneers”,⁴⁴ and are thus relatively selected for in premise plumbing.

Variable plumbing materials: A wider array of materials are often used in building plumbing than in typical WDSs. Using dissimilar metallic materials in plumbing accelerates galvanic corrosion,⁴ which may in turn provide additional habitat for *L. pneumophila*.⁴⁵ Additionally, certain elastomeric materials have been shown to support additional microbial growth.⁴⁶

1.3 Contemporaneous Factors

Water conservation efforts have significantly reduced the rate of water consumption in the United States. The United States Geological Survey (USGS) has reported total water

withdrawals have decreased from 42.1 billion gallons (159 billion L) per day in 2010 to 39.2 billion gallons (148 billion L) per day in 2015, a decrease of seven percent.⁴⁷ During this period the US population increased by four percent, and the proportion of the population using public-supply water systems increased from 86% to 87%. Despite this growth, the national average for domestic water use declined from 88 GPCD in 2010 to 82 GPCD in 2015.⁴⁷ This USGS data suggests that total water withdrawals have been decreasing since approximately 1980. Water withdrawals in the United States are currently at their lowest point since 1965 despite continued increases in population.⁴⁸ Residential water demand decreased 22% per household from 1999-2016, and 73% of this change can be attributed to efficiency increases of toilets and showerheads.⁴⁹

Existing water distribution infrastructure in the United States has been designed to accommodate higher flows than currently experienced. The Hunter Fixture Unit Method was developed in 1940 and remains widely used to determine pipe sizing for water mains and building distribution systems. However, this method is outdated and consistently overestimates water flows and resulting pipe size.⁴⁹ Further, plumbing design guidance from state and local ordinances as well as design codes such as the Uniform Plumbing Code and the International Plumbing Code, have required high-efficiency water fixtures, leading to additional reductions in water use. At the same time, these plumbing codes have not addressed the decrease in demand.⁵⁰ These factors have led to oversized plumbing in both WDSs and PPSs, and further contribute to increases in HRT.

Much of the water infrastructure in the United States is now nearing the end of its expected lifespan.^{51,52} The American Society of Civil Engineers (ASCE) estimates that 6 billion gallons (23 billion L) of treated drinking water are lost each day to leaking infrastructure.⁵³ The

AWWA has estimated that existing water infrastructure will require \$1 trillion for repairs and upgrades over the next 25 years.⁵⁴ Reductions in demand have prolonged the useable life of water distribution infrastructure and reduced utility spending on operation and maintenance activities. However, these reductions have also decreased water sales and limited utility revenue.⁵⁵

Water is becoming increasingly unaffordable for many in the United States, driving down demand. It has been projected that water rates will exceed the EPA's water unaffordability index for 35.6% of United States households by 2022.⁵⁶ Utilities may shut off water for delinquent accounts resulting in immediate risks to that household, but also economic risks to the broader community. Infrastructure maintenance costs are largely fixed, meaning that shutoffs cause maintenance costs to be distributed over a smaller population and thus an increased marginal rate on water. This is likely to further increase water rates for the remaining customers, creating a positive feedback loop that may cause cascading rate increases and water shut-offs.⁵⁶ Shutoffs also limit the throughput of utility distribution systems leading to increased HRT, presumably increasing concentrations of OPPPs.

In addition to factors which increase the prevalence of OPPPs in potable water, risk factors for the general population in the United States have also increased. Advanced age and immune status are critical risk factors for *Legionella* infections.^{1,7,40} Records show that the United States population is growing older, with the proportion of people over age 65 increasing from 13.7% in 2002 to a projected 20.3% by 2030,⁵⁷ and evidence suggests immunosuppression is also becoming more common.⁵⁸ Given the increased prevalence of *Legionella* spp. in potable water and increased risk factors of those using it, it is perhaps unsurprising that incidence of

Legionellosis has increased. Between 2000 and 2015, cases of Legionnaires' disease in the United States have increased by 450%.²⁷

1.4 Modeling Water Age

Water age, the duration that water spends in a plumbing system prior to use, cannot be measured directly but may be indirectly estimated using either tracer studies or mathematical simulations.^{4,14,59} Tracers studies use changes in water chemistry, typically induced by injecting tracing compounds such as salts or radionuclides at a reference point, and monitoring for those tracing compounds downstream.¹⁴ Tracer studies are labor intensive and provide water age results only for a single snapshot in time based on the water usage pattern at the time of the study. In contrast, mathematical simulations use plumbing network information (e.g. pipe segment lengths and diameters) and flowrate data to simulate the movement of water through plumbing networks. These simulations are limited by the accuracy and representativeness of the input data and the assumptions made by the model.¹⁴ Mathematical simulations can be used to estimate water age during an extended period of time, but require much more data about the plumbing system and significant effort to calibrate and validate the results.

With the exception of the lead and copper rule, federal monitoring requirements do not extend to building plumbing.⁴ Perhaps unsurprisingly then, currently available hydraulic modeling software is designed specifically for distribution systems, not necessarily building plumbing. The most widely cited of these software is EPANET, a public-domain software which simulates hydraulic and water quality behavior in pressurized water delivery piping.⁶⁰ Several other software packages have been developed to achieve the same goals, but most alternatives rely on EPANET's computational model⁶¹ which uses Lagrangian transport theory to model water flow, subdividing water volume into "fronts" and tracking their movement over time.

During advective transport, volumes of water occasionally must be created or destroyed to maintain assumptions regarding the fixed volume of the plumbing.^{59,60} EPANET, and alternates for that matter, are configured for pipe diameters and use patterns common in water distribution plumbing which reduces the impact of this error. However, the plumbing found in buildings is typically of smaller diameter, leading to a much smaller ratio of volume to pipe length.⁴ This exacerbates the issue of creating and/or destroying water volume. Further, the age of water in premise plumbing is anticipated to be more stratified than in distribution plumbing due to the intermittent operation of premise plumbing fixtures.

Schück⁵⁹ provides a demonstration of erroneous results when using EPANET to model water age in model premise plumbing system, and discusses modifications to EPANET source code and pipe network description to reduce the impact of these errors and generate plausible water age results.⁵⁹ However, these are complicated and do not fundamentally address the error. All hydraulic modeling software built using EPANET's computational engine are assumed to suffer from the same error, and no similar software specifically designed for building plumbing were identified. Thus a simple, accurate tool to determine water age in premise plumbing was desired to better examine the relationship between OPPP concentrations and water age.

1.5 Assessing Risks with QMRA

Quantitative microbial risk assessment (QMRA) is a framework used to quantify human health risks associated with exposure to pathogenic microorganisms. QMRA is typically conducted in five steps: hazard identification, dose response, exposure assessment, risk characterization, and risk management. During hazard assessment, a specific microbial hazard of concern is selected for analysis. Dose-response establishes the relationship between the dose of microorganism received and the probability of an individual developing ill-effects, such as an

infection or even death. An exposure assessment is conducted to determine the dose of microbial hazard delivered to individuals. Risks are then characterized, providing an overview of potential risks. The final step, risk management, uses each of the previous steps to find methods which reduce risks. These steps build upon one another and are conducted sequentially, except for dose-response and exposure assessment, which can be addressed simultaneously. This process, along with data and tools to help complete such assessments, is further detailed in *Quantitative Microbial Risk Assessment: Second Edition*.⁶²

QMRA is typically carried out stochastically using Monte-Carlo simulations, and has been used to quantify human health risks associated with *Legionella* exposure.^{63,64} For example, concentrations of *Legionella* spp. in water may be taken from primary research to develop a distribution. Samples may then be taken from this distribution to simulate concentrations of *Legionella* spp. in water to estimate human health risks. By conducting this assessment stochastically as described, the aleatory and epistemic uncertainty associated with natural variation in stochastic processes and the imperfect measurements of the data, respectively, are retained in analysis.

Data regarding *Legionella* spp. concentrations is generally sparse, and given the impacts of water quality on *Legionella* spp. concentrations, it may not be appropriate to apply concentrations measured in literature to a particular scenario. Modeling to determine concentrations of *Legionella* spp. based on water quality data may prove useful in addressing this gap in information, and help to resolve uncertainty regarding human health risks associated with exposure to water containing *Legionella* spp.

A significant body of literature has been published investigating factors that contribute to the growth of OPPPs, methods to reduce OPPP numbers, and in quantifying OPPP risks in

PPSs.^{3,5,65,66} However, PPSs are inherently heterogeneous due to differences in configurations, materials of construction, water use patterns, typical temperature range, and quantity and type of residual disinfectant. This degree of heterogeneity makes drawing generalized conclusions from plumbing studies difficult and potentially inappropriate. Bench scale studies have thus far been the primary tool to investigate the impacts of individual factors on OPPP prevalence. Bench scale studies allow specific variables, such as water temperature, to be isolated for their effects on pathogen concentrations to be better evaluated. However, this leaves the interactive effects of these variables unknown despite the dynamic nature of plumbing systems. The work presented here draws from data collected from a full-scale residential home, enabling a more holistic investigation of factors which influence *Legionella* spp. prevalence.

2 Chapter 2 – Objectives

Concerns about water scarcity and degraded source water quality have generated much interest in water efficiency and conservation programs in recent decades. Policies such as the Energy Policy Act of 1992 have set mandatory efficiency limits for consumer goods, while other programs like the Environmental Protection Agency's (EPA) WaterSense are voluntary and meant to encourage adoption of more water-efficient technologies.⁶⁷ Water use rates in the United States have changed significantly in recent years. Residential water consumption decreased 22% on a per household basis from 1999 to 2016.⁴⁹ Additionally, water withdrawals in the United States decreased 7% between 2010 and 2015 despite a population increase of 4% over the same period.⁴⁸

Ninety-five percent of people in the United States receive their water from a community water system.³⁴ In these water systems, drinking water is typically collected by a utility or municipality, treated, and then delivered to consumers via water distribution system (WDS) infrastructure. Water delivery piping within a WDS or within the plumbing of a building (i.e. premise plumbing system (PPS)), is frequently difficult to access. Hence, water infrastructure is designed to accommodate flows over the full lifespan of the pipes, thus sizing and capacity over this time period must be predicted during the design phase. Unfortunately, methodologies for predicting water demand are outdated and consistently overestimate water demand,⁴⁹ especially in light of the pervasiveness of water-efficient technologies. Additionally, residential plumbing codes have not been updated to address reduced water consumption, further compounding this issue. WDS and PPS plumbing are typically oversized as a result, increasing the duration of time that water spends in the piping, or the hydraulic retention time (HRT), also referred to as “water age”. Water quality is affected by elevated HRT through a variety of mechanisms including;

decay of residual disinfectant, the formation of carcinogenic disinfectant by-products, leaching of pipe materials into water, and the proliferation of opportunistic pathogens.¹⁴ In the United States, WDS plumbing has a combined total length of roughly one million miles (1.4 million km), while PPSs account for greater than 6 million miles (9.7 million km) in combined length,⁴ Further, while municipalities or utilities are responsible for water quality in WDSs, this responsibility in PPSs falls to the property owner who may not have the skills and expertise to properly address risks.

Health risks are climbing due to elevated concentrations of waterborne opportunistic plumbing pathogens (OPPPs) in treated drinking water as well as increased risk factors such as advanced age and chronic illness.^{57,68} Several OPPPs, such as *Legionella* spp., *Pseudomonas aeruginosa*, and Non-tuberculosis Mycobacteria, are commonly identified in drinking water⁸. This research focuses on *Legionella* spp., which has become the most common cause of waterborne illness in the United States.³² Evidence suggests that concentrations of OPPPs, including *Legionella* spp., are increasing.⁶ The proportion of the United States population over age 65, who are at higher risk of infection,⁷ is projected to increase from 13.7% in 2012 to 20.3% by 2030.⁵⁷ The incidence of Legionnaire's disease, caused by *Legionella pneumophila*, a common waterborne opportunistic pathogen, has increased by 450% in the United States between 2000 and 2015.²⁷ Healthcare costs associated with treating Legionnaire's disease and other such infections has been estimated to be \$850 million per year,^{6,69} while Naumova et al.¹³ estimated these costs at over \$2 billion per year.

A more complete understanding of actual water demand and the impacts of plumbing components and configurations, such as pipe configuration, pipe materials, water use patterns, water age, and temperature, on waterborne disease is required to develop improved

recommendations regarding plumbing design and ultimately combat rising infection rates.

Developing this knowledge prior to updating plumbing codes and repairing much of the United States water infrastructure has the potential to reduce incidence of drinking water associated disease.

Key knowledge gaps currently act as barriers to better informed plumbing design guidance and to ultimately reducing the health impacts and financial burden presented by *Legionella* spp. This research aims to address the following knowledge gaps:

1. Monitoring of *Legionella* spp. in PPSs is not required by any federal regulation. Additionally, such monitoring often not even feasible in many PPSs. Identifying key water quality properties which contribute to *Legionella* spp. prevalence could help to assess risks without widespread resource-intensive laboratory testing required to directly monitor bacterial concentrations. This research aims to identify a set of water quality variables which most-effectively inform *Legionella* spp. concentrations in a full scale PPS.
2. Water age has consistently been noted as a variable with significant impact on *Legionella* spp. concentrations, but computational methods or models to determine HRT are currently lacking. This research presents a novel method to estimate water age in a full-scale PPS, and investigates the utility of these results in predicting *Legionella* spp. concentrations.
3. Water temperature is also frequently cited in the literature as a variable with an impact on *Legionella* spp. prevalence. However, like water age, computational methods or models to assess the impacts of water temperature in premise plumbing systems have yet to be developed. This work presents a method to

determine whether water conforms to thermal guidance to control *Legionella* spp., and investigates its significance on resulting *Legionella* spp. concentrations.

3 Chapter 3 – Introduction to Methodological Approach and Exploratory Data Analysis

3.1 Primary Data Source

Experimental data for this study was collected from the Retrofitted Net-zero Energy, Water, and Waste (ReNEWW) house.⁷⁰ This home, located in West Lafayette, IN, was originally constructed in 1928 and underwent a complete plumbing retrofit in 2016. During the retrofit, all piping in the home was replaced with cross-linked polyethylene (PEX) type A pipe using a trunk-and-branch design. Brass fittings and valves are present in the plumbing. Additionally, several flowmeters and thermistors were installed to monitor water use and temperature. The house was designed to conserve water and energy compared to conventional homes. Fixtures throughout the home were selected for their efficient water use. Due to these design considerations, the ReNEWW house cannot be considered a typical residential building. However, these design elements are becoming more prevalent in homes across the United States. Further detail regarding the plumbing design and data collection efforts can be found in Salehi et al.,^{71,72} respectively. This data set is unique in scope due to the resources required to collect it. Instrumentation required to record flowrates and water temperatures cost approximately \$100,000 and over 220,000 labor hours were spent on sample collection and analysis over a one year period.⁷²

Plumbing at the ReNEWW house was surveyed to determine pipe length, pipe diameter, and to summarize plumbing configuration for this research. A piping and instrumentation diagram (P&ID) was developed using these results and is presented as **Figure 1**. The premise plumbing consists of ¾-inch and ½-inch nominal PEX-A style pipe, as denoted in **Figure 1**. The inner diameters of these pipes were assumed to be 1.73 and 1.23 cm, respectively. The locations of flowmeters, thermocouples, and sample locations are shown in **Figure 1**. Flowmeter

calibration was confirmed in May 2017 by manually collecting a measured volume of water from each plumbing fixture and comparing the volume to that recorded by the flowmeter. This process was repeated thrice for each flowmeter and showed the flowmeters were operating consistently and accurately. The plumbing configuration employed at the ReNEWW house is somewhat atypical for a residential PPS as it includes a hot-water recirculation loop to limit cooling of water at distal fixtures. This PPS also uses a thermostatic mixing valve to enable the water heater to achieve pathogen-killing temperatures of $60^{\circ}\text{C}^{1,8,73}$ while limiting the potential for scalding users with excessively hot water.

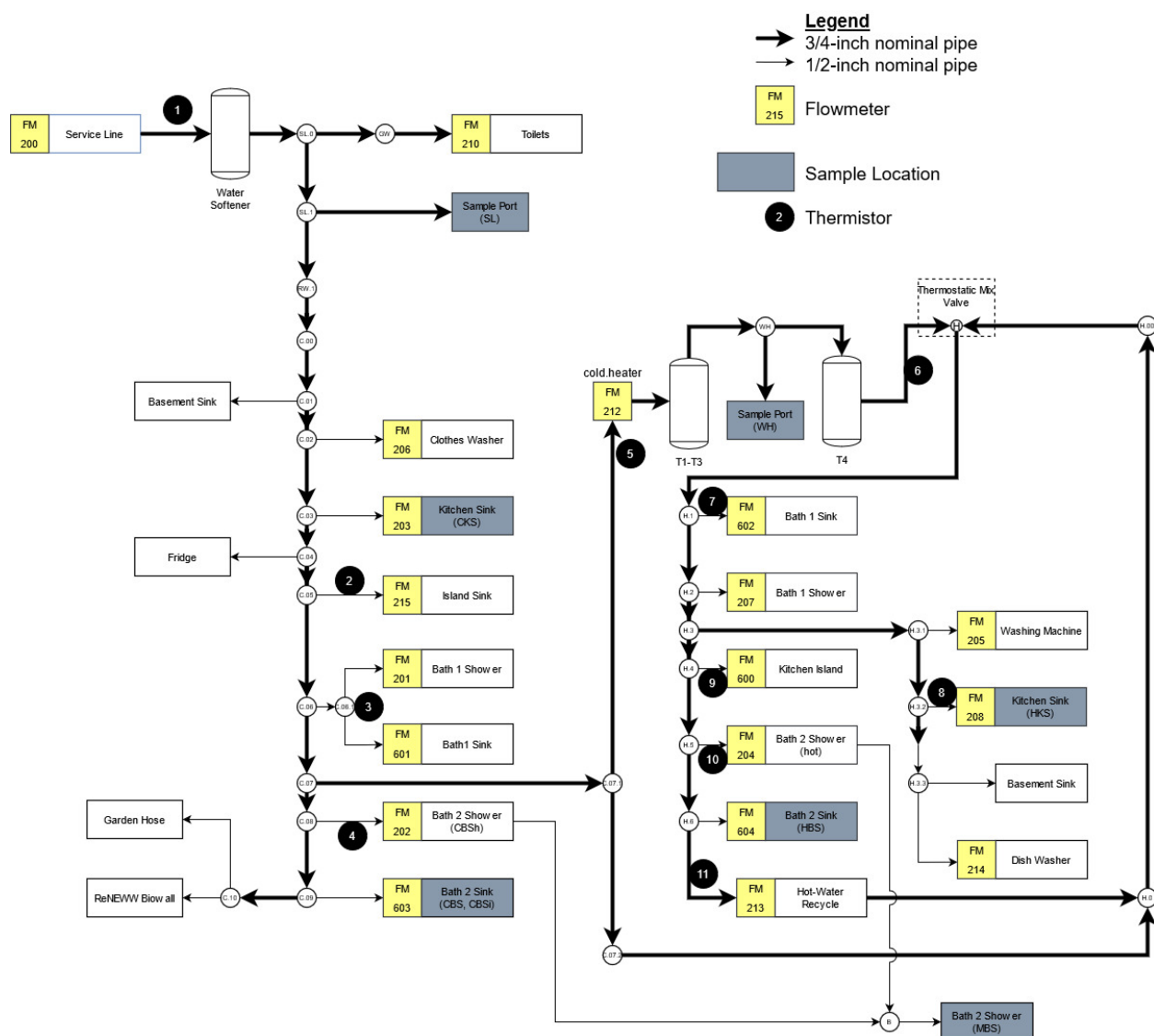


Figure 1 - ReNEWW house piping and instrumentation diagram. Pipe diameter is depicted with line width, flowmeters with squares, sample locations in shaded rectangles, and the approximate location of thermistors shown in numbered circles.

3.1.1 Analytical Data

Data from water samples were collected during 58 separate sampling events from October 2017 to October 2018. Each of the seven fixtures listed in **Table 1** was sampled during every event for a total of 406 total samples. Samples were collected from each of the fixture locations listed in **Table 1** in descending order during each sampling event. The design flowrate

of each sampled fixture was identified, as well as the cumulative volume and percent of total whole-home water consumption were calculated. Each of the collected samples was analyzed for each of the water quality variables listed in **Table 2**. Additional detail regarding water quality sampling and analysis can be found in Salehi et al.⁷² Additional data regarding the total and dissolved metals, as well as chromatography for ions commonly identified in water, were also collected. These data, such as iron and manganese concentrations, have been linked to increased *Legionella* concentrations.^{74–76} However, these data were ultimately not used in analysis because most results were below detection limits, providing little utility in data analysis. *Legionella* spp. were quantified using qPCR for *Legionella*'s 23s gene. Specific tests for the *Legionella pneumophila mip* gene were also performed, however each of these results was below detection limits⁷⁷. Further detail regarding the enumeration of *Legionella* spp. can be found in Ley et al.⁷⁷

Table 1 - Sample Collection Locations

Fixture Name	Abbreviation	Design Flowrate (LPM)	Total Volume (m³)	Percent of Total
Service Line	SL	NA	130.7	100%
Kitchen Sink - Cold	CKS	6.8	5.9	4%
Bathroom Sink - Cold	CBS	4.5	2.0	2%
Water Heater	WH	NA	40.6	31%
Kitchen Sink - Hot	HKS	6.8	5.2	4%
Bathroom Sink - Hot	HBS	4.5	16.2	12%
Bathroom Shower - Mixed	BSM	7.6	36.6	28%

Table 2 - Monitored Variables

Variable Name	Variable Description	Units	Log transformed	Percentile (natural scale)			Number of Observations
				2.5%	50.0%	97.5%	
pH	pH	NA	No	7.36	8.00	9.04	406
Temp	Temperature	C	No	15.63	22.90	26.30	406
DO	Dissolved oxygen	mg/L	No	4.30	8.40	10.56	406
Total.Cl	Total Chlorine	mg/L	Yes	BDL	0.10	1.00	406
Free.Cl	Free Chlorine	mg/L	Yes	BDL	0.01	0.75	259
TOC	Total Organic Carbon	mg/L	Yes	0.42	0.81	15.36	406
DOC	Dissolved Organic Carbon	mg/L	Yes	0.42	0.73	18.97	371
Alka	Alkalinity	mg/L as CaCO ₃	Yes	264.15	287.25	332.65	377
TTHM	Total Trihalomethanes	mg/L	No	0.05	15.57	31.55	399
TCC	Total Cell Count	#cells/mL	Yes	1.54E+03	3.77E+04	1.56E+06	406
HPC	Heterotrophic Plate Count (by culture)	CFU/100mL	Yes	4.03E+00	1.01E+04	3.60E+07	390
Leg.sp	<i>Legionella</i> spp. (by qPCR)	gene copies/100mL	Yes	2.29E+01	4.02E+03	1.78E+05	258

Despite best efforts to ensure complete data were collected, several results are missing for individual variables, as indicated in **Table 2**. Many results were found to be below method detection limits for free and total chlorine, TTHM, and *Legionella* spp. The lower detection limit (LDL) for chlorine and TTHM testing was 0.1 mg/L. The LDL for *Legionella* spp. was variable depending on the final sample concentration, and ranged from 13.3 to 38.9 gene copies per 100 mL. For each set of these results below the respective LDL, half of that LDL was taken as the result of the analysis.

The Shapiro-Wilks diagnostic⁷⁸ was used to test the data for normality and was performed in R,⁷⁹ implemented in RStudio.⁸⁰ The test diagnostic, *W*, was used to determine whether natural or log-transformed data resulted in a higher *W*-statistic. This tests whether the data from each variable more closely conforms to a normal distribution before or after the log transformation. In each case where *W* for the log-transformed data was less than that of the natural-scale data, indicating the log-transformed data more closely conformed to a normal

distribution, further analyses were conducted on that log-transformed data. This distinction is noted in **Table 2**. Log-transformation of these variables was performed to better linearize the relationships between variables for subsequent analyses which assume linear relationships (e.g. principal component analysis and generalized linear modeling). Log transformation has no effect on Spearman's rank-order correlation coefficient.

3.1.2 Electronically Recorded Data

Water flowrates were monitored in the ReNEWW house with 19 Omega FPR300⁸¹. These flowmeters recorded data with a one-second resolution between August 2015 and May 2019. The location of each of these flowmeters is depicted in **Figure 1**. Water use events were defined from flowmeter data as any time water was used for longer than three consecutive seconds from a single water fixture.⁷¹ Further, any use events with less than five seconds between them were combined to reduce noise in the data. During data analysis two flowmeters, located at the hot kitchen island sink and hot bath 2 sink, were found to have errant results far greater than the design flowrate of these fixtures, likely as a result of a poor electrical connection with the sensors. These data were evaluated and ultimately replaced with a ratio multiple of the cold water data from adjacent fixtures as a reasonable estimate for water use. This process is detailed in the **Appendix**.

3.1.3 Development of water use metrics

The time that water spends in building plumbing (i.e., the difference in time from entry to the home and exiting the tap) is referred to as water age. However, it is important to note that water age cannot be directly measured. To approximate water age, four water use metrics were developed as surrogates from water use records: the number of uses (num.events), the cumulative volume used (vol.events), and the mean and maximum time since last water use

(meanTSL and maxTSL). It is important to note that none of these measures accurately represent water age because the age of water stored in the plumbing immediately preceding water use is not considered. The metrics num.events, meanTSL, and maxTSL are instead related to the *frequency of water use*, and vol.events is related to the *total water volume consumed* at each fixture. Each of these water use metrics are calculated over a specific time period.

While it is well-established that water age impacts water quality,^{4,14} it is largely unknown how long elevated water age must be maintained in building plumbing for degradation of water quality to manifest in sampling results. To select an appropriate time period, Spearman correlation coefficients calculated between analytical results and water use metrics were evaluated over the time period time period from 1-120 days, as shown in **Figure 2** to determine a relationship between these correlations and time. Variables appear to be affected by time period differently. As the time period increases from a single day the strength of correlation with water use metrics increases for most variables. The absolute value of the correlation begins to decline for most variables, including total chlorine, temperature, HPC, and *Legionella* spp., after 5-20 days of use data are incorporated into the calculation of the use metrics. The strength of correlation increases again for several variables when using a time period >60 days. Ultimately, a period of 14 days, shown as a red vertical line in **Figure 2**, was selected with the goal of capturing strong correlations for as many variables as possible while still using a consistent time period for further analyses. A single time period, rather than multiple and/or parameter-specific periods, was chosen for simplicity and to ease further analyses.

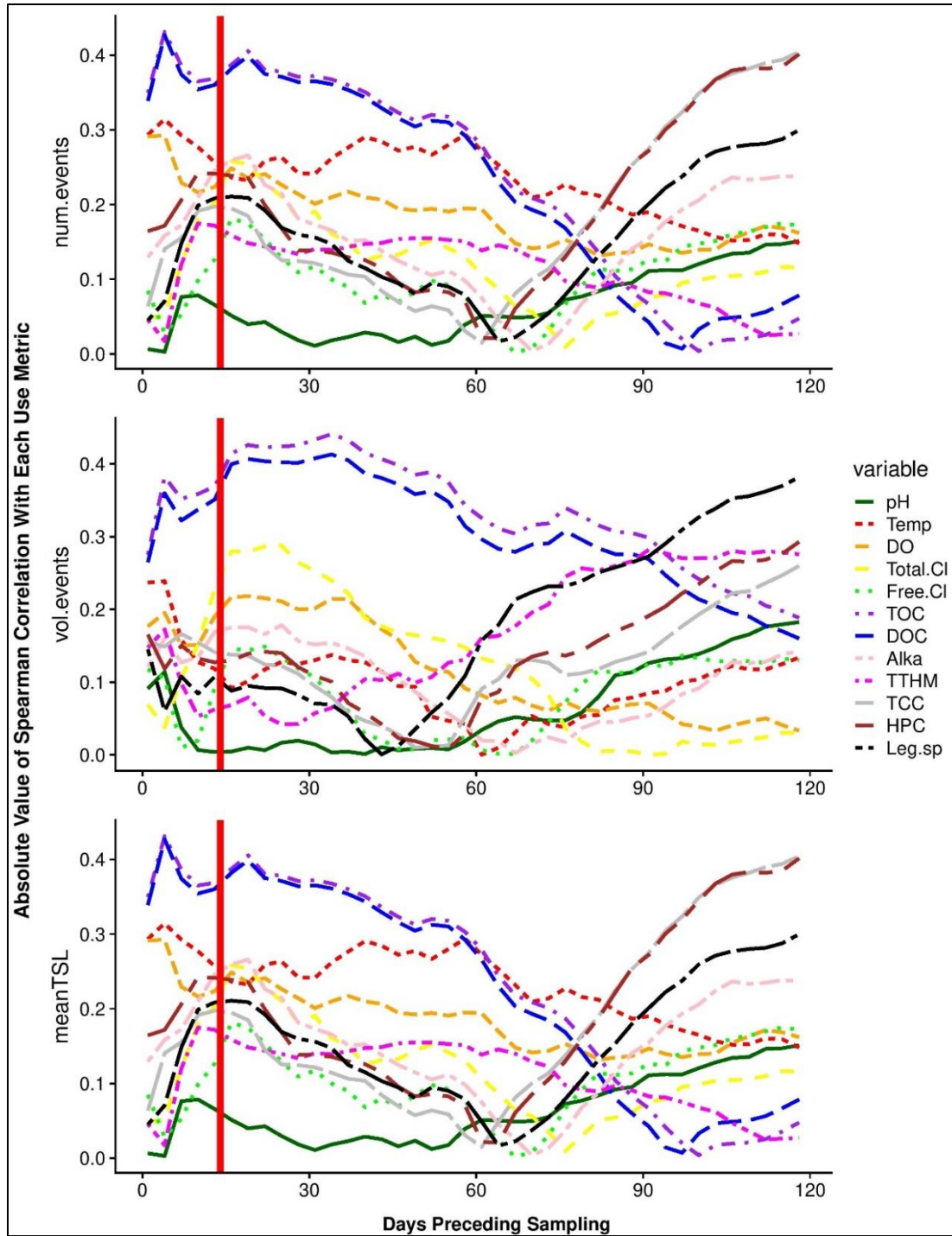


Figure 2 - Absolute value of Spearman correlation coefficients between water quality variables and water use metrics, calculated using varying time periods from 1 to 120 days prior to sample collection. The red vertical line represents the time period selected for analysis, 14 days.

These findings demonstrate that the strength of correlations between water quality and water use metrics vary depending on the duration of water use considered. Based on visual inspection of **Figure 2**, there appear to be two local maxima for many variables, which may be explained by two types of processes that have been discussed in literature: (i) those directly affected by use such as leaching,⁸² flushing,⁴ and biofilm detachment;^{15,22} and (ii) those mediated by established biofilm^{46,83} Several of these correlations, including free and total chlorine, TTHM, and DO, appear less sensitive to changes in time period as that period becomes larger (i.e. greater than 100 days). This may suggest that long-term trends at each fixture have an important role regarding changes in water quality as well as to the biological and mechanical stability of biofilm as has been noted in the literature.^{15,23,24} Note that variables are affected differently by the choice of time period. Selecting an alternative time period may impact relationships described in subsequent analyses that rely on water use metrics.

3.2 Statistical Analyses Conducted

Multiple statistical methods were used to evaluate data collected from the ReNEW house, including correlation analysis, principal component analysis (PCA), and generalized linear modeling (GLM). Spearman correlation coefficients provide a sense of the interdependence between variables, and were used as a screening mechanism to eliminate highly correlated analytical variables with missing observations from PCA and GLM, allowing additional data to be considered in these analyses. PCA was used to identify phenomena influential to water quality which were not directly measured in the study. Multiple iterations of GLM were considered to identify variables with the strongest predictive power to *Legionella* spp. concentrations. The value of each variable in predicting *Legionella* spp. was also evaluated using a Bayesian generalized linear regression (BGLR) variable selection technique. The results

of each of these analyses were considered in concert to provide a more holistic understanding of the interactions which drive changes in building water quality. Results of these analyses are presented in **Chapter 4**, and are used to assess the significance of relationships measured variables have on concentrations of *Legionella* spp. These analyses are then again repeated in **Chapter 5** to assess the impact of including a metric to represent water age on the overall results.

3.2.1 Spearman's Rank Correlation

Spearman's rank correlation coefficient is a bivariate, non-parametric measure which expresses the degree of statistical dependence of a pair of variables based on a comparison of their ordinal ranks. As such, Spearman's rank makes no assumption about the linearity of relationship between variables as is required of alternative methods such as Pearson's correlation coefficient. Spearman's rank has been used in previous literature to explore similar water quality relationships in PPSs as those evaluated in this dissertation, such as *Legionella* relationships with HPC, TCC, water temperature, and chlorine concentration.^{77,84,85} It is also important to recognize that Spearman's rank, much like Pearson's correlation coefficient, is not suited to assess non-monotonic relationships. Bivariate scatterplots were visually inspected for obvious signs of nonmonotonicity, though no such relationships were identified. Spearman's rank correlation coefficients (denoted as ρ) were calculated using the "rcorr()" function of the Hmisc library⁸⁶ in R,⁷⁹ implemented in RStudio.⁸⁰

3.2.2 Principal Component Analysis

PCA analysis is a linear dimensionality reduction technique that maps data into a subspace with fewer dimensions. Orthogonal vectors called principal components (PCs) are selected within that subspace to maximize variance attributed to each PC, effectively indicating their significance to the data as a whole. In some instances it is possible to relate a PC with a

known physical phenomenon.⁸⁷ PCA was selected for these analyses for its potential to identify key features in the data and important physical phenomena associated with interactions among variables. PCA was conducted in RStudio.⁸⁰ Bartlett's Sphericity Test⁸⁸ was used to verify that PCA efficiently reduces the dimensionality of each data set analyzed. PCA was utilized in each study to investigate the influence of individual variables on the overall water quality in the plumbing. In **Chapter 4**, this includes water quality measurements as well as water use metrics developed from flowmeter data. PCA is also used in **Chapter 5** to determine whether the addition of a mechanistically plausible water age metric clarifies any relationships. The goal of PCA in this application was to identify processes which result in the greatest variability in the water quality data to better describe factors associated with increased concentrations of *Legionella* spp. Specific hypothesis regarding this PCA were not developed beforehand to allow the data to drive the analysis and to better identify risk factors.

3.2.3 Generalized Linear Modeling and Generalized Linear Mixed Modeling

Generalized linear modeling (GLM) is a multivariate linear regression technique in which multiple independent continuous variables are used to predict a single continuous dependent variable. Relationships between water quality variables in this study were not presumed to be linear, and as such, the results from GLM are not intended for use as a predictive model of *Legionella* spp. concentrations. Instead, GLM is used here to determine how effectively water quality variables explain the variance observed in *Legionella* spp. concentrations. GLM has been used in published literature to explore similar relationships as those investigated in this dissertation⁸⁹

Only data with complete observations (i.e., no observations are missing for any variables) may be analyzed using GLM. With all analytical predictors included (i.e. water quality variables

and water use metrics), a total of 133 observations were available and used in the GLM analysis due to missing data. No data developed from flowmeter data (e.g. meanTSL, age.mean) were missing. Analytical, variables were selectively eliminated based on the results of a preliminary GLM to increase the robustness of the data set and improve model performance. Free.Cl and DOC were eliminated from GLM analyses due to frequent missing data and a high degree of correlation with alternative variables. Free.Cl was missing 147 observations and exhibited a correlation of 0.791 with Total.Cl, for which all data were observed. DOC was missing 35 observations and exhibited a correlation of 0.966 with TOC, for which all data were observed. As such, Total.Cl and TOC were selected as proxies for Free.Cl and DOC, respectively. This allowed for the inclusion of 89 additional observations, for a total of 222, in GLM analysis.

GLM was conducted in an iterative fashion, using the second-order Akaike Information Criterion (AIC_C) to further eliminate insignificant variables as described in the results of **Chapter 4** and **Chapter 5**. AIC_C is a modified version of the Akaike Information criterion, modified to address small sample sizes. As a general rule of thumb, when the ratio of the number of observations to the number of variables in the model is less than 40, AIC_C is preferred to AIC.⁹⁰ All possible combinations of first-order effects were evaluated using the ‘dredge()’ function from the R library, “MuMin”.⁹¹ In each case, the model with the lowest AIC_C was identified as the top-performing model (m.top). A competing model (m.comp) was defined as including all the variables found in models with a difference in AIC_C (ΔAIC_C) of less than two. The variable sets identified in m.top and m.comp were each further investigated using a generalized linear mixed model (GLMM) which included a random effect to address the categorical impact sample location.

This process was conducted in **Chapter 4** and **Chapter 5** to identify variables that predict *Legionella* spp. concentrations with statistical significance. It was again conducted in the water age modeling study to determine whether the inclusion of plausible water age metrics resulted in a selection of alternative variables.

3.2.4 Bayesian Generalized Linear Regression

A Bayesian variable selection method was utilized to further investigate relationships between water quality variables and *Legionella* spp. concentrations. Each of the two models defined from GLMM analysis were fit using the R package BGLR.⁹² Studies such as O'Hara and Sillanpää⁹³ and Woznicki et al.⁹⁴ have reported the effectiveness of Bayesian variable selection. In this research, the BGLR library was used to calculate the probability of each parameter having a non-zero estimate, which is taken as evidence that the variable has a significant impact on *Legionella* spp. concentrations. This library relies on Gibbs sampling and employs scalar updates in parameter estimation.⁹² The BGLR library includes several options for the selection of prior probabilities in the Bayesian framework, including Gaussian and scaled-t mixtures that include a large point-mass at zero and are suitable for variable selection.⁹⁵

4 Chapter 4 – Variable Selection

4.1 Introduction

Quantifying concentrations of *Legionella* is critical to assessing and managing human health risks. Identifying a limited set of factors which most influence *Legionella* spp. concentrations in full-scale building plumbing systems is expected to alleviate some of the challenge and expense of conducting water quality monitoring. This would enable sampling for less-costly analytes to inform *Legionella* spp. risks. Several studies have been conducted to this end but have not identified suitable surrogate monitoring strategies.^{84,96–98} Building plumbing environments undergo significant spatiotemporal variation in chemical and biological conditions.^{3,72} As such, large data sets are necessary to distinguish between mechanistic effects and variability inherent to building plumbing.

4.2 Methods

This study relies on a rich data set from a residential home equipped with high-efficiency fixtures and appliances. Previous literature has identified water quality degradation in bench-scale studies pertaining to building plumbing, such as the leaching of carbon from pipes⁸² and proliferation of *Legionella* spp. over time.⁹⁹ These previously unavailable, high-resolution data increase confidence in these bench-scale assessments. Further, this study presents an opportunity to confirm phenomena observed at bench-scale are also apparent at full-scale and to assess the relative strength of these processes. The objectives of this study were to (i) identify water quality variables that are most strongly related with *Legionella* spp.; and (ii) elucidate interactions between variables that ultimately influence *Legionella* spp. concentrations.

To achieve these goals, relationships in the data were analyzed using a suite of statistical tools. Spearman's correlation coefficient was selected to evaluate bivariate relationships between

each pair of variables. Spearman's correlation was selected because it is non-parametric and makes no assumption about the distribution of the variables. Spearman's correlation coefficient is suited only to characterize monotonic relationships. Scatterplots of each variable were visibly inspected for signs of nonmonotonicity, and no such patterns were identified. Principal component analysis was used to reduce the dimensionality of the data and identify phenomena related to the variation across all variables. GLM was selected to evaluate the statistical significance of each variable on *Legionella* spp. concentrations. These results informed a GLMM, which included the sample location as a random variable to account for variation between fixtures. The significance of each variable in this analysis are taken as evidence of the predictive value of each variable on *Legionella* spp. Finally, a Bayesian generalized linear regression technique used in variable selection was performed. These results present the probability of each variable having a non-zero parameter in a linear model to predict *Legionella* spp., which is taken as evidence the variable has a significant relationship with *Legionella* spp. concentrations.

4.3 Results

4.3.1 Correlation Coefficients

Spearman's Rank was selected to measure the degree of association between each of the variables in this analysis. These results were used as a screening technique to inform further analyses and can be found in **Table 3**. These results support the notion that building plumbing

environments are complex. Changes in one variable likely propagate to other variables or may be muted due to equilibria and buffering.

Table 3 - Spearman's rank correlation coefficients

Variable	pH	Temp	DO	Total.Cl	Free.Cl	TOC	DOC	Alka	TTHM	TCC	HPC	Leg.sp	vol.events	num.events	meanTSL	maxTSL
pH	1.00	0.09	-0.30	-0.20	-0.19	0.19	0.17	0.07	0.24	-0.06	0.19	0.15	-0.11	-0.19	0.18	0.14
Temp	0.09	1.00	-0.40	-0.51	-0.41	0.49	0.53	0.31	0.20	0.48	0.46	0.35	-0.27	-0.29	0.34	0.47
DO	-0.30	-0.40	1.00	0.27	0.22	-0.42	-0.45	-0.14	-0.34	-0.23	-0.35	0.08	0.34	0.38	-0.39	-0.50
Total.Cl	-0.20	-0.51	0.27	1.00	0.79	-0.35	-0.42	-0.21	-0.29	-0.28	-0.30	-0.48	0.19	0.16	-0.21	-0.26
Free.Cl	-0.19	-0.41	0.22	0.79	1.00	-0.25	-0.27	-0.12	-0.33	-0.14	-0.16	-0.32	0.15	0.04	-0.07	-0.12
TOC	0.19	0.49	-0.42	-0.35	-0.25	1.00	0.97	0.36	0.65	0.53	0.61	0.53	-0.57	-0.53	0.53	0.61
DOC	0.17	0.53	-0.45	-0.42	-0.27	0.97	1.00	0.36	0.62	0.56	0.60	0.52	-0.57	-0.53	0.54	0.63
Alka	0.07	0.31	-0.14	-0.21	-0.12	0.36	0.36	1.00	0.32	0.56	0.50	0.54	-0.17	-0.12	0.14	0.29
TTHM	0.24	0.20	-0.34	-0.29	-0.33	0.65	0.62	0.32	1.00	0.26	0.37	0.33	-0.47	-0.31	0.30	0.38
TCC	-0.06	0.48	-0.23	-0.28	-0.14	0.53	0.56	0.56	0.26	1.00	0.70	0.54	-0.16	-0.19	0.23	0.38
HPC	0.19	0.46	-0.35	-0.30	-0.16	0.61	0.60	0.50	0.37	0.70	1.00	0.62	-0.24	-0.40	0.42	0.49
Leg.sp	0.15	0.35	0.08	-0.48	-0.32	0.53	0.52	0.54	0.33	0.54	0.62	1.00	-0.22	-0.35	0.39	0.22
vol.events	-0.11	-0.27	0.34	0.19	0.15	-0.57	-0.57	-0.17	-0.47	-0.16	-0.24	-0.22	1.00	0.75	-0.74	-0.72
num.events	-0.19	-0.29	0.38	0.16	0.04	-0.53	-0.53	-0.12	-0.31	-0.19	-0.40	-0.35	0.75	1.00	-0.99	-0.79
meanTSL	0.18	0.34	-0.39	-0.21	-0.07	0.53	0.54	0.14	0.30	0.23	0.42	0.39	-0.74	-0.99	1.00	0.79
maxTSL	0.14	0.47	-0.50	-0.26	-0.12	0.61	0.63	0.29	0.38	0.38	0.49	0.22	-0.72	-0.79	0.79	1.00

Several variables exhibit a relatively high degree of correlation because the variables are intrinsically related. For example, the four water use metrics were highly correlated with one another. The number of water uses and meanTSL are nearly perfectly negatively correlated (-0.989). The high strength of this correlation is expected because meanTSL is essentially an inversion of num.events. num.events and maxTSL show a similar correlation, albeit weaker (-0.793) as maxTSL represents a single value rather than a mean over the two-week period. num.events and vol.events are also strongly correlated (0.754) because both the volume and frequency of use increase whenever water is used. Typical duration and flowrate of water use differ across fixtures, which are not accounted for in the metrics, making the correlation weaker.

As a consequence of these two points, vol.events has very similar correlations to num.events (0.754), meanTSL (-0.740), and maxTSL (-0.715). Some pairs of analytical results exhibit high correlation as well, such as DOC and TOC (0.966), as well as Free.Cl and Total.Cl (0.791), which is expected given that these tests have some overlap (e.g., DOC is a constituent of TOC). These correlations were considered as a data screening technique to determine which variables to include in linear modeling and principal component analysis. For any pair of variables with missing data, if the absolute value of the correlation coefficient was greater than 0.75, the variable with more missing data was excluded from subsequent analysis. As such, Free.Cl and DOC were not included in PCA or GLM analyses. As such, TOC and Total.Cl can be considered surrogates for DOC and Free.Cl, respectively, in the subsequent analyses.

4.3.2 Principal Component Analysis

A summary of the PCA results can be found in **Table 4**. PCs one through three (i.e., PC1, PC2 and PC3) have a standard deviation of greater than one and were considered relevant, cumulatively accounting for 62% of the variance in the data. Loading factors for each of the first three PCs were reviewed in concert, incorporating knowledge gained through expert knowledge and published literature, to synthesize plausible hydraulic, chemical, or biological interpretations of each PC.⁸⁷ PC1 appears to be related to elevated water age due to loading factors such as those on vol.events (-0.280) and meanTSL (0.206). In other words, PC1 is positively associated with the time between water uses and negatively associated with flushing the plumbing with water from the service line. Relationships between other variables offer further support to interpret PC1 as a metric for water age. PC2 has strong loading factors of 0.466 and -0.545 on pH and DO, respectively, which have been related to biofilm sloughing.^{23,24} PC2 has a negative loading factor on vol.events (-0.282), with minimal loadings on meanTSL (0.018), or Temp (0.038) further

supporting this interpretation. Additionally, all three microbial measures have negative PC2 loading factors (-0.430, -0.283, -0.161), suggesting microbial washout. Taken together, these loading factors suggest PC2 is related to biofilm detachment. The strongest loadings on PC3 were on meanTSL (-0.519), Total.Cl (-0.430), TTHM (0.347), pH (-0.342), and vol.events (-0.300). No specific phenomena were identified in the literature to definitively explain these loadings.

Table 4 - Principal Component Analysis Results

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
pH	0.16	0.47	-0.34	0.47	-0.10	0.05	-0.14	0.61	-0.07	-0.12	-0.05	-0.04
Temp	0.32	0.04	0.14	0.08	0.62	-0.16	0.38	0.03	0.23	-0.35	-0.38	0.04
DO	-0.17	-0.55	0.34	-0.05	0.15	-0.04	-0.24	0.65	-0.14	0.02	-0.09	0.15
Total.Cl	-0.29	-0.14	-0.43	-0.34	-0.29	-0.13	0.03	0.16	0.57	-0.25	-0.29	-0.05
TCC	0.32	-0.43	-0.08	0.09	-0.12	0.19	0.01	-0.01	-0.09	-0.42	0.20	-0.65
HPC	0.37	-0.28	-0.19	0.16	-0.07	0.09	-0.05	-0.06	0.27	-0.17	0.42	0.66
Leg.sp	0.37	-0.16	-0.01	0.19	0.01	0.15	-0.44	-0.15	0.35	0.49	-0.43	-0.14
TOC	0.33	0.04	0.03	-0.36	-0.14	0.38	0.55	0.35	0.07	0.40	0.07	-0.03
Alka	0.26	-0.25	-0.21	0.10	-0.27	-0.71	0.26	-0.01	-0.31	0.23	-0.15	0.03
TTHM	0.30	0.18	0.35	-0.24	-0.48	0.11	-0.16	-0.06	-0.23	-0.38	-0.41	0.24
vol.events	-0.28	-0.28	-0.30	0.32	-0.02	0.47	0.30	-0.16	-0.32	-0.01	-0.41	0.20
meanTSL	0.21	0.02	-0.52	-0.53	0.40	0.06	-0.31	0.00	-0.38	-0.01	-0.04	0.06
Standard deviation	2.17	1.28	1.05	0.98	0.92	0.85	0.69	0.66	0.60	0.56	0.49	0.43
Proportion of Variance	0.39	0.14	0.09	0.08	0.07	0.06	0.04	0.04	0.03	0.03	0.02	0.02
Cumulative Proportion	0.39	0.53	0.62	0.70	0.77	0.83	0.87	0.91	0.94	0.96	0.98	1.00

4.3.3 General Linear Model

An initial GLM was constructed to reduce the number of variables. This model relied on 13 variables and included 222 observations of; pH, Temp, DO, Total.Cl, TCC, HPC, TOC, Alka, TTHM, num.events, vol.events, meanTSL, maxTSL; as input variables to predict the corresponding concentration of *Legionella*. All possible combinations of these variables as main-effects were evaluated using the ‘dredge’ function of the R library, “MuMin”,⁹¹ and sorted using the second-order Akaike Information Criterion (AIC_C), to identify variables in best-performing

models. Eleven models were identified as having a difference in AIC_C (ΔAIC_C) of less than 2.0 from the top-performing model. Of the 13 predictor variables included in the GLM, only Alka did not appear within these top-performing models. As a result, Alka was excluded from further analysis, allowing for 13 more observations to be included in the next iteration of the GLM.

A second GLM was initiated to predict Leg.sp using the 12 variables identified in top-ranked GLM models: DO, HPC, maxTSL, meanTSL, num.events, pH, TCC, Temp, TOC, Total.Cl, TTHM, and vol.events as predictors. This GLM utilized 235 observations of the data. Combinations of variables were again evaluated using the “dredge” function from the “MuMin” library available for R,⁹¹ and the top-ranked model included the variables pH, DO, HPC, maxTSL, meanTSL, num.events, TCC, TOC, and Total.Cl. Competing models with a ΔAIC_C of less than 2.0 were also considered, identifying only TTHM in addition to variables identified in the top-ranked model.

Finally, two generalized linear mixed models (GLMMs) were constructed, including the variables from the top-ranked model (m.top) and variables from competing models (m.comp), as well as a random effect to describe sample location. Both models exhibited model convergence, homoscedasticity, and low multicollinearity, and are summarized in **Table 4**. Both models exhibited similar performance. DO and HPC were identified as statistically significant at the $p < 0.001$ level in both models. TOC was significant at the $p < 0.001$ level in m.top but only at the $p < 0.01$ level in m.comp. TCC was significant at the $p < 0.01$ level, and Total.Cl, meanTSL, and pH were each significant at $p < 0.05$ in both models. maxTSL and num.events were not significant. m.comp was unique in that it used TTHM as a predictor variable. However, the significance of TTHM was limited as it had a p-value of > 0.1 . m.top and m.comp exhibited

AIC_C values of 992.7 and 994.2, respectively. m.top was selected as the preferred model based on the lower AIC_C and fewer variables.

Table 5 - Comparison of Generalized Linear Mixed Models m.top and m.comp

Variable	m.top				m.comp			
	Estimate	Std. Error	z-value	Pr(> z)	Estimate	Std. Error	z-value	Pr(> z)
Intercept	-5.14E+00	1.46E+00	-3.52	4.38E-04	-5.34E+00	1.49E+00	-3.59	3.37E-04
DO	3.03E-01	4.99E-02	6.06	1.34E-09	3.05E-01	5.00E-02	6.10	1.05E-09
HPC	2.64E-01	5.39E-02	4.91	9.32E-07	2.70E-01	5.44E-02	4.96	7.11E-07
maxTSL	-1.21E-06	7.09E-07	-1.71	0.09	-1.15E-06	7.13E-07	-1.62	0.11
meanTSL	6.76E-06	2.93E-06	2.30	0.02	6.66E-06	2.94E-06	2.27	0.02
num.events	3.37E-07	7.92E-05	0.00	1.00	7.02E-08	7.90E-05	0.00	1.00
pH	3.34E-01	1.53E-01	2.18	0.03	3.47E-01	1.55E-01	2.25	0.02
TCC	3.26E-01	1.04E-01	3.13	1.72E-03	3.18E-01	1.04E-01	3.05	2.28E-03
TOC	7.17E-01	2.14E-01	3.35	8.04E-04	6.52E-01	2.34E-01	2.79	0.01
Total.Cl	-3.12E-01	1.26E-01	-2.47	0.01	-2.95E-01	1.28E-01	-2.31	0.02
TTHM	NA	NA	NA	NA	6.49E-03	9.53E-03	0.68	0.50
AIC _C	992.7				994.2			

4.3.4 Bayesian Variable Selection Method

The probability that each parameter had a non-zero probability was estimated for each of the linear models using the BGLR library in RStudio.⁹² This estimate, as well as estimates for parameter and standard deviation, are presented in **Table 6** for each of the three variable sets analyzed using the frequentist GLM presented above. The resulting parameter estimates were much closer to zero than frequentist GLM results. All variables except for maxTSL and meanTSL showed probabilities for non-zero parameter estimates of between 52% and 57%. meanTSL showed probabilities of a non-zero parameter of between 98% and 100% depending on the variable set, whereas probabilities for maxTSL were lower, between 14% and 38%.

Table 6 - BLGR Variable Selection Results (BayesC priors)

Variable	m.full			m.top			m.comp		
	Prob.	β	SD(β)	Prob.	β	SD(β)	Prob.	β	SD(β)
DO	0.54	6.99E-09	1.89E-05	0.53	-1.26E-07	1.41E-05	0.56	1.12E-07	1.35E-05
HPC	0.54	3.40E-08	1.92E-05	0.52	-2.25E-08	1.42E-05	0.57	-5.77E-08	1.35E-05
maxTSL	0.22	3.36E-07	1.76E-05	0.14	1.07E-07	1.32E-05	0.38	5.38E-07	1.13E-05
meanTSL	0.98	1.43E-05	5.20E-06	1.00	1.51E-05	3.30E-06	0.99	1.33E-05	4.31E-06
num.events	0.56	-5.45E-06	2.08E-05	0.52	5.73E-07	1.40E-05	0.56	3.92E-07	1.33E-05
pH	0.53	-1.71E-07	1.89E-05	0.52	1.40E-07	1.41E-05	0.57	-3.52E-08	1.35E-05
TCC	0.54	-2.08E-07	1.92E-05	0.53	7.85E-08	1.43E-05	0.57	2.36E-07	1.35E-05
TOC	0.54	2.81E-07	1.89E-05	0.52	-4.95E-08	1.41E-05	0.56	-1.28E-08	1.35E-05
Total.Cl	0.54	-1.65E-08	1.89E-05	0.52	-2.09E-07	1.41E-05	0.57	-9.28E-08	1.36E-05
TTHM	0.54	4.10E-08	1.89E-05	NA	NA	NA	0.57	-3.46E-09	1.35E-05
vol.events	0.55	-3.31E-06	2.01E-05	NA	NA	NA	NA	NA	NA
Temp	0.54	2.21E-07	1.90E-05	NA	NA	NA	NA	NA	NA

While these results suggest that meanTSL has a strong influence on Leg.sp, no other variables were identified as likely contributors. The same analysis was conducted using flat, uninformative priors, which resulted in parameter estimates nearly identical to those obtained using the frequentist GLMs, lending support to the methods employed by BGLR. It is hypothesized that the relationships each variable have with Leg.sp are not strong enough to overcome the zero-biased prior assumptions used in variable selection, leading to near-zero parameter estimates due to the high degree of correlation and interrelatedness of the variables in this analysis.

4.4 Discussion

To date, published research on this topic has been most often conducted at the bench-scale. Results from the analyses for a full-scale plumbing system largely corroborate effects observed in published literature conducted at bench-scale, such as the positive association of water age with increased concentrations of *Legionella* spp.^{2,4,7,10} and TOC.^{82,99,100} Building plumbing environments are highly variable due to differences in several factors such as

operation, materials of construction, piping layout, incoming water quality, and water temperature. This variability makes it difficult to develop generic guidance for reducing pathogen risks regarding building plumbing.

This study is unique in that it was conducted in a full-scale residential home with online monitoring capabilities unmatched by other buildings across the United States based on discussions with the plumbing and green building industries. In addition to the high level of resources expended to collect this rich data set for analyses, the close proximity of the sampling location to the analytical lab facilitated increased data collection by eliminating the logistical requirements associated with sample handling such as long-term storage and transportation. The wealth of analytical and electronic water use data collected provide a unique opportunity to study how these effects interact with real plumbing and water use patterns. The scale of this study and its data collection efforts provide a unique opportunity to examine variable interactions in fully operational building plumbing. Our results indicate that *Legionella* spp. concentrations are most closely related to a limited subset of the water quality variables measured in this study. These findings have the potential to reduce labor and analysis costs of future OPP studies by identifying variables most relevant to *Legionella* spp. Further, the results of this study corroborate several relationships identified in previous literature, lending credibility to these findings. These relationships are explored in the following subsections.

4.4.1 Flushing, Stagnation, and Water Age

Water consumption at any fixture draws water through the system, refreshing the pipes with water from the service line. This effect also dictates the age of water within the plumbing, which has implications on how each of the measured constituents and properties accumulate or degrade within the water. PCA appears to have identified these effects as PC1. By flushing

plumbing with water from the service line dissolved oxygen and chlorine, which are consumed in building plumbing, can be replenished.⁴ Flushing also reduces water temperature, which is known to increase with age due to the indoor building temperature.⁴ These relationships are apparent in the data as DO, Free.Cl, and Total.Cl were each positively correlated with vol.events and num.events, and negatively correlated with meanTSL and maxTSL. Correlations of the opposite sign were noted for temperature. As further support, PC1 also exhibited loading factors of -0.23, -0.31, and 0.36, on DO, Total.Cl, and Temp, respectively.

4.4.2 Fate of Residual Disinfectant

Free and total chlorine were observed at each sample location. However, both forms of chlorine were often found to be below the method detection limit of 0.1 mg/L. Free and total chlorine were below detection limits in 44% and 27% of respective samples. These data were replaced with one half the detection limit (i.e. 0.05 mg/L) for analysis. A challenge encountered by monitoring a full-scale plumbing system was that more than 10% of the discrete water samples from the service line had no detectable chlorine. Meaning, the utility was delivering water of different quality than is typically expected. This degree of variation of water entering the service line has not previously been widely reported in the literature. Though, intensive water sampling as conducted in this study has not been previously reported in the literature either. Notwithstanding these challenges, free and total chlorine were strongly correlated (0.791) and correlations with other variables each had the same sign. Correlations with total chlorine were generally larger than those observed with free chlorine. The rate of chlorine decay is known to increase with temperature^{101,102} which was observed with negative correlations with total chlorine (-0.506) and free chlorine (-0.408).

Chlorine concentrations were negatively correlated to TCC, HPC, and Leg.sp. These results corroborate the susceptibility of the organisms which comprise these measures to chlorine disinfectants found in the literature.^{1,103} Additionally, Total.Cl was negatively associated with Leg.sp in the final GLMM and statistically significant at $p < 0.05$, demonstrating *Legionella*'s susceptibility to chlorine disinfection.^{1,7} Chlorine is known to react with organic matter to produce disinfection by-products (DBPs) such as TTHMs.¹⁰⁴ The reaction rate is affected by temperature^{38,102} and the concentration of available carbon.¹⁰⁵ TTHM production has been associated with elevated water age.^{4,14,102} Correlations suggest the same effects were present in these data. TTHM was correlated with TOC (0.652) and DOC (0.618), demonstrating a link between TTHM formation and available carbon. Correlations with free and total chlorine were much lower and negative, -0.334 and -0.288, respectively, supporting literature indicating that chlorine is consumed in forming TTHMs.¹⁰⁴ This same consumptive effect is not apparent in TOC and DOC, perhaps due to the relative abundance of organics and the relative lack of disinfectant, or that only a fraction of organics are reactive with chlorine. Results of these analyses show that TTHM is negatively correlated with num.events (-0.305) and vol.events (-0.470), positively correlated with meanTSL (0.227) and maxTSL (0.379), and has a loading factor of 0.300 on PC1, all which imply TTHM concentrations increase with water age.

PEX has been shown to leach carbon into water with time,^{82,99,100} implying organic carbon concentration may increase with water age. Leaching rates have also been positively associated with temperature.^{9,72} TOC and DOC are negatively correlated with vol.events and num.events, and positively correlated with meanTSL and maxTSL, implying that water use flushes TOC and DOC from the plumbing and resultingly that this carbon originates in the plumbing itself, whether through pipe leaching or biofilm sloughing. TOC's positive loading

factor of 0.325 on PC1 is supportive of this as well. Further, temperature is positively correlated with TOC (0.488) and DOC (0.526). TOC proved to be an influential predictor of *Legionella* in both evaluated general linear models. These results may be evidence that microbial growth was carbon-limited in the plumbing, which is common in drinking water,¹⁰⁶ or that the leaching of carbon from PEX allows TOC to act as an indicator of elevated water age.

4.4.3 Microbial Contaminants

TCC was correlated with HPC (0.701), HPC with Leg.sp (0.617), and TCC with Leg.sp (0.538). TCC and HPC also had positive parameter estimates in both GLMMs, indicating both are potential predictors of *Legionella* in this system. These relationships provide evidence that conditions favoring *Legionella* also favor higher concentrations of TCC and HPC. TCC, HPC, and Leg.sp are each expected to increase with water age.^{2,4,7,10} This is supported by correlation results and positive loading factors on PC1. Further, Bayesian methods showed that meanTSL, a proxy for water age, was most predictive of *Legionella* concentrations. Frequentist GLMM results included meanTSL with a statistical significance of $p < 0.05$. Bayesian linear modeling using BGLR suggests that meanTSL is the most likely contributor to *Legionella* concentrations, with the probability of a non-zero parameter estimate of $\geq 98\%$ for all three evaluated models. The probability of a non-zero estimate for all other parameters was $\leq 57\%$.

HPC and TCC had similar correlations to Temp (0.477 and 0.460, respectively), though Leg.sp had a weaker correlation (0.347). This is perhaps because *Legionella* spp. are better-suited to the low-temperature environments found in plumbing.⁸ TCC, HPC, and Leg.sp were each negatively correlated with vol.events and num.events, and positively correlated with meanTSL, and had positive loading factors on PC1 of 0.324, 0.368, and 0.368, respectively. These results corroborate previous findings that these constituents increase with water age.^{100,107}

TCC and HPC were each more strongly correlated to maxTSL than with meanTSL, though this relationship was reversed for Leg.sp. Further, GLMM results showed that Leg.sp increased with meanTSL. However, the association with maxTSL was weaker ($p < 0.1$) and inverted. Pathogens such *Legionella* spp. are relatively slow growing.⁶ These results suggest that consistent stagnation may be required for elevated concentrations of *Legionella* to develop compared to other measured microbes in this example water system. TCC, HPC, and Leg.sp had similar correlations with DOC, TOC, and Alka, implying these organisms favored similar conditions. Negative correlations with DO are stronger for HPC and TCC than with Leg.sp. The difference in these correlations may serve as evidence of selective pressures which favor *Legionella* in low-oxygen environments.⁸

Biofilm becomes detached for two primary reasons; cellular erosion due to shear stress induced by moving water, and bulk sloughing which may be induced by rapid changes in nutrient concentrations, temperature, shear stress, or a variety of other forces.^{22,23} Bulk sloughing is responsible for a majority of biofilm detachment.^{22,25} Correlation analysis shows that TCC, HPC, and Leg.sp are more strongly correlated with num.events than with vol.events, perhaps suggesting that the disruption caused by initiating each water use is more related to washout than the volume. DO exhibits a negative loading factor (-0.545) on PC2. Given the literature and the data, it appears that bulk sloughing may be occurring and that it may be driven by oxygen scarcity.

TCC, HPC, and Leg.sp are anticipated to be washed out from pipes due to biofilm detachment, causing a transient increase in bulk water microbial concentrations as the sloughed biofilm is carried out by the bulk water, followed by a period with lower microbe concentrations as the biofilm becomes reestablished. Negative loading factors on PC2 were identified for TCC

(-0.430), HPC (-0.283), and Leg.sp (-0.161). The weaker loading factor on Leg.sp may be indicative of *Legionella*'s adaptations to resist washout⁶. This sloughing of biofilm may also be associated with loading factors on Total.Cl (-0.143) and TTHM (0.176) as organic matter in detached biofilm is anticipated to consume chlorine and result in the formation of TTHM. TOC was presumed to increase with biofilm detachment but has a weak loading factor of 0.037 on PC2. The 14-day time period selected to calculate water use metrics appears to have captured a decreased microbial population resulting from the removal of biofilm but did not capture the slug of the biofilm itself as it exited the pipe network. Selecting a shorter time period may capture this as a temporary increase in microbial concentrations. Likewise, a longer time period may instead capture an increase in the same concentrations as the biofilm becomes well-established. Identifying the timeframe, which best corresponds with this increase resulting from biofilm sloughing was not attempted in this analysis.

4.4.4 Research Objectives

Data in this study were evaluated using multiple iterations of GLM, with both frequentist and Bayesian methods, to identify variables strongly related to *Legionella* spp. concentrations. Variable selection using Bayesian methods failed to identify an adequate subset of variables for *Legionella* prediction; however, both Bayesian and frequentist methods were used to develop parameter estimates with similar results. The top-performing GLM includes DO, HPC, TOC (each significant in the frequentist GLMM at $p < 0.001$), TCC ($p < 0.01$), Total.Cl, meanTSL, pH ($p < 0.05$), maxTSL, and num.events (not significant) as variables. Based on literature describing the influence of water age on *Legionella* concentrations,^{1,2,5} we anticipated that at least one water use metric would prove significant. However, PCA revealed that 39% of the total variance could be attributed to PC1, which was identified as water age. Additionally, although

maxTSL and num.events were determined to not be statistically significant in the top-ranked GLM, both were selected for inclusion in that model. Bayesian results for m.top indicated the highest probability of a non-zero parameter estimate for meanTSL at >99%. Taken together, these results illustrate the need for a more realistic indicator of water age.

This study also aimed to elucidate interactions which influence *Legionella* spp. concentrations. Analyses presented here associated increased water age with decreased concentrations of chlorine and dissolved oxygen as well as increases in temperature, TOC, DOC, TTHMs, TCC, HPC, *Legionella* spp., alkalinity, and pH through correlation coefficients, loadings on PCs, and GLMM fitting. Concentrations of *Legionella* spp. appear primarily driven by water age. The effect of water age is manifested in this data set as a combination of the water age metrics, where age is negatively related with num.events and vol.events, but positively related to meanTSL and max TSL. Further, PC1, which accounted for 39% of the total variance, appears to represent water age in PCA. Further analysis, including an accurate representation of water age, is necessary to further investigate and distinguish these effects.

4.4.5 Limitations of this Study

While the ReNEW house is a single family home, it differs from most homes in that it was specifically designed to reduce water and energy consumption. Water is heated in the home using waste heat from a roof-top photovoltaic system, which provides electrical power to the home. Water absorbs this waste heat in a series of large water heating. The combined volume of these tanks is much larger than would be expected for a typical residence. A thermostatic mixing valve then combines this heated water with unheated water to limit scalding risks throughout the plumbing. Additionally, the building experiences transient occupancy as it is used by students who live in the home during the school year. These factors may make generalizing results to

other homes difficult. However, the unique design elements at ReNEWW are expected to become more common as more water-efficient technologies are adopted. Further, focus on the analyses is placed on relationships between water quality characteristics. As such, this study was focused on conditions that contribute to OPP concentrations and not the water use that drive those conditions.

Another challenge is that water age cannot be directly measured, and instead, the water use metrics vol.events, num.events, meanTSL, and maxTSL were considered for analysis. While each of these metrics are expected to inform water age, it is important to recall that none consider the age of water in the plumbing . This is thought to have a relatively minor impact on the cold-water fixtures in the home due to the limited storage capacity of the water delivery piping (i.e., the system is effectively flushed with “new” water). While hot water delivery piping is similar, these pipes receive water from the water heating system rather than drawing water from the service line. The large storage volume (approximately 1,378 liters) and limited hot water use (mean 107.6 L per day) lead to a high hydraulic retention time (mean 12.8 days), which is unaccounted for in all four water use metrics. The variables meanTSL and vol.events were both used in PCA, but each has slightly different implications to water age. Given the importance of water age to *Legionella* spp. concentrations, an accurate, practical means of evaluating water age in a PPS is desirable. The development of such a model is described in the next chapter of this dissertation.

5 Chapter 5 – Water Age Modeling

5.1 Introduction

Water age, the time water spends in contact with a plumbing system, has been noted as a key factor influencing OPPP concentrations in **Chapter 4** as well as published literature.^{1,4,5,8} Water consumption rates have declined in recent decades.⁴⁸ However, this trend has not been addressed in plumbing design guidelines,³⁶ leading to increased water age.^{1,37} Guidance to limit the impacts of elevated water age include limiting stagnation (i.e. long periods of no flow or lack of water use) and implementing flushing protocols¹. However, this advice is often prescriptive due to the inherent heterogeneity of premise plumbing and as a result may not adequately address risks in all circumstances.

The goals of this work were to (i) produce a tool to determine the duration of time that water spends within the premise plumbing, referred to as water age, of a full-scale residential home; (ii) compare results to those produced by EPANET; and (iii) review the resulting water age to evaluate effects of water age on concentrations of *Legionella* spp. and common water quality variables.

This study relies on a wealth of data collected from the ReNEW house. Data include electronic water use records from September 2015 to May 2019, and detailed water quality analyses from sampling which ranged from October 2017 to October 2018. This data set is unique in scope due to the resources required to collect it. Instrumentation required to record flowrates cost approximately \$100,000 for a three-bedroom home and sample collection efforts required over 220,000 labor hours⁷². Water samples were collected 58 separate times from seven different fixtures in the house for a total of 406 samples, which were analyzed for a variety of common water quality variables. Flowrate data was recorded at one-second resolution from 18

flowmeters situated throughout the building plumbing as shown in **Figure 1**. The plumbing design and data collection efforts are described in additional detail in Salehi et al.^{71,72}, respectively. Simulating water age was preferred in this instance due to the labor requirements of a tracer study, availability of flowmeter data, and desire to determine water age over extended periods of time.

5.2 Methods

Plumbing configuration and water use data are utilized to simulate the flow of water through the premise plumbing. The storage volume of the entire plumbing network is divided into equal-volume (5mL) segments. Water flowrates are also discretized by the same unit volume. Flowmeter data is then used to direct each unit-volume of water through a series of states which represent the plumbing network using a plug-flow regime. Each unit of water is labeled with the time it is added to the plumbing from the service line. Analytical samples were collected from seven fixtures in the premise plumbing. Water age is calculated at each of these seven nodes as water exits the plumbing in the simulation. This simplified Eulerian approach eliminates the ability to model hydraulics, but also fundamentally eliminates the advective mixing error generated by EPANET. The process to simulate water flow is described in additional detail in the following sections.

Plumbing at the ReNEW house was surveyed to determine pipe length, pipe diameter, and to identify plumbing configuration. A piping and instrumentation diagram (P&ID) was developed using these results and is presented as **Figure 1**. The premise plumbing consists of ¾-inch and ½-inch nominal PEX-A style pipe, as denoted in **Figure 1**. The inner diameters of these pipes were assumed to be 1.73 and 1.23 cm, respectively. The volume of each pipe in the network was determined by data collected during a plumbing survey, and then expressed in

integer counts of 5 mL units. The volume of the water heater tanks was assumed to be part of the pipe immediately downstream in the plumbing model. For example, tanks 1-3 have a combined volume of 1227 L (324 gallons), which was incorporated into the volume of pipe segment T3-WH (see **Figure 1** Error! Reference source not found.). The upstream and downstream node of each pipe are also noted, which informs how water passes from one pipe to the next.

5.2.1 Water Age Model Development Process

Cumulative flow was discretized rather than instantaneous flow to ensure the full water volume passing through the pipe network was accounted for, regardless of choice of unit volume. The finest resolution of any flowmeter used in the premise plumbing was 5.9 mL. Thus, a 5 mL unit volume was selected to balance the need to capture flow with a fine resolution and with the computational complexity that arises from using a smaller unit volume.

Cumulative flow recorded by the 19 flowmeters in the pipe network were assigned to corresponding nodes as shown in **Figure 1**. Fixtures with no associated flowmeter (e.g., basement sink) were assumed to experience no water use. During the sampling period, residents of the home were asked to minimize water use at each of these unmetered fixtures, and their contribution to water use is thus assumed negligible. Further, it was assumed that the flowrate into the water heater never exceeded the sum of all monitored hot water use (i.e., water does not enter the heater unless there is equal downstream demand).

The cumulative flow was then calculated for each of the nodes not associated with a flowmeter. When the cumulative flow of all immediately downstream nodes was known, the sum of the cumulative flow of those nodes was taken as the cumulative flow of the upstream node. Likewise, when the cumulative flow of all immediately upstream nodes was known, the

cumulative flow at the downstream node was taken to be the sum of that flow, minus any flow directed to any alternative nodes.

A list is then prepared which contains a vector for each pipe segment in the network. Each of these vectors are to be populated by the units of water as they enter the pipe. Each unit of water that passes through a pipe is represented by a single element in that pipe's vector, the value of which is equal to the time that unit of water entered the premise plumbing. The order of each unit represented in these vectors is preserved to model a plug-flow regime. The water age of every unit water leaving the seven monitored fixtures is calculated as the difference between the time that unit exited that node and time it entered the building from the service line.

To initiate the simulation, the first n elements of the entry time vector for each pipe are populated with the starting time of the simulation, where n is equal to the pipe volume, in units of water. This has the effect of filling the pipes with water upon initiating the simulation. Due to the large volume and elevated hydraulic retention time (HRT) of the water heater tanks, the volume corresponding to those tanks began the simulation with an entry time equivalent to the mean HRT of the tanks (13.82 days). This was done to generate more realistic water age results for hot-water fixtures. This water with an assumed age is replaced as hot water is consumed in the house. Another vector is initialized to track the cumulative flow experienced by each pipe, which is used to identify water which is still available to be passed downstream. This tracker is used to essentially shift water forward in pipes to accommodate flow. Recall that water is not mixed during the simulation.

The model loops through each row of the cumulative flow matrix and identifies all nodes and pipes which experience flow during that second. Each second, the script then loops through each active node in the network sequentially, starting at the service line and moving downstream.

All upstream flow to be delivered to the node is identified by indexing the entry time vectors of those upstream pipes using the cumulative flow matrix. The volume of the pipe is included as an offset to identify only water which has passed through the full volume of the upstream pipe. Likewise, the cumulative flow experienced by the pipe is used as an offset to ensure only flow which is still present in the pipe is available to be moved downstream. When the active node is the service line, new water is added to the premise plumbing by adding a vector with length equal to the change in cumulative flow at that pipe at that second, equal to the current time. All upstream flow is combined as a single vector to represent the slug of water that passes through the node at that second. Next, if the node is in the list of monitored fixtures, water age is calculated as the difference between the current time in the simulation and the entry time of water at the node. Finally, the slug of water is then delivered to downstream pipes by directing elements of that slug to account for the volume of water moving through those pipes. An equivalent volume is added to the cumulative tracker, effectively shifting the water forward in the pipe. This process is then repeated for every node, and for each second of recorded data.

5.2.2 Comparison with EPANET

Water age was calculated using a simplified version of the ReNEWW plumbing network by both EPANET 2.2⁶⁰ and the novel model presented here to draw comparisons between the two. Water age was calculated using both methods for three cold water fixtures, cold kitchen sink (CKS), as well as cold bathroom sink (CBS) and the cold portion of mixed bathroom shower (MBS) (referred to in this section as CBSi and CBS_h, respectively). This simplified plumbing network eliminates all water use except from these three fixtures, as well as the hot water recirculation loop and thermostatic mixing valve, which are incompatible with EPANET.

EPANET calculates the water age at each fixture using a time step of one-minute. The water age model presented here, however, calculates the water age of each 5mL parcel to exit each fixture. To more directly compare the two models, the water age calculated by EPANET was assigned to each parcel based on the time of water use. This serves to weight the age of water by volume rather than time. Scatterplots, boxplots, and correlation coefficients were used to draw general comparisons between the two methods.

5.2.3 Variable Selection

Published literature highlights the importance of water age to the growth of pathogens such as *Legionella* spp.^{1,4} To gauge the importance of the results from the water age modeling presented here, the variable selection process implemented in **Chapter 4** was repeated with the inclusion of results from the water age model presented here to investigate its influence on water quality. Mean, median, and 95th percentile of water age were calculated for each of the 406 water samples collected. This summarized water age for two weeks prior to the collection of each sample. Correlation between variables was assessed using the Spearman correlation coefficient.

Principal component analysis (PCA) was conducted in R on the variables Temp, DO, Total.Cl, TCC, HPC, Leg.sp, TOC, Alka, TTHM, vol.events, meanTSL, age.mean, age.median, and age.95per. As in **Chapter 4**, the variables Free.Cl and DOC were excluded due to high correlation with other variables and the number of missing observations which limit the data which can be used in PCA. The suitability of PCA to efficiently reduce the dimensionality of this data set was confirmed with Bartlett's sphericity test.¹⁰⁸ The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA)¹⁰⁹ was also calculated, indicating an overall MSA of 0.72 which is considered "middling".¹¹⁰ These suggest that PCA offers value in dimensionality reduction on these data.

Generalized linear modeling (GLM) was conducted in an iterative fashion to exclude variables with low importance and high degree of missing data, as in **Chapter 4**. A specific description of this process is provided in the along with results. Bayesian generalized linear regression (BGLR) was also utilized to assess the significance of variables, including water age results from this model. BGLR was conducted using the BGLR library in R.⁹⁵ This analysis calculated the probability that each variable in the linear model has an associated non-zero parameter estimate, taken here as evidence the variable influences *Legionella* spp. concentrations.

5.3 Results

5.3.1 Water Age

A summary of the calculated water age at each fixture is presented in **Figure 3**. Results indicate that water age was lower at cold fixtures than hot. This is especially true of the service line, where water enters the building. As expected, the highest and most consistent water age was identified at the water heater. These results indicate that water age at the hot fixtures is typically highest at hot kitchen sink (HKS) followed by the hot bathroom sink (HBS). The ReNEWW house relies on a thermostatic mixing valve downstream of the water heater to regulate the temperature of water. Thus, a significant volume delivered to HKS and HBS bypassed the large heating system that feeds the water heater (WH), resulting in lower water age. The mixed bathroom shower (MBS) location is a shower fixture, where the user may change the proportion of hot and cold water. This is expected to have implications on water age, and indeed, 56.8% of water from MBS has an age of less than one week. Further, water use at MBS is generally over a prolonged period of time in comparison to other fixture types, thus flushing older water from nearby plumbing. This may help to explain the lower age of HBS relative to HKS. Water age

results, when evaluated with consideration of the configuration of the system, appear to explain general trends in observed water age.

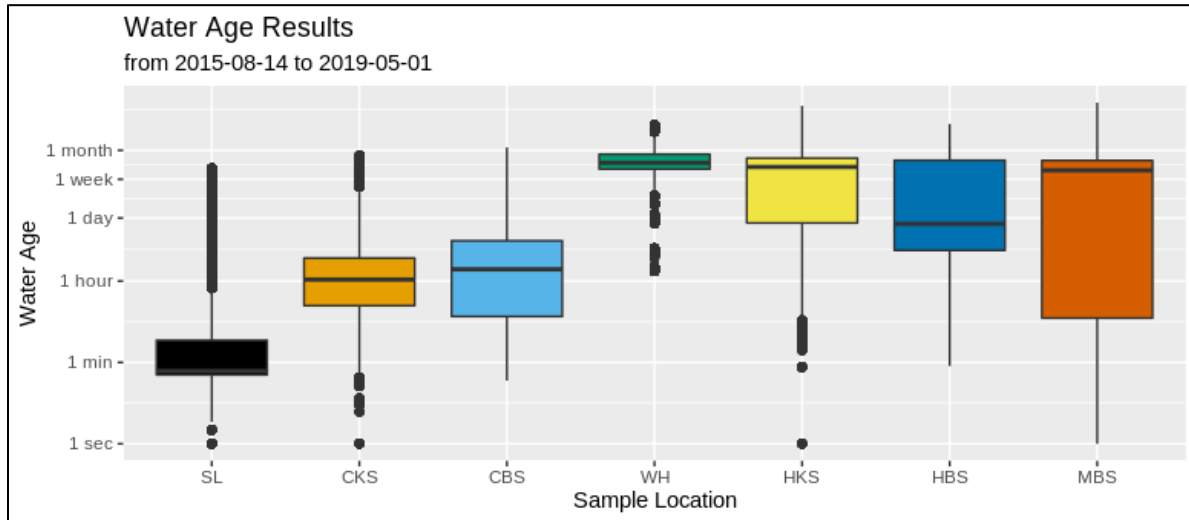


Figure 3 - Boxplot of water age results

Mean water age as calculated by this method (age.mean) was found to be similar to the mean time since last use (meanTSL) for the cold water locations, as shown in **Figure 4**. The cold water sample locations; the service line (SL), cold kitchen sink (CKS), and cold bathroom sink (CBS); exhibit a linear trend between age.mean and meanTSL. Spearman correlation coefficients between these two metrics were found to be 0.91, 0.86, and 0.73 were calculated for SL, CKS, and CBS, respectively. The hot water samples, however, do not visually follow this trend, and tend to have lower meanTSL than would be implied by a linear trend with age.mean. Spearman correlations for the hot water sample locations, WH, HKS, HBS, MBS are lower, at 0.70, 0.72, 0.60, 0.52, respectively.

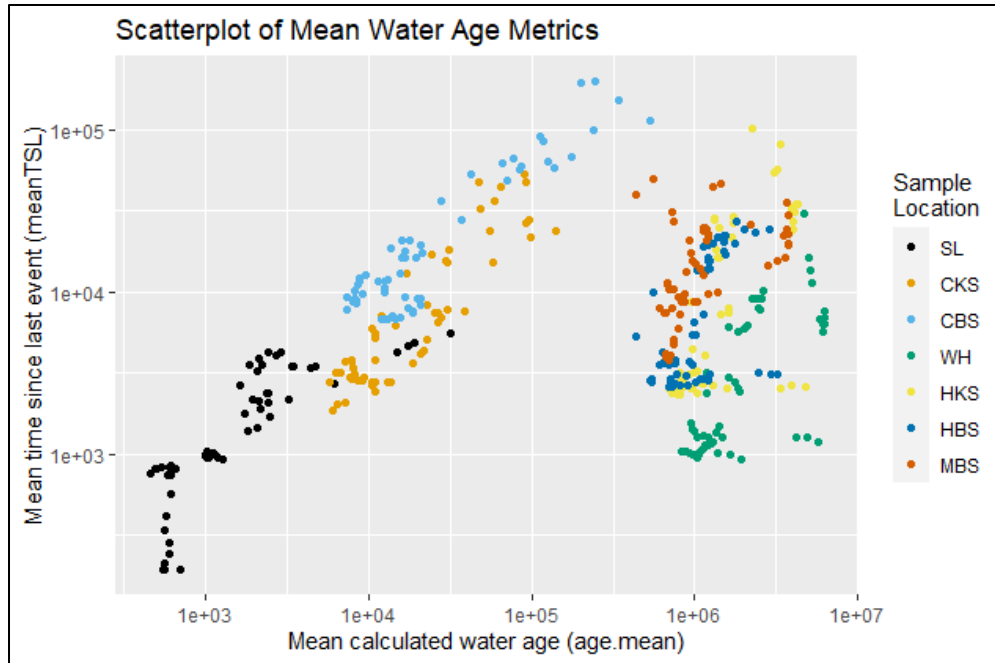


Figure 4 - Scatterplot comparison of the water age metric age.mean with the water use metric meanTSL. The color of each point indicates the location from which each sample was collected, as shown in the legend.

The water age model produced results which were more distinct between fixtures than use metrics. age.mean and meanTSL are compared by fixture in **Figure 5**. Water age model results show stratification between the cold and hot water sample locations, whereas substantial overlap exists across sample locations for the water use metric results (meanTSL). This again demonstrates the impact of water heater HRT on water age model results which is unaccounted for in each of the water use metrics (meanTSL, maxTSL, num.events, vol.events).

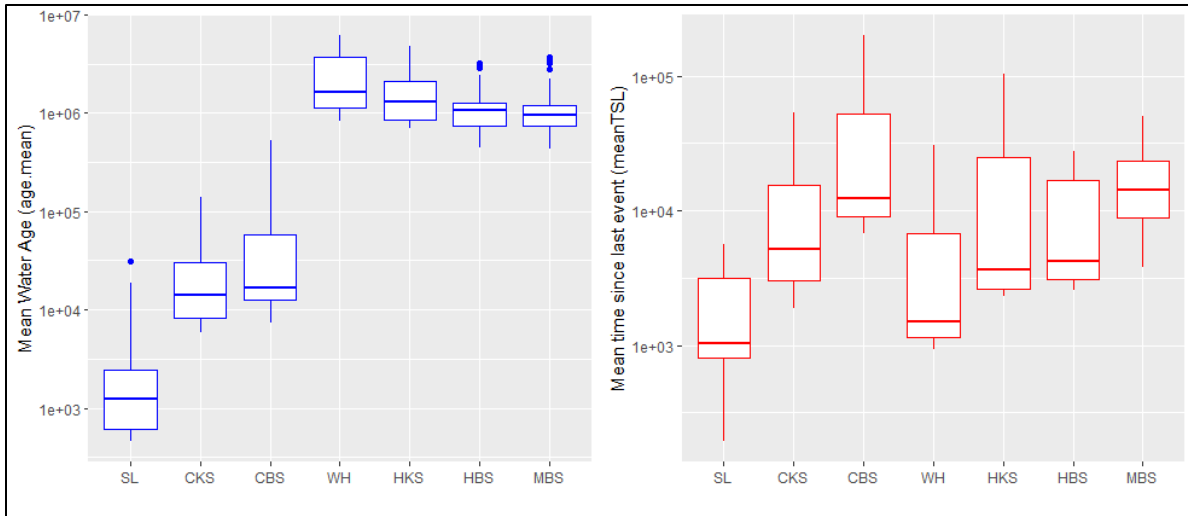


Figure 5 - Boxplots comparing the water age metric $age.mean$ (left) with the use metric $meanTSL$ (right)

Histograms of water age at each fixture sampled for water quality parameters are presented in **Figure 6**. Each location exhibits a unique pattern, though CBS and HBS appear to have a similar shape for water age below one week. This effect appears present between CKS and HKS as well, but to a much lesser extent. Recall that water use at HKS and HBS were assumed to be a ratio of use at meters for CKS and CBS, respectively. These similar use patterns appear to have manifested as similar patterns in water age. This seems to suggest that water use patterns, rather than the HRT of the upstream plumbing, has a more substantial influence on water age in this PPS.

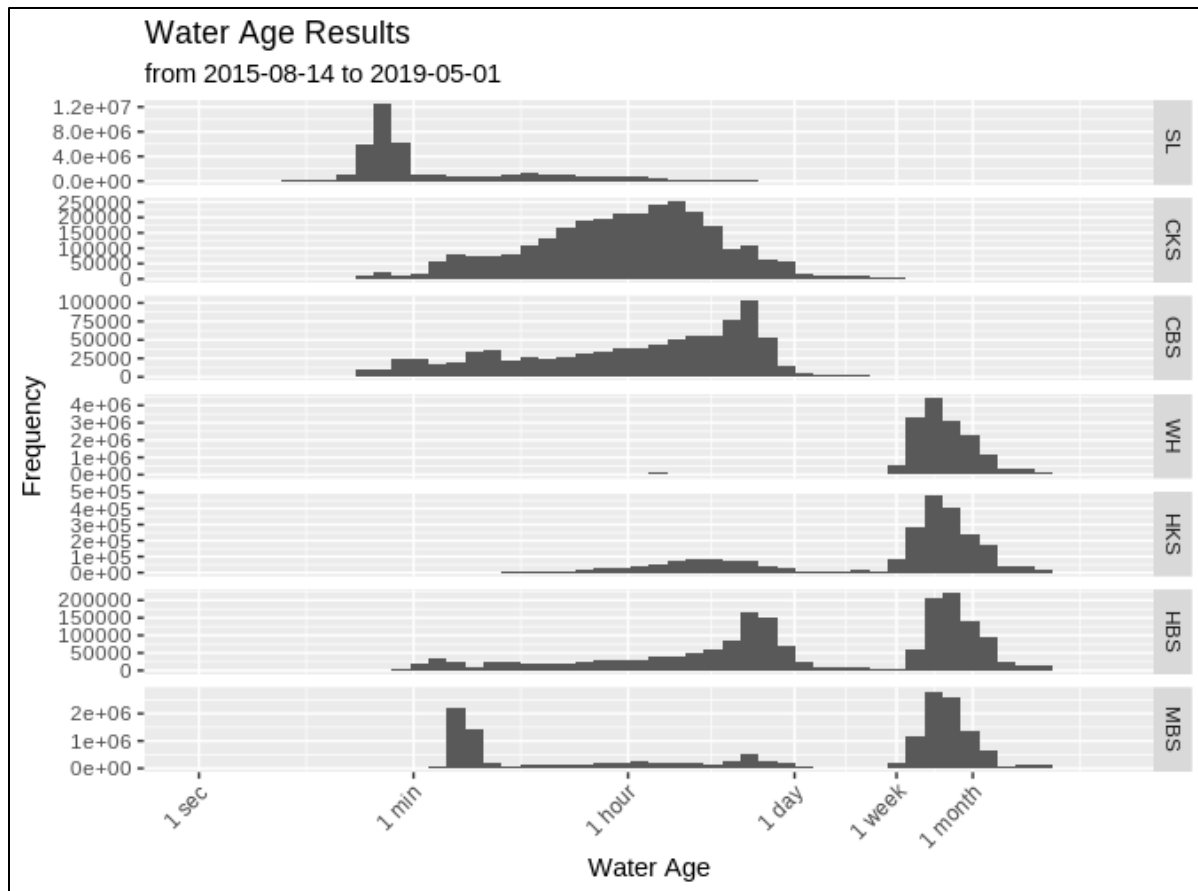


Figure 6 - Histogram of water age results by fixture

31.5% and 81.2% of water delivered to HKS and HBS respectively had an age of less than one week. This could be evidence of a problem with the assumption about cold water coming from the thermostatic mix valve whenever hot water is used by not metered at WH.

5.3.2 Variable Selection Including Water Age

5.3.2.1 Correlations

Spearman correlation coefficients were calculated for each pair of analytical and water use variables as in **Chapter 4** with the inclusion of mean, median, and 95th percentile of water age (referred to as age.mean, age.median, and age.95per, respectively) calculated during the same two week period prior to sample collection. All three metrics had the same sign for all correlations, and the magnitude of these correlations was generally higher for age.mean than for

age.median or age.95per. These water age metrics showed relatively weak correlations with the water use metrics evaluated in **Chapter 4**. For example, age.mean showed correlation coefficients of -0.156, -0.312, 0.330, and 0.379 with the cumulative volume (vol.events), number of water uses (num.events), mean time between water uses (meanTSL), and maximum time between uses (maxTSL), respectively, with each of these four metrics calculated over the same two-week period prior to sample collection. These weak correlations suggest that water age calculated in this paper represents unique data.

5.3.2.2 Principal Component Analysis

PCs one through three (i.e. PC1, PC2, and PC3) exhibited a standard deviation of greater than one. As such, these PCs were considered for interpretation using loading factors and variable relationships identified in published literature.⁸⁷ PCs one through three have a combined variance of 0.57, and are presented in **Table 7**.

Table 7 - PCA results for PC1 through PC3

Variable	PC1	PC2	PC3
pH	0.080	-0.370	-0.495
Temp	0.271	-0.240	0.181
DO	-0.116	0.402	0.404
Total.Cl	-0.225	0.292	-0.136
TCC	0.359	0.170	0.018
HPC	0.294	0.221	-0.024
Leg.sp	0.140	0.019	-0.473
TOC	0.028	-0.236	0.368
Alka	0.278	0.037	-0.031
TTHM	0.199	-0.328	0.315
vol.events	-0.227	0.348	-0.219
meanTSL	0.144	-0.208	-0.092
age.mean	0.396	0.248	0.004
age.median	0.363	0.259	0.066
age.95per	0.378	0.150	-0.140
Standard deviation	2.22	1.51	1.13
Proportion of Variance	0.33	0.15	0.08
Cumulative Proportion	0.33	0.48	0.57

The water age metrics age.mean, age.95per, and age.median had the highest magnitude loadings of any variable on PC1 at 0.396, 0.378, and 0.363, respectively. TCC and HPC have the next highest loading factors at 0.363 and 0.298. These results suggest that PC1 is related to water age. The sign of each loading factor aligns with expectations about how water age effects each variable. For example, literature suggests that with increasing water age, HPC concentrations and water temperature will increase and concentrations of residual chlorine and DO will decline.^{2,14,38,98,111} The water metrics vol.events and meanTSL are negatively and positively related to PC1, respectively, further associating PC1 with water age. Alka showed a loading factor of 0.278. A literature search was conducted to identify other instances relating alkalinity to water age in premise plumbing, however, no meaningful results were identified. Alkalinity may

be produced or consumed in some biological reactions.¹¹² While no direct mechanistic link between water age and alkalinity is identified here, it appears plausible that alkalinity could increase with water age as the result of a process such as denitrification. Additional investigation would be required to conclusively determine this relationship in the ReNEWW plumbing.

Loading factors on PC2 were of the highest magnitude for DO (0.402), pH (-0.370), vol.events (0.348). Each of these variables have documented impacts on biofilm detachment, either by weakening the structure of biofilm or by encouraging washout due to the increased shear stress caused by changes in water velocity.^{15,23} Leg.sp had the lowest PC2 loading factor at 0.019, which is perhaps evidence of *Legionella*'s adaptations to resist washout.^{6,83} As such, PC2 appears to be related to biofilm detachment, just as in **Chapter 4**. The water age metrics age.median, age.mean, and age.95per have lower-magnitude loading factors on PC2 at 0.259, 0.248, and 0.150). These weaker factors may be related to the time that it takes biofilm to become reestablished following biofilm washout. The highest-magnitude loading factors on PC3 were pH (-0.495), Leg.sp (-0.473), and DO (0.404). Weaker loadings were found for age.95per (-0.140), age.median (0.066), and age.mean (0.004), suggesting PC3 has little to do with water age. PC3 may then also be related to biofilm detachment, and appears to be more closely related with inducing structural failure of biofilm than with hydraulic washout.

5.3.2.3 *General Linear Modeling*

An initial GLM to predict Leg.sp was implemented using 16 independent variables; pH, Temp, DO, Total.Cl, TCC, HPC, TOC, Alka, TTHM, num.events, vol.events, meanTSL, maxTSL, age.mean, age.median, age.95per; and 222 useable observations. Combinations of these variables were reviewed using the 'dredge' function from the "MuMin" R library⁹¹. The second-order Akaike Information Criterion (AIC_C) was used to identify the best-performing

models. A total of 39 models with a difference in AIC_C (Δ AIC_C) of less than 2.0 from the top-performing model were identified in this way, which included the variables age.95per, age.mean, age.median, HPC, maxTSL, meanTSL, num.events, pH, TCC, TOC, and vol.events.

A second GLM was then implemented using only those 11 variables as predictors, allowing a total of 242 observations to be included. Variable combinations were again evaluated based on Δ AIC_C and the “dredge” function from the “MuMin” R library.⁹¹ The top performing model (m.top) incorporated age.95per, HPC, maxTSL, mean.TSL, num.events, pH, TCC, and TOC as independent variables to predict Leg.sp. A competing model (m.comp) was defined to include all variables which were included in models with a Δ AIC_C of less than 2.0, which included all 11 independent variables. Finally, two GLMMs were defined using the “glmTMB” library in R.¹¹³ These included variables defined in the top and competing models of the previous GLM, as well as a random effect to describe sample location. The top and competing models are defined as m.top and m.comp, respectively. m.top included age.95per, HPC, maxTSL, mean.TSL, num.events, pH, TCC, TOC, and sample location as independent variables to predict Leg.sp. In addition to the independent variables in m.top, m.comp also included age.mean, age.median, and vol.events as independent variables to predict Leg.sp.

m.top exhibited model convergence, homoscedasticity, and low-multicollinearity. m.comp, however, failed to converge. Fitted results for m.top are provided in **Table 8**. These results show that age.95per is statistically significant at the $p < 0.05$ level, that meanTSL is significant at $p < 0.01$, and HPC, maxTSL, TOC, and TCC were significant at the $p < 0.001$ level.

Table 8 - GLMM Results for m.top

Variable	Estimate	Std. Error	z value	Pr(> z)	Significance
(Intercept)	-9.30E-01	1.34E+00	-0.70	4.87E-01	
age.95per	8.56E-08	3.47E-08	2.47	1.36E-02	p < 0.05
HPC	2.64E-01	5.96E-02	4.44	9.20E-06	p < 0.001
maxTSL	-3.20E-06	7.20E-07	-4.44	9.03E-06	p < 0.001
meanTSL	8.72E-06	3.22E-06	2.71	6.66E-03	p < 0.01
num.events	3.24E-05	8.76E-05	0.37	7.12E-01	
pH	9.08E-02	1.56E-01	0.58	5.60E-01	
TCC	4.39E-01	1.06E-01	4.13	3.68E-05	p < 0.001
TOC	8.05E-01	2.30E-01	3.50	4.69E-04	p < 0.001

5.3.2.4 Bayesian Generalized Linear Regression

The BGLR library⁹² in R was utilized to estimate the probability that each variable is associated with a non-zero parameter in the linear model. The probability of a non-zero parameter, the parameter estimate, and the standard deviation are presented for both m.top and m.comp in Table 9. m.top showed a near 100% probability of a non-zero parameter estimate for age.95per, and a probability of 0.785 for maxTSL. The remaining parameters had probabilities ranging from 0.641 to 0.674. m.comp showed less certainty in any one parameter estimate, and instead the highest probabilities were associated with age.mean (0.843), maxTSL (0.711), age.95per (0.623), and meanTSL (0.610). The remaining parameters in m.comp had non-zero probabilities between 0.442 and 0.587. BGLR parameter estimates were closer to zero than GLMM parameter estimates for all except the larger parameter estimate for age.95per in m.top.

Table 9 - BGLR Results

	m.top			m.comp		
	Prob.	β	SD(β)	Prob.	β	SD(β)
age.95per	1.000	2.39E-07	3.85E-08	0.623	9.06E-08	3.38E-07
maxTSL	0.785	5.06E-07	5.69E-07	0.711	3.25E-07	4.58E-07
meanTSL	0.674	2.93E-07	7.18E-07	0.610	1.56E-07	5.42E-07
HPC	0.642	-4.57E-09	6.39E-07	0.581	-3.27E-10	4.55E-07
num.events	0.640	2.03E-09	6.32E-07	0.579	-4.95E-09	4.43E-07
pH	0.639	7.99E-09	6.44E-07	0.581	4.53E-09	4.47E-07
TCC	0.638	-3.60E-09	6.37E-07	0.587	-1.98E-09	4.47E-07
TOC	0.641	7.21E-10	6.37E-07	0.580	-2.47E-09	4.54E-07
age.mean	NA	NA	NA	0.843	2.89E-07	2.64E-07
vol.events	NA	NA	NA	0.580	-4.86E-09	4.46E-07
age.median	NA	NA	NA	0.442	-4.99E-08	3.67E-07

These results are markedly different than those presented in **Chapter 4**. Water age is highlighted as relevant in these results with high probabilities of non-zero parameters for age.95per and maxTSL in m.top and age.mean, maxTSL, and meanTSL in m.comp. This corroborates literature indicating water age is positively related to *Legionella* concentrations^{1,4}. Parameter estimates differ substantially between GLMM and BGLR results, and parameter estimates for maxTSL, HPC, and TCC even have different signs. It is hypothesized that including the water age metrics in this analysis allows the water use metrics presented in **Chapter 4** to primarily convey information about hydraulic disruption.

5.3.3 Comparison with EPANET

Water age was generated using both EPANET as well as the water age model presented in this paper. This comparison was conducted on data recorded between 1/22/2018 and 2/5/2018. A Spearman correlation coefficient of 0.666 was measured between the two methods. Water age results were generally lower for the water age model than for EPANET (**Figure 7**). Notably, the median of CBSH is much lower for the water age model (2.9 minutes) than EPANET (35.2 minutes). The low water age at this fixture is unsurprising, as showers can be expected to

operated continuously for a longer period than sinks, potentially consuming enough volume to completely flush the plumbing all the way back to the service line. The design flowrate of the shower is 7.6 LPM. When operating at this flowrate, the HRT of the cold water plumbing leading directly to this shower is 1.5 minutes. However, the shower flowrate is composed of both hot and cold water. Assuming an equal mix of hot and cold water, the HRT of the cold water plumbing to the shower is then 3.0 minutes. This HRT is nearly identical to the median age of CBSH calculated by the water age model presented here, and more than an order of magnitude below that calculated by EPANET.

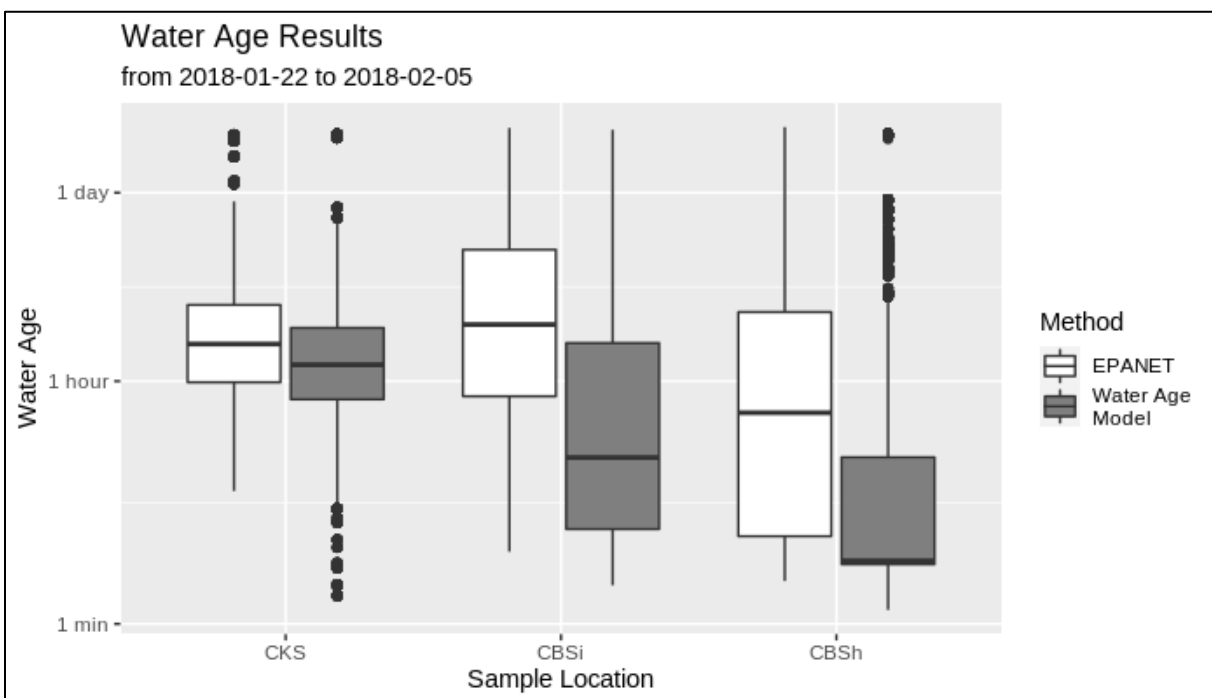


Figure 7 - Boxplot comparison of water age results for both models

5.4 Discussion

5.4.1 Water Age Results

Water age results appear plausible given the description of the building monitored by the data in this study. ReNEWW is unique due to its extensive water and energy efficiency measures

and expected to have greater water age than typical homes. Differences in water age appear primarily driven by water use patterns. Cold water fixtures exhibited substantially lower water age than of hot-water fixtures, primarily due to the large volume of the water heating system. This large volume causes water age in those hot water fixtures to be higher than would be expected in conventional homes. Calculated water age for hot-water fixtures is substantially higher than the water use metric meanTSL. This is expected, as the HRT of the water heating system is accounted for by age.mean but not by meanTSL, and illustrates that the water age metrics calculated by this model are more representative of actual water age.

Water age results from HKS and HBS are suspect due to the interference between the assumptions that hot water use at the sinks is a ratio of the cold, and that water use at hot fixtures that exceeds that metered at the water heater must bypass the heater via the thermostatic mixing valve. These assumptions interact, in that assumed hot water use is unlikely to occur at the exact same time as actual water use, leading the simulation to route water through the thermostatic mixing valve to make up the appropriate volume at the fixture. This effect appears to primarily affect water age at HKS and HBS, however, it is important to note that this could drive water age down within all hot water plumbing upstream of these locations. Additional study with complete flowmeter data would be useful in evaluating the impact of these interfering assumptions.

5.4.2 Comparison with EPANET

Substantial variations were found between EPANET and the model presented here. Water age calculated by the water age model was generally lower than that by EPANET for all three fixtures. Results show these two methods are most similar for CKS, with larger differences in the bathroom fixtures. Notably, the median of CBSH is much lower for the water age model (2.9 minutes) than EPANET (35.2 minutes). The low water age at this fixture is unsurprising, as

showers can be expected to operate continuously for a longer period than sinks, potentially consuming enough volume to completely flush the plumbing all the way back to the service line. Assuming the plumbing between the shower and service line are completely flushed during a typical shower, the HRT of the plumbing between the service line and shower is 1.45 minutes at the 7.6 LPM design flowrate of the shower. While the cold water makes up only a fraction of the water delivered to the shower, this HRT helps demonstrate much of the water delivered to the shower is expected to have a lower age than EPANET results suggest.

The difference in methods stems from water mixing within pipes, which is modeled only in EPANET, where water from inactive pipes becomes mixed with water flowing through pipes actively in use. However, these results should cause only local variations in age, as any mixing that results in lower water age in one part of the plumbing would act to increase age in another. These results indicate EPANET resulted in higher water age for all three monitored fixtures.

5.4.3 Variable Selection

Microbial variables were more closely correlated with water age metrics than with the water use metrics. For example, age.mean had correlations of 0.513, 0.653, and 0.627 with TCC, HPC, and Leg.sp, respectively. The magnitude of these correlation coefficients with meanTSL was lower at just 0.225, 0.412, and 0.391, respectively. TOC and DOC were more highly correlated with use metrics like meanTSL (coefficients of 0.528 and 0.535) than with mean water age (coefficients of 0.514 and 0.472). This may suggest that TOC and DOC concentrations are more influenced by the dynamics of water use (i.e. flushing pipes) than by water age itself. However, the differences between these correlations are small. Free and total chlorine exhibited larger magnitude correlations with age.mean (-0.368 and -0.336) than with meanTSL (-0.210, -

0.065). However, correlations with temperature, carbon concentrations, and microbial constituents remain highly relevant for chlorine.

PCA results here are similar to those presented in **Chapter 4**, with little change in interpretation of the PCs. Water age metrics appear to have taken on the primary role as indicators of water age, with water use metrics (e.g. meanTSL, vol.events) perhaps representing information more related to hydraulic disruption and its effects on biofilm detachment.

GLMM results show that age.95per as significant ($p < 0.05$) to *Legionella* concentrations, as well as HPC, maxTSL, TCC, TOC (all at $p < 0.001$) and meanTSL ($p < 0.01$). The inclusion of the water age metrics developed here eliminated DO, Total.Cl and pH as significant variables, and added maxTSL and age.95per. These results are more representative of literature on the topic as water age, HPC, and carbon concentrations have been correlated with *Legionella* concentrations. However, no direct relationships showing DO or pH were identified in the literature. Concentrations of residual disinfectant are indeed related to *Legionella* concentrations,^{1,28,114} but the concentrations typically observed during this study period, 95% of which ranged from below detection limits (0.1 mg/L) to 0.71 mg/L, may be low enough to not effectively inhibit *Legionella* growth¹. These results expand upon the findings of **Chapter 4** by further supporting water age as a critical to modeling *Legionella* spp. concentrations and allowing less-important factors such as DO or pH, to be excluded from the selection process. No literature was identified that supports a direct mechanistic relationship between *Legionella* spp. These results, including metrics for water age, better corroborate the literature by identifying only variables that have been previously identified as being related to *Legionella* spp.^{1,4,9,77,97,99} Further, BGLR results show that age.95per and age.mean are more better predictors of *Legionella* spp. concentrations than water use metrics in m.top and m.comp, respectively.

5.4.4 Conclusions

The novel water age model presented herein results in plausible estimates of water age for the full-scale, water-efficient home in this study. Water age results exhibit merit in prediction the concentration of *Legionella* spp. due to higher correlations with the measured variable Leg.sp and loading factors on PC1 than water use metrics. GLMM results also suggest that age.95per is a significant variable in predicting Leg.sp. While age.95per is not more significant than the use metric meanTSL, PCA results suggest this could be due to an alternative interpretation of meanTSL as having to do with hydraulic disruption, which is important to biofilm detachment and *Legionella* washout. As such, the water age model provides benefit in predicting *Legionella* spp. concentrations. Validating these results as an accurate means of describing water age would require additional research.

5.4.5 Limitations

This novel method to determine water age in a residential PPS provides several advantages over existing methods to estimate age, such as tracer studies or hydraulic modeling software designed for WDSs. However, several limitations exist including the inability to model hydraulic pressure and its effects, in-pipe mixing, and discretization errors arising from using a unit volume. Further, recall that water use at the hot sink fixtures was assumed to be a ratio of the use at corresponding cold fixtures (**Appendix**). This assumption causes the timing of hot water use to become misaligned with that recorded by the flowrate into the water heater. Further, a thermostatic mixing valve is present in the plumbing, allowing a portion of water to bypass the water heater. Because the flow at HBS and HKS was assumed from CBS and CKS, respectively, this flow will not have been recorded as passing through the water heater, and is thus assumed to be delivered through the thermostatic mixing valve's bypass line, resulting in lower water age

than would have been observed if those flowmeters had been operational. Resolving these issues would require additional metering and falls outside the scope of this work.

6 Chapter 6 – Assessing Compliance with Thermal Guidance

6.1 Introduction

Published literature,^{1,4,5,36} as well as the results presented in **Chapter 4**, indicate that water temperature has a significant effect on *Legionella* spp. concentrations. Past studies have investigated the effect of water temperature, as measured during sample collection, on *Legionella* spp. concentrations.^{71,84} However, using only the temperature at end use ignores the temporal thermal profile experienced by the water as it moved through the plumbing. Literature has shown that water temperatures fluctuate while stored in pipes between uses, tending toward the ambient temperature of the building.³⁸ Accounting for the water temperature during stagnation is anticipated to better inform effluent concentrations of *Legionella* spp.

Current guidance suggests that unheated water be kept below 20°C, that heated water achieve a temperature of no less than 60°C, and that heated water is delivered for end use at no less than 55°C.^{1,28,39} This ensures that water spends little time at *Legionella*'s ideal growth temperature range, approximately 25°C-43°C.^{1,7} Maintaining water temperatures outside of this range is expected to limit the potential for *Legionella* growth in PPSs by limiting growth in cold water (<20°C), achieving elevated water temperatures sufficient to inactivate *Legionellae* during heating (≥60°C), and maintaining sufficient temperatures (≥55°C) throughout distal plumbing to prevent regrowth. However, these guidelines do not account for the cooling or heating of water stored in plumbing between uses.

Data collected from the ReNEW house includes measurements of flowrate and temperature from flowmeters and thermocouples situated throughout the plumbing as shown in **Figure 1**. Further, samples were collected while electronic data were being monitored, and analyzed for *Legionella* spp. concentrations. These data provide a unique opportunity to evaluate

the efficacy of adhering to common temperature guidelines on resulting *Legionella* spp. concentrations, and is the first such study conducted at full-scale. The goals of this study were to develop a model to evaluate compliance with thermal guidance to limit *Legionella* proliferation by considering the temperature of water samples prior to sample collection.

6.2 Methods

The water age model presented in this dissertation was modified to evaluate whether each 5 mL parcel of delivered water complied with thermal guidance designed to limit *Legionella* spp. proliferation. In the initial water age model, presented in **Chapter 5**, a list is prepared, with a vector for every pipe segment, to track the entry time of each water parcel. This list is then used to determine the water age of each parcel as the difference in time from entry into the home to exit at end use.

To develop the temperature model, a second list containing a vector representing each pipe segment, much like the list for water age, was created to track temperature compliance. Like the water age list, each element represents a 5 mL parcel of water. However, instead of being populated with the entry time, each element contains a Boolean value, true if the parcel complies with temperature guidance, and false if it ever violates that guidance while retained in the plumbing system. There are 11 thermistors located through the plumbing, whose approximate locations are shown in Figure 1. Each pipe in the plumbing was assigned a temperature zone, as shown in **Figure 8**, by considering the proximity to the nearest thermistor, the flow direction, and fixtures which generally consume the most water. While temperature was recorded at 1 second intervals, the mean temperature every three minutes was taken as the temperature instead to reduce the impact of momentary deviations in temperature and noise, as well as to reduce the computational complexity of the model.

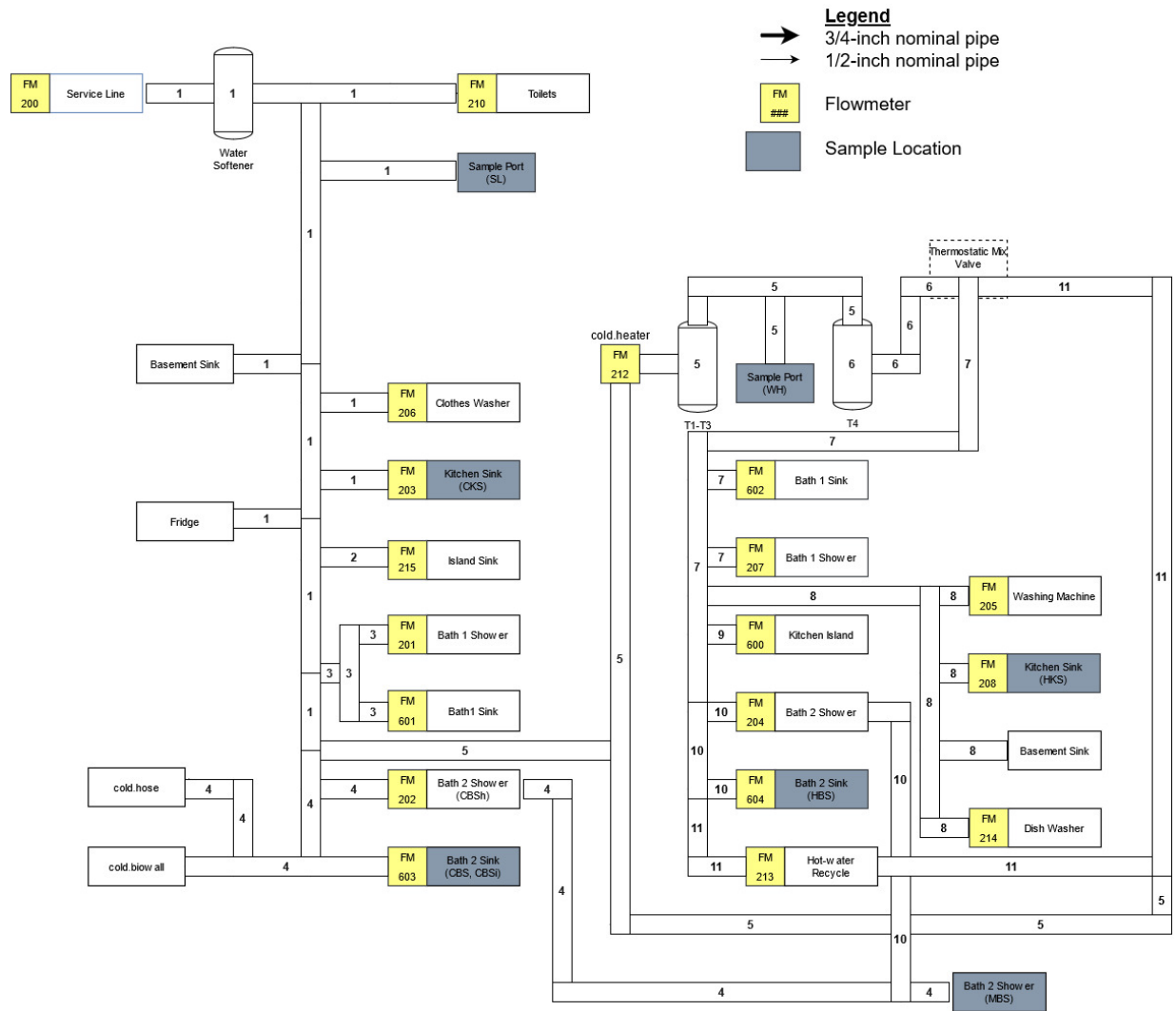


Figure 8 - ReNEW P&ID with temperature zone information. Each pipe is labeled with a number indicating the temperature zone, with each number corresponding to the thermistors depicted in Figure 1.

Parcels of water are directed through the plumbing using the methods described in the water age modeling chapter of this dissertation. This movement is applied to both the lists of water age entry times and of Boolean temperature compliance. Upon entry to the home from the service line each parcel is labeled as true if the water temperature in zone 1 is below 20°C, and false if above. Each time a parcel is shifted forward, it is evaluated for temperature compliance. By checking the three-minute mean of the thermistor associated with the temperature zone that parcel is currently in. If that temperature is between 20°C and 55°C, the parcel is marked false. If

that temperature is above 60°C, the parcel is marked true. If that temperature is either below 20°C or above 55°C, the parcel's value is left as-is. Thus, each parcel exits the system marked true if it either remains below 20°C or achieves a temperature of at least 60°C and maintains a temperature of at least 55°C before end use.

Water samples were collected from the ReNEW house from 10/10/2017 to 10/9/2018. Water use and temperature data from 9/1/2017 through 11/1/2018 were input to the model to calculate temperature compliance over the sampling period, with a period analyzed prior to sample collection to flush the PPS of water whose age and temperature compliance status were assumed at the start of the simulation. Temperature compliance was associated with each collected sample by taking the mean compliance of all parcels delivered to that location during the two-week period prior to sample collection, referred to here as compliance rate. Parcels which complied with guidance took on the value 1.0, with non-compliant parcels equal to 0.0.

6.3 Results

The mean compliance rate over the entire evaluated period was calculated for each fixture, as shown in **Figure 9**. The service line (SL), where water enters the home, showed a compliance rate of 96%. The cold kitchen sink (CKS) had a higher compliance rate (95%) than the cold bathroom sink (CBS) (60%). Compliance generally decreased for downstream sample locations, as expected, with cold water more compliant than hot. However, the water heater (WH) exhibited higher compliance (91%) than CBS. This is a result of water temperature exceeding 60°C at the WH, which resets the compliance variable. Compliance decreased for fixtures downstream of the WH at the hot kitchen sink (HKS), hot bathroom sink (HBS) and the mixed bathroom shower (MBS) at 18%, 8%, and 0.7%, respectively.

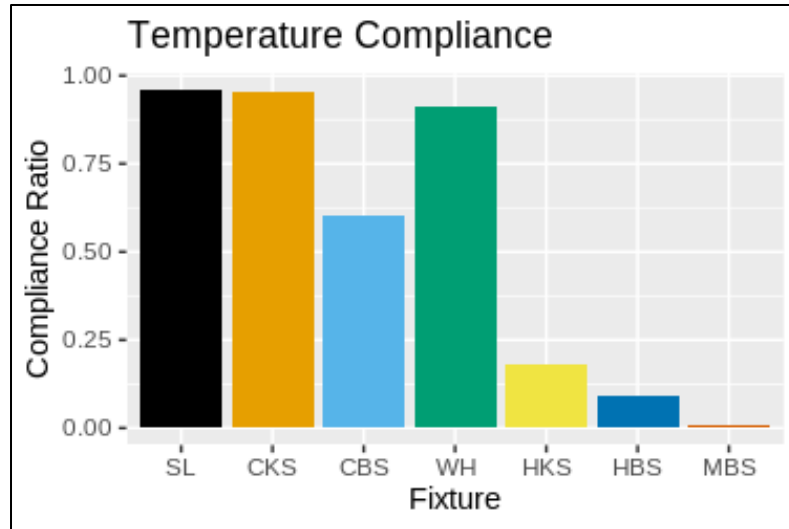


Figure 9 – Mean temperature compliance by fixture from 9/1/2017 through 11/1/2018

Temperature compliance was also compared with *Legionella* spp. concentrations observed in the samples collected at the ReNEW house. A Spearman correlation of $\rho = 0.22$ ($p < 0.001$) was measured between these data. These data are presented in **Figure 10**. Visual inspection suggests that *Legionella* spp. concentrations remain relatively low ($< 10^4$ GC/100 mL) when temperature compliance is held above 30%. However, some samples, especially from WH and MBS show high compliance ($>80\%$) as well as high concentrations of *Legionella* spp.

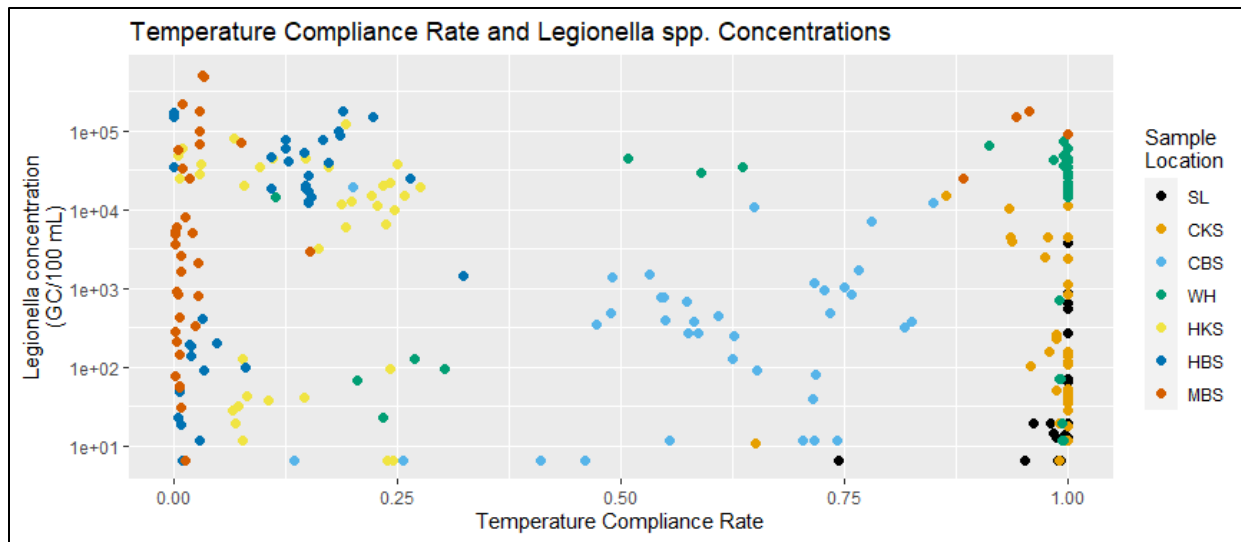


Figure 10 - Scatterplot of mean temperature compliance during two-week period preceding each sample and measured *Legionella* spp. concentration. The color of each point represents the location from which that sample was collected.

6.4 Discussion

Results of this model show substantially higher compliance with temperature guidance in cold-water than in hot. These results are unsurprising, as it is known that hot-water temperatures are known to cool to the ambient temperature of the building given time.³⁸ Hot-water sample locations exhibited very low compliance rates with thermal guidelines, especially MBS with only a 0.7% compliance rate. Each of these hot-water fixtures are downstream of the water heater, with relatively long (>9 meters) section of uninsulated pipe between the heater and each end use. Results showing poor compliance with temperature guidance reflects the problematic cooling which occurs, especially in longer pipes and between water uses. The low compliance rate at MBS may also result from the longer period between shower uses compared to sinks, and potential heat transferred from heated to unheated water at the mixing valve of the shower. The low compliance at MBS may present human health concerns as water used in showers is aerosolized, presenting inhalation risks for the user.

Especially for hot and mixed water samples, it is important to consider that *Legionella* spp. is measured using qPCR. This test method quantifies concentrations of genetic material unique to the analyzed organism (in this case, the 23s gene of *Legionella* spp.) by identifying RNA segments present in the sample. Achieving sufficient temperature to inactivate *Legionella* is indeed expected to prevent additional growth, but does not necessarily destroy the *Legionella* RNA already present in the water. As such, samples from heated water fixture may contain limited concentrations of viable bacteria. Further investigation is required to interpret these results and their influence on concentrations of *Legionella* spp.

7 Chapter 7 – Conclusions

The research presented in this dissertation explores factors in premise plumbing systems which support the growth and proliferation of the opportunistic pathogen *Legionella* spp. Multiple variable selection techniques were applied to a rich, novel data set to uncover relationships associated with increased *Legionella* spp. concentrations. These findings prompted the development of a novel water age model used to quantify the hydraulic residence time, also referred to as water age, experienced by water in the ReNEWW house, a full-scale water-efficient home. This model was then expanded to determine whether water adhered to common thermal maintenance strategies designed to limit *Legionella* growth in premise plumbing.

Each of these contributions to the science were performed to better understand the concentrations of *Legionella* spp. in the water of PPSs. A better understanding of the concentrations of pathogens such as *Legionella* in PPSs is critical to assessing and reducing associated human health risks.

7.1 Implications for Quantitative Microbial Risk Analysis

QMRA has been used in previous literature to assess OPPP risks.^{63,64,115} The basic approach used in Hamilton et al.⁶³ to describe *Mycobacterium Avium* Complex (MAC) risks was modified to quantify risks human health risks from *Legionella* spp. A distribution of *Legionella* spp. concentrations was developed from Filipis et al.,¹¹⁶ and dose-response data was assumed from Hamilton et al.⁶⁴ Exposure dose and risk were each calculated for showering, for both conventional and low-flow showerheads. Aerosol generation for these fixtures was modeled using data from O'Toole et al.¹¹⁷ Three styles of toilets were also evaluated to assess inhalation risks from toilet flushing. Aerosol generation for these three toilets was adapted from Johnson et

al.¹¹⁸ MAC concentrations were assumed from Donohue et al.¹¹⁹ to additionally evaluate MAC risks for each water use.

Results from that model are summarized in the **Supplemental Information**, and show that both exposure dose and resulting risk are highly correlated to the concentration of *Legionella* spp. For example, exposure dose and risk estimated during a low-flow shower show Spearman correlations with *Legionella* spp. concentrations of 0.99 and 0.96. Similar correlations were found in the conventional shower and toilet flushing results. These results highlight the importance of pathogen concentrations of potable water in PPSs on determining human health risks. The research presented in this dissertation aims to advance the science regarding understanding of *Legionella* spp. concentrations in building water systems, and is expected to prove useful in improving in the capacity for more case-specific QMRA related to premise plumbing systems.

7.2 Limitations of These Studies

Despite care and best efforts to address gaps in data and methodology, several limitations regarding the research presented herein persist:

- Several observations of the measured water quality analytes (e.g. free chlorine, DOC) were missing from this dataset. A more complete dataset is likely to improve the confidence of these findings.
- Data for two flowmeters used in this study (located at HKS and HBS) were shown to intermittently record implausible flowrates, leading to those data being removed from analysis and replaced with a ratio-multiple of the cold water metered at those locations.

- The plumbing configuration at the ReNEWW house allows unheated water to bypass the water heating system and is added at a thermostatic mixing valve downstream of the water heater to reduce the risks of scalding users. However, this bypass did not include a flowmeter, and was instead flow through this piping was assumed as the difference between the sum of all hot water consumed and the water passing through the water heater. Adding a flowmeter to this bypass line would have eliminated the need to make assumptions about the proportion of water bypassing the heater, and would increase confidence in water age and temperature modeling results.

The following conclusions are drawn from the body of research presented in this dissertation:

- Water age is critical: Water age is consistently identified as a significant predictor of *Legionella* spp., whether using water use metrics or modeled water age. As such, water age should be considered when inferring *Legionella* concentrations from alternative data to directly-measured *Legionella* counts.
- Analytical variables of importance: The water quality variables total chlorine, total organic carbon, and the alternative microbial metrics total cell count and heterotrophic plate count, were found to have significant impacts on *Legionella* spp. concentrations. However, these variables alone are not sufficient to produce realistic predictive estimates of *Legionella* spp. concentrations.
- Modeled water age predictive of *Legionella*: Water age as modeled in this dissertation is more indicative of residence time experienced by individual parcels of water than metrics describing stagnation between uses, especially for hot-water

fixtures. This modeled age incorporates the HRT of the water heating system, making it a more plausible descriptor of water age. Results show modeled water age as a significant predictor of *Legionella* spp. concentrations, supporting existing literature,^{1,5,36} and suggesting an increased confidence in these results.

- Compliance with temperature guidelines significant to *Legionella* spp.:

Compliance with common temperature guidelines,^{1,28,39} as determined by the novel model presented herein, are statistically significant to *Legionella* spp. concentrations.

8 Chapter 8 – Future Research

Complex interactions between water quality parameters, water use patterns, and water temperature have been demonstrated to influence *Legionella* concentrations. This dissertation explores a rich, novel data set and presents a foundation by which to make comparisons between these impacts. However, much work is yet to be done to quantify the effects of these changing conditions on pathogen concentrations and ultimately, to human health risks. The following research would help advance the science towards these goals:

- Compare findings from the ReNEW house with alternative premise plumbing systems:

Premise plumbing systems are inherently heterogeneous due to countless differences in water use patterns, fluctuating temperatures, varying influent water quality, plumbing configuration, pipe materials, etc. While the relationships identified in this research are largely corroborated by published literature, the interactions between effects and relative impact of each on *Legionella* concentrations could be different in alternative plumbing systems. Comparing results from similar full-scale research other premise plumbing systems will help to generalize these findings.

- Investigate the predictive value of metals: Concentrations of metals have been linked to concentrations of *Legionella* in previous research.⁷⁴⁻⁷⁶ These data were collected from the ReNEW house, but were not included in this analysis due to excessive observations below detection limits. Conducting further analyses after accounting for the limited detection of these metals could help to identify additional factors associated with increased *Legionella* spp. concentrations.

- Validate water age model: Results for water age were presented based on flowmeter data. Conducting additional study, including a tracer study, could be used to validate these results and explore the effects of in-pipe water mixing.
- Extend thermal modeling to address *Legionella* growth in biofilm: Modeling to assess compliance with temperature guidance to limit *Legionella* proliferation was conducted using parcels of water as the base unit, which is thus focused on the bulk-phase of water. However, *Legionella* growth occurs primarily in biofilm.¹⁵ Developing a temperature compliance heuristic based on the temperature of the spatially-fixed pipes could prove useful in predicting concentrations of *Legionella* measured in samples.
- Investigate the influence of biofilm detachment: Biofilm sloughing events distribute biofilm previously attached to pipe walls into potable water. These discrete sloughing events are often precipitated by changes in water quality (e.g. DO or pH),¹⁵ and are expected to significantly contribute to the concentrations of *Legionella* spp. and other such pathogens at the tap. Future research investigating the frequency and quantity of sloughed biofilm, especially considering the factors which drive sloughing, would improve the scientific understanding of *Legionella* spp. variability and concentrations at end use.
- Simulate water quality and its effects on *Legionella* risks: Future research should be conducted to simulate water quality and its influences on *Legionella* prevalence and growth in PPSs. The relationships investigated in this dissertation can be used to inform and calibrate simulations of *Legionella*. The results of these simulations may be passed to QMRA models which determine health risks based on the anticipated use of each fixture. Conducting this work at full-scale will inform risk management strategies by highlighting

plumbing configurations, use patterns, or water quality conditions that most contribute to *Legionella* risks.

APPENDIX

During data analysis, it was discovered that some flowmeters recorded data in excess of what was plausible. Data from each meter was subsequently reviewed to determine the ratio of the total flow received at a rate less than 15 LPM over the cumulative volume recorded by the flowmeter. This check ensured that observed flowrates fell within a plausible range (0-15 LPM). While this range is not plausible for all fixture types due to differing design flowrates, it provides a consistent basis to evaluate all flowmeter data. This ratio was > 0.999 for all but three flowmeters; cold.heater (0.976), hot.kitchen.isl (0.030), and hot.bath2.sink (0.287). Because cold.heater captures the flowrate into the water heater, it may occasionally experience water use from a combination of hot-water fixtures simultaneously. The instantaneous flowrate from cold.heater over the entire span of data was reviewed visually, and no indications of noise were identified. Flowmeter data from the hot kitchen island sink included intermittent signals indicating far higher flowrates than were plausible at the fixture. Similar issues were identified at the hot bathroom 2 sink flowmeter.

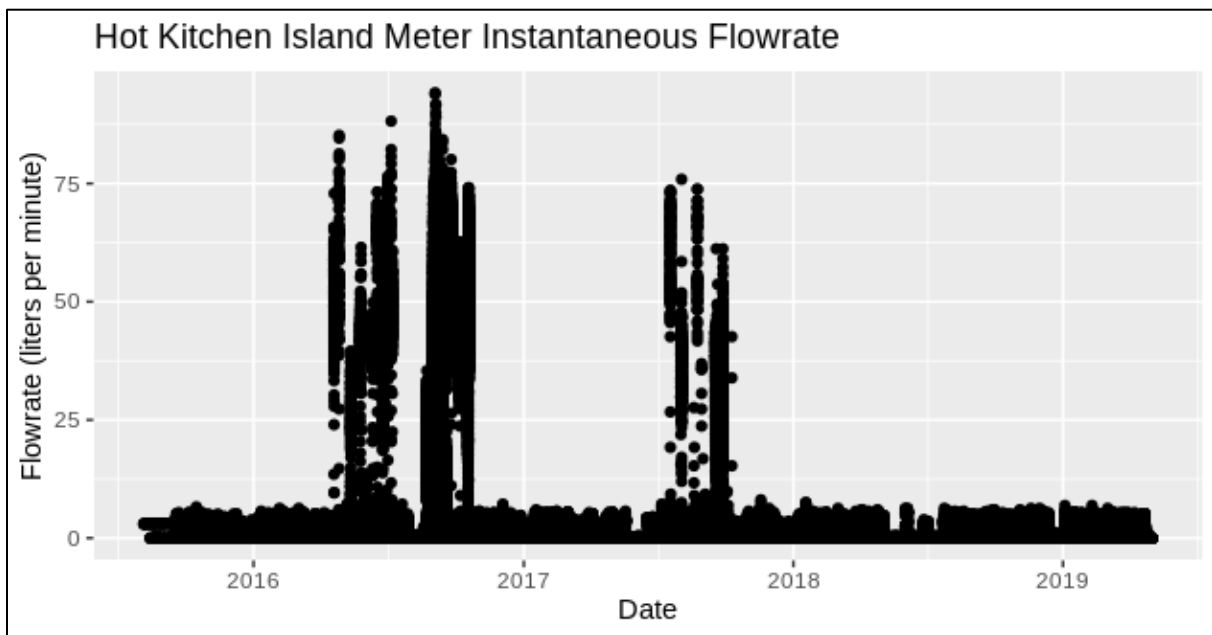


Figure 11 - Raw instantaneous flowrate at hot.kitchen.isl

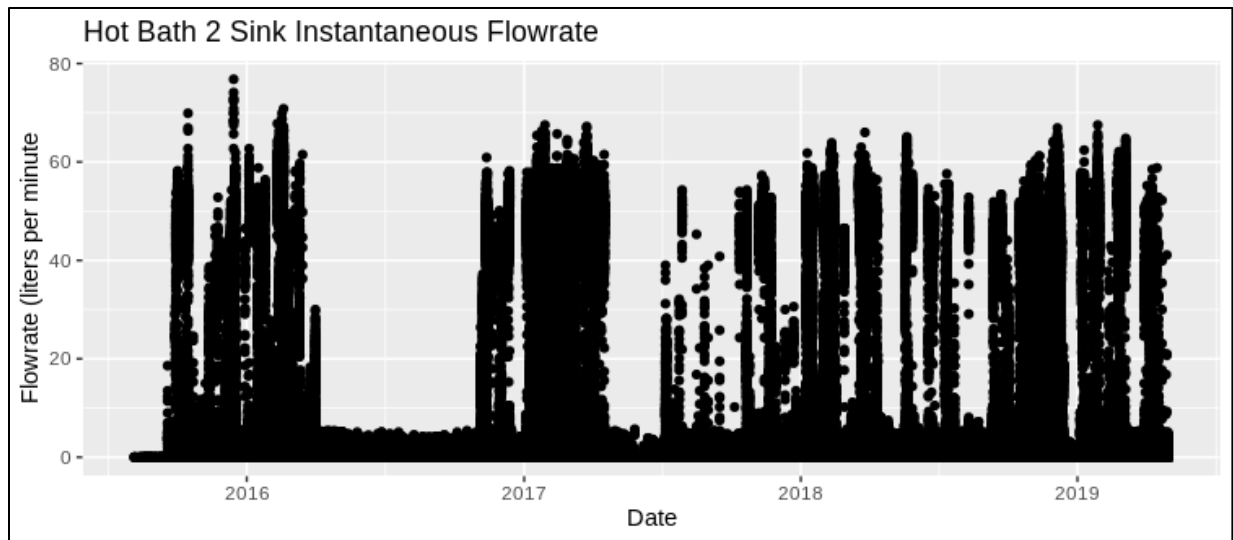


Figure 12 - Raw instantaneous flowrate at hot.bath2.sink

Both of these flowmeters were installed on hot water lines. Hot water that is left stagnant in pipes between uses cools toward the ambient temperature of the house. As water is consumed at a hot fixture, it draws hot water through the plumbing thus increasing water temperature. In addition to flowrate, water temperature was measured at the locations depicted in Error! Reference source not found. with one-second resolution. The decrease in hot water temperature as a result of stagnation can be seen in these data shown in **Figure 13**. Note how water use at hot.kitchen.isl aligns with water use at the service line, and with deviations in water temperature measured just upstream of hot.kitchen.isl's fixture.

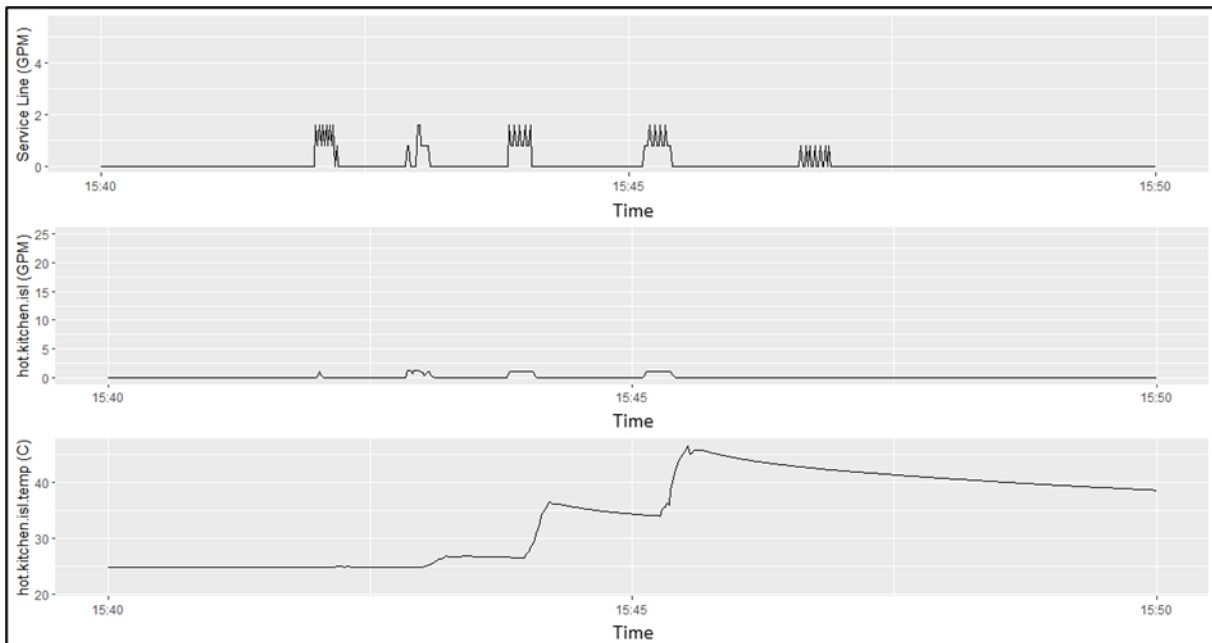


Figure 13 - Example data showing the impact of flowrate on temperature on 5/4/2017

However, flow at this meter does not always align with changes in service line flowrates or temperature data as shown in **Figure 14**. Here, water use at hot.kitchen.isl does not correspond to water use at the service line. Further, water use at hot.kitchen.isl has no apparent influence on measured water temperature, and changes in flowrate at hot.kitchen.isl appear to loosely correspond to flowrates observed at the service line potentially indicating poor sensor grounding or electrical interference. As such, it appears that all water use at hot.kitchen.isl over the time period shown in **Figure 14** is erroneous.

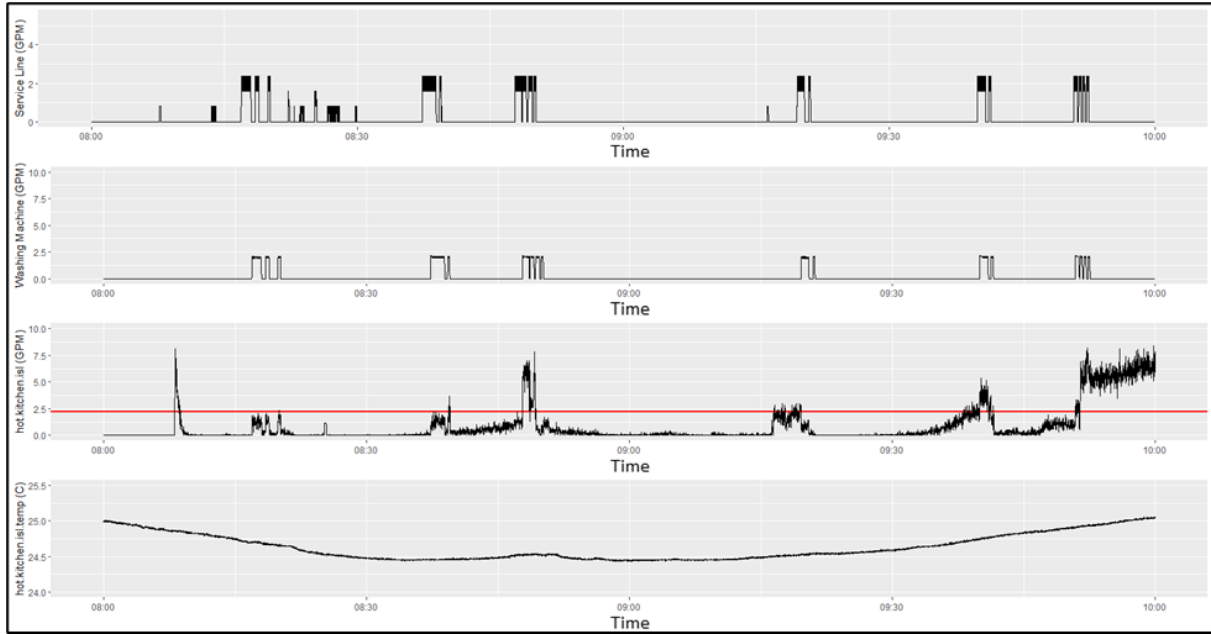


Figure 14 - Example data showing flowmeter noise with no influence on water temperature on 9/5/2016. The design flowrate of hot.kitchen.isl is shown as a horizontal red line.

These two meters exhibiting noise were each installed on the hot line of a sink, and thus each had a cold-water counterpart. Further, two kitchen sinks (kitchen.sink and kitchen.isl) and two bathroom sinks (bath1.sink and bath2.sink) were monitored during the study. It was assumed that the ratio of hot to cold water consumed was equal at each type of sink (i.e. kitchen or bathroom). Thus, the flowrate data from hot.kitchen.isl and hot.bath2.sink were replaced as follows:

$$\text{hot.kitchen.isl} = \text{cold.kitchen.isl} * \text{sum}(\text{hot.kitchen.sink}) / \text{sum}(\text{cold.kitchen.sink})$$

$$\text{hot.bath2.sink} = \text{cold.bath2.sink} * \text{sum}(\text{hot.bath1.sink}) / \text{sum}(\text{cold.bath2.sink})$$

While accurate flowmeter data for all sample locations is obviously preferable, that was not available for this data set. Replacing the data from these two flowmeters exhibiting noise with a ratio multiple of their noise-free cold counterparts provides several advantages, such as achieving plausible flowrates from all monitored fixtures and realistic stagnation between water

uses. This method is also simple to understand and implement, with no need to model the house occupancy or time of use.

REFERENCES

REFERENCES

1. National Academies of Sciences Engineering and Medicine. *Management of Legionella in Water Systems*. (The National Academies Press, 2019). doi:10.17226/25474
2. Rhoads, W. J., Pruden, A. & Edwards, M. A. Survey of green building water systems reveals elevated water age and water quality concerns. *Environ. Sci. Water Res. Technol.* **2**, 164–173 (2016).
3. Wang, H. *et al.* Methodological approaches for monitoring opportunistic pathogens in premise plumbing: A review. *Water Res.* **117**, 68–86 (2017).
4. National Research Council. *Drinking water distribution systems: Assessing and reducing risks*. National Academies Press (2006).
5. Julien, R. *et al.* Knowledge gaps and risks associated with premise plumbing drinking water quality. *AWWA Water Sci.* 1–18 (2020). doi:10.1002/aws2.1177
6. Falkinham, J., Pruden, A. & Edwards, M. Opportunistic Premise Plumbing Pathogens: Increasingly Important Pathogens in Drinking Water. *Pathogens* **4**, 373–386 (2015).
7. Falkinham, J., Hilborn, E. D., Arduino, M. J., Pruden, A. & Edwards, M. A. Epidemiology and ecology of opportunistic premise plumbing pathogens: *Legionella pneumophila*, *Mycobacterium avium*, and *Pseudomonas aeruginosa*. *Environ. Health Perspect.* **123**, 749–758 (2015).
8. Falkinham, J. O. Common Features of Opportunistic Premise Plumbing Pathogens. *Int. J. Environ. Res. Public Health* **12**, 4533–4545 (2015).
9. Proctor, C. R., Dai, D., Edwards, M. A. & Pruden, A. Interactive effects of temperature, organic carbon, and pipe material on microbiota composition and *Legionella pneumophila* in hot water plumbing systems. *Microbiome* **5**, (2017).
10. LeChevallier, M. W. Conditions favouring coliform and HPC bacterial growth in drinking- water and on water contact surfaces. in *Heterotrophic Plate Counts and Drinking-water Safety* (eds. J., B., J., C., M., E., C., F. & A., G.) 177–97 (TJ International (Ltd.), 2003).
11. Lendenmann, U., Snozzi, M. & Egli, T. Growth kinetics of *Escherichia coli* with galactose and several other sugars in carbon-limited chemostat culture. *Can. J. Microbiol.* **46**, 72–80 (2000).
12. Cirillo, J. D., Falkow, S., Tompkins, L. S. & Bermudez, L. E. Interaction of *Mycobacterium avium* with environmental amoebae enhances virulence. *Infect. Immun.*

- 65**, 3759–3767 (1997).
13. Naumova, E. N., Liss, A., Jagai, J. S., Behlau, I. & Griffiths, J. K. Hospitalizations due to selected infections caused by opportunistic premise plumbing pathogens (OPPP) and reported drug resistance in the United States older adult population in 1991–2006. *J. Public Health Policy* **37**, 500–513 (2016).
 14. AWWA. *Effects of Water Age on Distribution System Water Quality*. US EPA (2002).
 15. Wingender, J. & Flemming, H. C. Biofilms in drinking water and their role as reservoir for pathogens. *Int. J. Hyg. Environ. Health* **214**, 417–423 (2011).
 16. Pryor, M. *et al.* Investigation of opportunistic pathogens in municipal drinking water under different supply and treatment regimes. *Water Sci. Technol.* **50**, 83–90 (2004).
 17. Cooper, I. R. & Hanlon, G. W. Resistance of *Legionella pneumophila* serotype 1 biofilms to chlorine-based disinfection. *J. Hosp. Infect.* **74**, 152–159 (2010).
 18. Lau, H. Y. & Ashbolt, N. J. The role of biofilms and protozoa in legionella pathogenesis: Implications for drinking water. *Journal of Applied Microbiology* **107**, 368–378 (2009).
 19. Ji, P., Rhoads, W. J., Edwards, M. A. & Pruden, A. Impact of water heater temperature setting and water use frequency on the building plumbing microbiome. *ISME J.* **11**, 1318–1330 (2017).
 20. Percival, S. L., Suleman, L. & Donelli, G. Healthcare-Associated infections, medical devices and biofilms: Risk, tolerance and control. *J. Med. Microbiol.* **64**, 323–334 (2015).
 21. R. Temmerman, H. Vervaeren, B. Nosedá, N. Boon & W. Verstraete. Necrotrophic Growth of *Legionella pneumophila*. *Appl. Environ. Microbiol.* (2006).
 22. Schrottenbaum, I. *et al.* Simple Model of Attachment and Detachment of Pathogens in Water Distribution System Biofilms. *World Environ. Water Resour. Congr. 2009* **41036**, 1–13 (2009).
 23. Flemming, H. C. Biofilms. *Encycl. Life Sci.* **1**, (2008).
 24. Flemming, H. C., Percival, S. L. & Walker, J. T. Contamination potential of biofilms in water distribution systems. *Water Sci. Technol. Water Supply* **2**, 271–280 (2002).
 25. Applegate, D. H. & Bryers, J. D. Effects of carbon and oxygen limitations and calcium concentrations on biofilm removal processes. *Biotechnol. Bioeng.* **37**, 17–25 (1991).
 26. Schoen, M. E. & Ashbolt, N. J. An in-premise model for *Legionella* exposure during showering events. *Water Res.* **45**, 5826–5836 (2011).
 27. Beer, K. *et al.* Surveillance for Waterborne Disease Outbreaks Associated with Drinking Water — United States, 2011–2012. *CDC: Morbidity and Mortality Weekly Report* **64**,

- (2015).
28. USEPA. Technologies for Legionella Control in Premise Plumbing Systems: Scientific Literature Review. *EPA, Off. Water* 139 (2016).
 29. Lin, Y. S. E., Vidic, R. D., Stout, J. E. & Yu, V. L. Legionella in water distribution systems. *J. / Am. Water Work. Assoc.* **90**, 112–121 (1998).
 30. Fraser, D. W. *et al.* Legionnaires' Disease. *N. Engl. J. Med.* **297**, 1189–1197 (1977).
 31. Logan-Jackson, A. R., Flood, M. & Rose, J. B. Enumeration and characterization of five pathogenic: Legionella species from large research and educational buildings. *Environ. Sci. Water Res. Technol.* **7**, 321–334 (2021).
 32. CDC. Legionnaires' Disease Surveillance Summary Report, United States-2014-2015. (2018).
 33. Agency, U. S. E. P. & Water, O. 2018 Edition of the Drinking Water Standards and Health Advisories Tables. (2018).
 34. United States Environmental Protection Agency. Factoids: Drinking water and groundwater statistics for 2007. *Environ. Prot. Agency* 1–15 (2008).
 35. Parr, A., Whitney, E. A. & Berkelman, R. L. Legionellosis on the rise: A review of guidelines for prevention in the United States. *J. Public Heal. Manag. Pract.* **21**, E17–E26 (2015).
 36. Persily, A. & Healy, W. Measurement Science Research Needs for Premise Plumbing Systems NIST Technical Note 2088 Measurement Science Research Needs for Premise Plumbing Systems.
 37. Pickering, R., Onorevole, K., Greenwood, R. & Shadid, S. *Measurement Science Roadmap Workshop for Water Use Efficiency and Water Quality in Premise Plumbing Systems : August 1-2 , 2018 Synthesis of a Workshop organized by the National NIST GCR 19-020 Measurement Science Roadmap Workshop for Water Use Efficiency.* (2018).
 38. Zlatanovic, L., Moerman, A., van der Hoek, J. P., Vreeburg, J. & Blokker, M. Development and validation of a drinking water temperature model in domestic drinking water supply systems. *Urban Water J.* **14**, 1031–1037 (2017).
 39. Bédard, E. *et al.* Temperature diagnostic to identify high risk areas and optimize Legionella pneumophila surveillance in hot water distribution systems. *Water Res.* **71**, 244–256 (2015).
 40. World Health Organization. *Legionella and the Prevention of Legionellosis.* (2007).
 41. Kragh, K. N. *et al.* Role of multicellular aggregates in biofilm formation. *MBio* **7**, 1–11

- (2016).
42. Lehtola, M. J. *et al.* The effects of changing water flow velocity on the formation of biofilms and water quality in pilot distribution system consisting of copper or polyethylene pipes. *Water Res.* **40**, 2151–2160 (2006).
 43. Proctor, C. R. *et al.* Considerations for Large Building Water Quality after Extended Stagnation. *AWWA Water Sci.* Accepted Author Manuscript (2020). doi:10.1002/aws2.1186
 44. Mullis, S. & Falkinham, J. O. Adherence and biofilm formation of *Mycobacterium avium*, *Mycobacterium intracellulare* and *Mycobacterium abscessus* to household plumbing materials. *J. Appl. Microbiol.* **115**, 908–914 (2013).
 45. Stout, J. E., Yu, V. L. & Best, M. G. Ecology of *Legionella pneumophila* within water distribution systems. *Appl. Environ. Microbiol.* **49**, 221–228 (1985).
 46. Rogers, J., Dowsett, A. B., Dennis, P. J., Lee, J. V. & Keevil, C. W. Influence of plumbing materials on biofilm formation and growth of *Legionella pneumophila* in potable water systems. *Appl. Environ. Microbiol.* **60**, 1842–1851 (1994).
 47. Dieter, C. A. & Maupin, M. A. Public Supply and Domestic Water Use in the United States, 2015. (2017).
 48. Dieter, C. A. *et al.* *Estimated use of water in the United States in 2015 Circular 1441*. (2018).
 49. Water Research Foundation. *Residential End Uses of Water, Version 2*. (2016).
 50. International Association of Plumbing and Mechanical Officials. 2018 Uniform Plumbing Code. (2018).
 51. Qureshi, N. & Shah, J. Aging infrastructure and decreasing demand: A dilemma for water utilities. *J. Am. Water Works Assoc.* **106**, 51–61 (2014).
 52. American Water Works Association (AWWA). 2017 State of the Water Industry Report. (2017).
 53. ASCE Committee on America’s Infrastructure. *2017 American Infrastructure Report Card: Drinking Water*. *ASCE Infrastructure Report Card* (2017).
 54. AWWA. Buried No Longer: Confronting America’s water infrastructure challenge. *Am. Water Work. Assoc.* 37 (2011).
 55. Beecher, J. A. The conservation conundrum: How declining demand affects water utilities. *J. / Am. Water Work. Assoc.* **102**, 78–80 (2010).

56. Mack, E. A. & Wrase, S. A burgeoning crisis? A nationwide assessment of the geography of water affordability in the United States. *PLoS One* **12**, 1–19 (2017).
57. Ortman, J. M., Velkoff, V. a. & Hogan, H. An aging nation: The older population in the United States. *Econ. Stat. Adm. US Dep. Commer.* **1964**, 1–28 (2014).
58. Harpaz, R., Dahl, R. M. & Dooling, K. L. Prevalence of Immunosuppression Among US Adults, 2013. *JAMA* **316**, 2547–2548 (2016).
59. Schuck, S. Water Age in Residential Premise Plumbing In the Graduate College. (University of Arizona, 2018).
doi:https://repository.arizona.edu/bitstream/handle/10150/630144/azu_etd_16515_sip1_m.pdf?sequence=1&isAllowed=y
60. Rossman, L. A., Woo, H., Tryby, M., Janke, R. & Haxton, T. EPANET 2.2 User Manual. (2020). doi:10.1177/0306312708089715
61. PANGULURI, S., GRAYMAN, W. M., CLARK, R. M., GARNER, L. M. & HAUGHT, R. *Water Distribution System Analysis: Field Studies, Modeling and Management*. (2005).
62. Haas, C. N., Rose, J. B. & Gerba, C. P. *Quantitative Microbial Risk Assessment*. (Wiley-Blackwell, 2014).
63. Hamilton, K. A., Weir, M. H. & Haas, C. N. Dose response models and a quantitative microbial risk assessment framework for the Mycobacterium avium complex that account for recent developments in molecular biology, taxonomy, and epidemiology. *Water Res.* **109**, 310–326 (2017).
64. Hamilton, K. A., Ahmed, W., Toze, S. & Haas, C. N. Human health risks for Legionella and Mycobacterium avium complex (MAC) from potable and non-potable uses of roof-harvested rainwater. *Water Res.* **119**, 288–303 (2017).
65. Proctor, C. R. & Pruden, A. Effect of Various Water Chemistry Factors on Legionella Proliferation and the Premise Plumbing Microbiome Composition. *Civ. Environ. Eng. Masters of*, 176 (2014).
66. Hayes-Phillips, D., Bentham, R., Ross, K. & Whiley, H. Factors influencing legionella contamination of domestic household showers. *Pathogens* **8**, (2019).
67. United States Environment Protection Agency (US EPA). Water-Efficient Single-Family New Home Specification Supporting Statement. (2008).
68. Buttorff, C., Ruder, T. & Bauman, M. *Multiple Chronic Conditions in the United States. Multiple Chronic Conditions in the United States* (2017). doi:10.7249/tl221

69. Collier, S. A. *et al.* HHS Public Access. **140**, 2003–2013 (2012).
70. Whirlpool Corp. Whirlpool ReNEWW House - Retrofit Net Zero: Energy. Water. Waste.
71. Salehi, M. *et al.* Case study: Fixture water use and drinking water quality in a new residential green building. *Chemosphere* **195**, 80–89 (2018).
72. Salehi, M. *et al.* An investigation of spatial and temporal drinking water quality variation in green residential plumbing. *Build. Environ.* **169**, 106566 (2020).
73. Wadowsky, R. M., Wolford, R., McNamara, A. M. & Yee, R. B. Effect of temperature, pH, and oxygen level on the multiplication of naturally occurring *Legionella pneumophila* in potable water. *Appl. Environ. Microbiol.* **49**, 1197–205 (1985).
74. Edagawa, A. *et al.* Detection of culturable and nonculturable *Legionella* species from hot water systems of public buildings in Japan. *J. Appl. Microbiol.* **105**, 2104–2114 (2008).
75. Bargellini, A. *et al.* Parameters predictive of *Legionella* contamination in hot water systems: Association with trace elements and heterotrophic plate counts. *Water Res.* **45**, 2315–2321 (2011).
76. Stout, J. E. *et al.* *Legionella pneumophila* in residential water supplies: Environmental surveillance with clinical assessment for Legionnaires' disease. *Epidemiol. Infect.* **109**, 49–57 (1992).
77. Ley, C. J. *et al.* Drinking water microbiology in a water-efficient building: Stagnation, seasonality, and physiochemical effects on opportunistic pathogen and total bacteria proliferation. *Water Res. Technol.* **6**, 2902–2913 (2020).
78. Shapiro, S. S. & Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **52**, 591 (1965).
79. R Core Team. R: A language and environment for statistical computing. (2021).
80. RStudio Team. RStudio: Integrated Development for R. (2021).
81. Omega Engineering Inc. FPR300-310 Series Low-Flow Meter. (2017).
82. Connell, M. *et al.* PEX and PP water pipes: Assimilable carbon, chemicals, and odors. *J. Am. Water Works Assoc.* **108**, E192–E204 (2016).
83. Abdel-Nour, M., Duncan, C., Low, D. E. & Guyard, C. Biofilms: The stronghold of *Legionella pneumophila*. *Int. J. Mol. Sci.* **14**, 21660–21675 (2013).
84. Völker, S., Schreiber, C. & Kistemann, T. Modelling characteristics to predict *Legionella* contamination risk - Surveillance of drinking water plumbing systems and identification

- of risk areas. *Int. J. Hyg. Environ. Health* **219**, 101–109 (2016).
85. Proctor, C. R. *et al.* Biofilms in shower hoses-choice of pipe material influences bacterial growth and communities. *Environ. Sci. Water Res. Technol.* **2**, 670–682 (2016).
 86. Harrell, F. Hmisc: Harrell Miscellaneous. (2021).
 87. Jolliffe, I. T. *Principal components analysis*. (Springer-Verlag, 2002).
 88. Snedecor, G. & Cochran, W. *Statistical Methods*. (Iowa State University Press, 1989).
 89. Ling, F., Whitaker, R., LeChevallier, M. W. & Liu, W.-T. Drinking water microbiome assembly induced by water stagnation. *ISME J.* **2018** 1 (2018). doi:10.1038/s41396-018-0101-5
 90. Burnham, K. P. & Anderson, D. R. *Model Selection and Inference: A Practical Information-Theoretic Approach*. (Springer, 2002). doi:10.2307/3803117
 91. Barton, K. MuMIn: Multi-Model Inference. (2019).
 92. Pérez, P. & de los Campos, G. BGLR : A Statistical Package for Whole Genome Regression and Prediction. *Genetics* **198**, 483–495 (2014).
 93. O’Hara, R. B. & Sillanpää, M. J. A review of bayesian variable selection methods: What, how and which. *Bayesian Anal.* **4**, 85–118 (2009).
 94. Woznicki, S. A. *et al.* Ecohydrological model parameter selection for stream health evaluation. *Sci. Total Environ.* **511**, 341–353 (2015).
 95. Pérez, P., Campos, G. de los, Crossa, J. & Gianola, D. Genomic-Enabled Prediction Based on Molecular Markers and Pedigree Using the Bayesian Linear Regression Package in R. *Plant Genome* **2**, 106–116 (2010).
 96. Duda, S., Baron, J. L., Wagener, M. M., Vidic, R. D. & Stout, J. E. Lack of correlation between Legionella colonization and microbial population quantification using heterotrophic plate count and adenosine triphosphate bioluminescence measurement. *Environ. Monit. Assess.* **187**, (2015).
 97. Serrano-Suárez, A. *et al.* Microbial and physicochemical parameters associated with Legionella contamination in hot water recirculation systems. *Environ. Sci. Pollut. Res.* **20**, 5534–5544 (2013).
 98. Pierre, D. *et al.* Water quality as a predictor of Legionella positivity of building water systems. *Pathogens* **8**, (2019).
 99. Van Der Kooij, D., Veenendaal, H. R. & Scheffer, W. J. H. Biofilm formation and

- multiplication of *Legionella* in a model warm water system with pipes of copper, stainless steel and cross-linked polyethylene. *Water Res.* **39**, 2789–2798 (2005).
100. van der Kooij, D., Bakker, G. L., Italiaander, R., Veenendaal, H. R. & Wullings, B. A. Biofilm composition and threshold concentration for growth of *Legionella pneumophila* on surfaces exposed to flowing warm tap water without disinfectant. *Appl. Environ. Microbiol.* **83**, (2017).
 101. Hua, F., West, J. R., Barker, R. A. & Forster, C. F. Modelling of chlorine decay in municipal water supplies. *Water Res.* **33**, 2735–2746 (1999).
 102. Zhang, X. L., Yang, H. W., Wang, X. M., Fu, J. & Xie, Y. F. Formation of disinfection by-products: temperature effect and kinetic modeling. *Huanjing Kexue/Environmental Sci.* **33**, 4046–4051 (2012).
 103. Kuchta, J. M., States, S. J., McNamara, A. M., Wadowsky, R. M. & Yee, R. B. Susceptibility of *Legionella pneumophila* to chlorine in tap water. *Appl. Environ. Microbiol.* **46**, 1134–1139 (1983).
 104. Clark, R. M. Chlorine demand and TTHM formation kinetics: A second-order model. *J. Environ. Eng.* **124**, 16–24 (1998).
 105. Golea, D. M. *et al.* THM and HAA formation from NOM in raw and treated surface waters. *Water Res.* **112**, 226–235 (2017).
 106. Neu, L. & Hammes, F. Feeding the Building Plumbing Microbiome: The Importance of Synthetic Polymeric Materials for Biofilm Formation and Management. *Water* **12**, 1774 (2020).
 107. Shamsaei, H., Jaafar, O. & Basri, N. E. A. Effects Residence Time to Water Quality in Large Water Distribution Systems. *Engineering* **05**, 449–457 (2013).
 108. BARTLETT, M. S. A FURTHER NOTE ON TESTS OF SIGNIFICANCE IN FACTOR ANALYSIS. *Br. J. Stat. Psychol.* **4**, 1–2 (1951).
 109. Kaiser, H. F. A SECOND GENERATION LITTLE JIFFY. *Psychometrika* **35**, 401–415 (1970).
 110. Kaiser, H. F. An Index of Factorial Simplicity. *Psychometrika* **39**, 31–36 (1974).
 111. Zlatanović, L., van der Hoek, J. P. & Vreeburg, J. H. G. An experimental study on the influence of water stagnation and temperature change on water quality in a full-scale domestic drinking water system. *Water Res.* **123**, 761–772 (2017).
 112. The Water Planet Company. *Nitrification & Denitrification*. (2010).

113. Brooks, M. E. *et al.* glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* **9**, 378–400 (2017).
114. Carlson, K. M., Boczek, L. A., Chae, S. & Ryu, H. Legionellosis and recent advances in technologies for Legionella control in premise plumbing systems: A review. *Water (Switzerland)* **12**, 1–22 (2020).
115. Hamilton, K. A. *et al.* Risk-Based Critical Concentrations of Legionella pneumophila for Indoor Residential Water Uses. *Environ. Sci. Technol.* **53**, 4528–4541 (2019).
116. De Filippis, P., Mozzetti, C., Messina, A. & D’Alò, G. L. Prevalence of Legionella in retirement homes and group homes water distribution systems. *Sci. Total Environ.* **643**, 715–724 (2018).
117. O’Toole, J., Keywood, M., Sinclair, M. & Leder, K. Risk in the mist? Deriving data to quantify microbial health risks associated with aerosol generation by water-efficient devices during typical domestic water-using activities. *Water Sci. Technol.* **60**, 2913–2920 (2009).
118. Johnson, D., Lynch, R., Marshall, C., Mead, K. & Hirst, D. Aerosol generation by modern flush toilets. *Aerosol Sci. Technol.* **47**, 1047–1057 (2013).
119. Donohue, M. J. *et al.* Increased frequency of nontuberculous mycobacteria detection at potable water taps within the United States. *Environ. Sci. Technol.* **49**, 6127–6133 (2015).