Contextualization and Diet Implications of Occoquan Reservoir Salinization

Caitlin M. Shipman

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Environmental Sciences & Engineering

Stanley Grant, Chair
Peter Vikesland
Todd Schenk

February 17, 2023
Manassas, Virginia

Keywords: Occoquan, freshwater, salinization, threshold, context

Contextualization and Diet Implications of Occoquan Reservoir Salinization

Caitlin M. Shipman

(ABSTRACT)

Freshwater salinization syndrome is a rising threat globally which results in increased ion concentrations in inland freshwaters. This syndrome threatens healthy aquatic ecosystems and can alter the perception of the potability of finished drinking water. The Occoquan Reservoir, located in Northern Virginia, is a freshwater system that is facing rising salinization. Stakeholders for the reservoir have been convened to address these rising salinization concerns. Among these stakeholders, there are a variety of viewpoints on the significance of the salinization, which is preventing a high level of convergence around this threat. To assist in contextualizing this system, empirical cumulative distribution functions were generated from data gathered from various governmental sources and compared the reservoir's watershed and finished drinking water ion concentrations. These analyses show that the watershed and finished drinking water have some of the highest concentrations of sodium and chloride statewide. Additional investigations determined the trend of sodium increases in finished drinking water since the 1980s. Monte Carlo simulations were ran to determined whether there would be risks to human from ingesting this water should this trend continued. Results from these analyses greatly varied due to the wide range in drinking water ingestion rates. The purpose of these analyses is to assist with stakeholder convergence around the level of threat salinization poses to the reservoir and to initiate discussions of what an acceptable threshold for management could be.

Contextualization and Diet Implications of Occoquan Reservoir Salinization

Caitlin M. Shipman

(GENERAL AUDIENCE ABSTRACT)

Freshwater salinization syndrome is a rising threat globally which results in increased ion concentrations in inland freshwaters. This syndrome threatens healthy aquatic ecosystems and can alter the perception of the potability of finished drinking water. The Occoquan Reservoir, located in Northern Virginia, is a freshwater that is facing rising salinization. Stakeholders for the reservoir have been convened to address these concerns. Among the stakeholders, there are a variety of viewpoints on the significance of salinization. Various analyses were done to compare the sodium and chloride concentrations in the reservoir's watershed and in the finished drinking water with respective statewide levels. These analyses show that the watershed and finished drinking water have some of the highest concentrations of sodium and chloride statewide. Additional investigations were conducted to determine if there was a human health risk to consuming the finished drinking water. Results from this analysis were highly dependent on how much water an individual consumed. The purpose of these analyses is to assist with stakeholder convergence around the level of threat salinization poses to the reservoir and to initiate discussions of what an acceptable threshold for management could be.

Dedication

I'd like to dedicate this thesis to my friends and family who encouraged me to take a chance and return to school.

Acknowledgments

Thank you to the students and professors of Virginia Tech's Environmental and Water Resources Engineering program who I've had the pleasure to learn from over the past few years. A special thank you to everyone at the Occoquan Watershed Monitoring Lab, especially Shantanu Bhide and Lauren Krauss, who assisted me with data collection. And lastly, thank you to my advisor, Dr. Stanley Grant, and committee members, Dr. Todd Schenk and Dr. Peter Vikesland, for their guidance and support.

Contents

List of Figures

List of Tables

List of Abbreviations

AAFP American Academy of Family Physicians

AHA American Heart Association

AI Adequate Intake: recommended average daily nutrient intake level

CDC Centers for Disease Control and Prevention

CDF Cumulative distribution function

CDRR Chronic Daily Risk Reduction Intake: lowest intake level of intake for which there was sufficient evidence to characterize a chronic disease risk reduction

DASH Dietary Approaches to Stop Hypertension

DEQ Virginia's Department of Environmental Quality

DRI Dietary Reference Intake

DWEL Drinking Water Equivalent Level

EPA United States Environmental Protection Agency

FDA Food and Drug Administration

FSS freshwater salinization syndrome

GCR Growing Convergence Research program

Integrated Report 305(b)/303(d) Water Quality Assessment Integrated Report

NASEM National Academies of Sciences, Engineering, and Medicine

NHANES National Health and Nutrition Examination Study

NIH National Institute of Health

NPDWRs National Primary Drinking Water Regulations

NSDWRs National Secondary Drinking Water Regulations

NSF National Science Foundation

OWML Occoquan Watershed Monitoring Lab

PWS Public water systems

TAC Technical Advisory Committee

TMDL Total Maximum Daily Load

TUL Tolerable Upper Limit

UOSA Upper Occoquan Service Authority

VPDES Virginia Pollution Discharge Elimination System Regulations (9VAC25-31)

VSCI Virginia Stream Condition Index

WQS Virginia's Water Quality Standards Regulations (9VAC25-260)

Chapter 1

Introduction

1.1 Freshwater Salinization Syndrome

Inland freshwater salinization is defined as an increase in salt concentrations. This trend is a part of the broader change in the chemistry of inland freshwaters worldwide, and includes rising pH, alkalinity, and base cation concentrations, which together is known as the ”freshwater salinization syndrome” (FSS).[11] A recent study of 422 streams within the United States by the United States Geological Survey determined that inland freshwaters are *salinized rapidly in all human-dominated land use types.*[20]

The FSS threatens healthy freshwater ecosystems and drinking water security.[4] Increasing ion concentrations in these streams are associated with declines in benthic macroinvertebrates, which is indicative of poor stream health.[18] Furthermore, increasing ion concentrations can alter the perception of potability, decreasing trust in public water supplies.[8] These increased concentrations can lead to the leaching of heavy metals in drinking water as well as increased corrosion of drinking water infracstructure.[15]

1.2 Study Site: The Occoquan Reservoir

The Occoquan Reservoir, located in Prince William County, Virginia, supplies drinking water for approximately 1 million people in Northern Virginia and the surrounding localities

(Figure 1.1).[9] The reservoir has been augmented by highly treated wastewater effluent from the Upper Occoquan Service Authority (UOSA) since the 1970s, in a process known as indirect potable reuse.[4] Indirect potable reuse is the practice of using highly treated wastewater effluent to supplement drinking water supply.[9] The United States Environmental Protection Agency (EPA) promotes indirect potable reuse as a tool to address projected water supply shortages over the next 10 years in 40 different states.[29]

This reservoir has been monitored by Virginia Tech's Occoquan Watershed Monitoring Lab (OWML) since 1972. The lab has monitoring stations throughout the reservoir's watershed, including along it's tributaries, downstream of UOSA's discharge's confluence with Bull Run (ST45), and immediately upstream of Fairfax Water's drinking water intake (ST01).

The reservoir has shown a general trend of increasing salinization over the past several decades. [4] The sodium concentration in the reservoir frequently exceeds EPA's lower taste threshold (Section 1.3.1), prompting the local water authority, Fairfax Water, to explore new treatment options to address this increase.[4] These treatments include constructing a desalinization facility, to in effect desalinate freshwater for an estimated cost of $1 billion.[4]

1.3 State and Federal Regulations

1.3.1 Drinking Water Regulations

Drinking water in the Unites States is protected by the United States Environmental Protection Agency (EPA), which promulgates primary and secondary drinking water standards. Primary drinking water standards are promulgated under he National Primary Drinking Water Regulations (NPDWRs) and are legally enforceable standards aimed at protecting human health.[22] There are no primary standards for either sodium or chloride.[22]

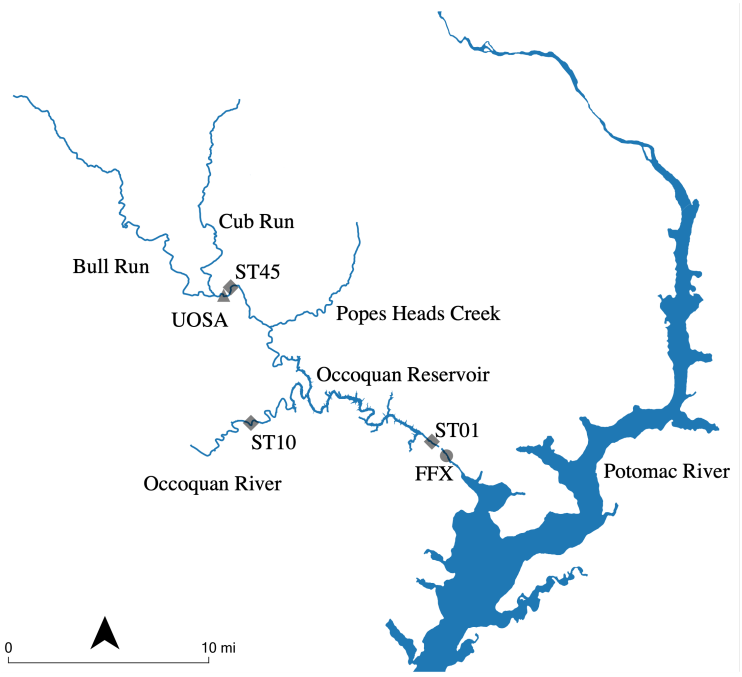


Figure 1.1: Map of the watershed, including tributaries to the reservoir (Bull Run & Occoquan River), as well as the location of OWML's monitoring stations (ST01, ST10, ST45), UOSA's discharge (UOSA), and the drinking water intake of Fairfax Water (FFX).

In addition to these standards, EPA promulgates secondary standards under the National Secondary Drinking Water Regulations (NSDWRs). Secondary standards are not mandatory or enforceable. Instead, they are established as guidelines for maintaining the aesthetic quality of drinking water (i.e. taste, color, and odor).[23] Chloride has a secondary standard of 250 mg/L, which protects against a salty taste.[23]

In 2003, a Drinking Water Advisory for sodium was issued to provide guidance to communities that are exposed to drinking water containing elevated levels of sodium.[24] This advisory recommended that sodium concentrations be between 30 and 60 mg/L to ensure there was no effect on taste.[25] This was concluded as only contributing 2.5% to 5% of a dietary goal of 2.4 g/day of sodium for an individual consuming 2 L/day of tap water.[25]

In addition, EPA has a Drinking Water Equivalent Level (DWEL) for sodium in drinking

water of 20 mg/L. A DWEL is a non-enforceable guideline that is considered protective against non-carcinogenic health effects.[24] This level was developed for individuals on a severely restricted total sodium intake of 500 mg/day.[25] This guidance level has been in effect since 1996.[8]

1.3.2 Environmental Regulations

Water Quality Standards and VPDES Regulations

The health of Virginia's environment is protected by the Virginia Department of Environmental Quality (DEQ). This agency is responsible for implementing a variety of laws and regulations aimed at protecting waterbodies within the Commonwealth, including Virginia's Water Quality Standards (WQS) and Virginia Pollution Discharge Elimination System permits (VPDES).

Virginia's Water Quality Standards, which undergo a comprehensive review every three years and are were last updated October 18, 2019, create qualitative and quantitative standards for all waterbodies within the Commonwealth.[33] The basis of these standards are the "designated uses" of a waterbody. Designated uses are defined as *those uses specified in water quality standards for each waterbody of segment whether or not they are being attained.*[33] These regulations state *all state waters, including wetlands, are designated for the following uses...to protect the propagation and growth of a balanced, indigenous, population of aquatic life, including game fish, which might reasonably be expected to inhabit them* (VAC25-260-10). This is referred to as the "aquatic life designated use". Other potential designated uses are *recreation, public water supply, human health, wildlife, and the production of edible and marketable natural resources* (commonly referred to as "shellfish" use).[33]

The aquatic life and human health use designations are the basis for the acute and chronic numerical criteria promulgated in Virginia's WQS (9VAC25-260).[33] The human health numerical criteria includes two categories: 1) for waterbodies that have been designated as public water supplies, criteria protect from toxic effects through drinking water and fish consumption, 2) all other waterbodies, where criteria protect human health from toxic effects from fish consumption.[33] An acute criteria protects against acute toxicity, which is defined in the Water Quality Standards as *an adverse effect that usually occurs shortly after exposure to a pollutant.*[33] Similarly, the chronic criteria protects against chronic toxicity, defined as *an adverse effect that is irreversible or progressive or occurs because the rate of injury is greater than the rate of repair during prolonged exposure to a pollutant.This includes low level, long-term effects such as reduction in growth or reproduction.*[33]

Permits for point source discharges into State waters are regulated under the Virginia Discharge Elimination System regulations, promulgated at 9VAC25-31.[34] A limit developed for a VPDES permit must be protective of all beneficial uses.[34] Therefore, while both criteria are considered in when developing VPDES permit limits, the aquatic life criteria is typically more stringent than human health criteria, and therefore drives the limit development.[34]

Virginia has promulgated acute and chronic aquatic life criteria for chloride of 860 mg/L and 230 mg/L, respectively, as well as an human health criteria of 250 mg/L for waters' designated as a public water supply in 9VAC25-260-390 through 9VAC25-260-540.[33] No Water Quality Standards have been promulgated for sodium.[33]

Water Quality Monitoring and TMDLs

DEQ conducts biological, ambient, and probablistic water quality monitoring. Biological monitoring samples a stream's benthic macroinvertebrate community, and from which species are present or absent in that community, determines a Virginia Stream Condition Index

(VSCI) score that represents the streams health. The ambient monitoring program samples pH, dissolved oxygen, and specific conductance of freshwater streams, while the probabilistic sampling program randomly selects monitoring stations from across the Commonwealth, providing an unbiased characterization of waters statewide.[30]

In accordance with Sections 305(b) and 303(d) of the Clean Water Act, DEQ provides a biennial 305(b)/303(d) Water Quality Assessment Integrated Report (Integrated Report) to EPA, which assesses and summarizes the current conditions of the State's waters.[32]

It is not unusual for a stream's ambient or probablistic monitoring data to be in compliance with numeric WQS, but the biological monitoring shows an impairment. This could be caused by a variety of factors, such as a high density of point source discharges or high level of urbanization. Should the biennial water quality assessment determine such a stream is impaired for its aquatic life designated use, DEQ will conduct a stressor analysis to determine what is causing the impairment.

A stressor analysis is a regulatory process involving a Technical Advisory Committee, public meetings, and a public comment period. If a pollutant is identified from the stressor analysis, DEQ will create a Total Maximum Daily Load (TMDL). A TMDL is a pollution diet for the stream and is also a regulatory process involving a TAC, public meetings, and public comment period. The TMDL determines Wasteload Allocations and Load Allocations that will protect the stream. Wasteload and Load Allocations are incorporated into VPDES Permits and Implementation Plans, respectively. Further water quality monitoring will determine whether the TMDL was successful. Due to the number of impaired streams, DEQ creates a TMDL priorities list for every two years.[31] (Appendix A).

1.4 Thresholds & Watershed Management

The research project "Catalyzing Stakeholder-Driven Solutions to Inland Freshwater Salinization", funded under the National Science Foundation's (NSF) Growing Convergence Research (GCR) program, utilizes convergence research approaches to catalyze a stakeholder-driven management of the salt budget in inland freshwaters, such as the Occoquan Reservoir.[9]

This project includes approximately 40 stakeholders, representing a broad range of perspectives, from local, state, and federal governmental organizations, nonprofit environmental organizations, and the wastewater and drinking water industries.[9] Conversations with these stakeholders show there is a very wide range in the levels of concern about rising levels of salinization in the reservoir.

The lack of any appropriate standards for either the reservoir's watershed or the drinking water that is derived from it indicate that a top-down regulatory approach to address this problem does not seem viable. Instead, the question is whether stakeholders can organize themselves and determine an appropriate bottom-up management strategy.[9]

Stakeholders have expressed the need for a threshold or thresholds to use as a basis for managing the reservoir's salt budget. Thresholds can be used to make management decisions, such as determining best management practices to implement within the watershed, where and how to develop spaces, and what level of industrial discharges should be allowed. However, agreement around salinization as an issue is a crucial first step towards convergence, and is necessary in determining how to manage this watershed.

1.5 Percent Dietary Intake of Sodium from Drinking Water

In Cruz et. al, the following equation was proposed for calculating the percent dietary intake for sodium intake in drinking water:

$$\% \text{ Dietary Intake for Sodium in Drinking Water} = \frac{(\text{Sodium in Drinking Water (mg/L)})\,(\text{Drinking Water Ingestion Rate (L/day)})}{\text{Sodium Tolerable Upper Limit (TUL) (mg/day)}} \tag{1.1}$$

Cruz, et. al used four variations of this equation: adult women, adult men, adult women on a low salt diet, adult men on a low salt diet. Daily drinking water consumption (L/day) came from the National Academies of Sciences, Engineering, and Medicine (NASEM) Adequate Intake value for total water intake, which is 2.7 L/day for women and 3.7 L/day for men.[7] NASEM defines total water intake as *the combination of fluids from beverages, drinking water, and food.*[12] Salt intakes were based on the U.S. Institute of Medicine of the National Academies Tolerable Upper Limit (2,300 mg/day) and the National Institute of Health's National Heart, Blood, and Lung Institute's Dietary Approaches to Stop Hypertension diet recommendations (1,500 mg/day).[7]

1.5.1 Review of Drinking Water Intake Recommendations

The daily water intake consumption rate used in Cruz, et. al, takes into account water that comes from sources other than publicly provided drinking water. Therefore, it is not an appropriate ingestion rate when determining the amount of sodium from drinking water.

A review of the governmental agencies (CDC, USDA, FDA, NIH), NASEM, and professional health organizations (AAFP) did not elicit a recommended value for how much water an individual should be consuming per day. The Center for Disease Control (CDC) goes as far as to state *there is no recommendation for how much plain water everyone should drink daily* and that *daily water intake recommendations vary by age, sex, pregnancy status, and breastfeeding status.*[6] Activity level and climate are also factors that may affect daily water intake recommendations.[6]

An additional issue is that any health recommendation, including the total water value provided by NASEM, would not differentiate between water consumed either from a community source, such as tap water, or water consumed from a private source, such as bottled water.

Since a single, recommended intake value was not available, we instead turned to the Environmental Protection Agency's Exposure Factors Handbook, which provides reference information on exposure factor information to assist with characterizing an individual's potential exposure to various agents.[26] EPA provides recommended ingestion rates for drinking water; these values are not health recommendations, rather, they represent the actual volume of water consumed by individuals.[27]

The recommended values for the general population were updated in 2019 using National Health and Nutrition Examination Survey (NHANES) data from 2005 – 2010. Historically, the EPA has assumed a drinking water ingestion rate of 2 L/day for adults and 1 L/day for infants and children under 10 years of age.[27] With this update, those recommendations were found to be more representative of a 90^{th} percentile intake value than an average value.[26] Rather than recommending a single intake value that should be used in exposure assessments, EPA provides a series of percentiles of drinking water intake values (mL/day) for various populations and situations.[27]

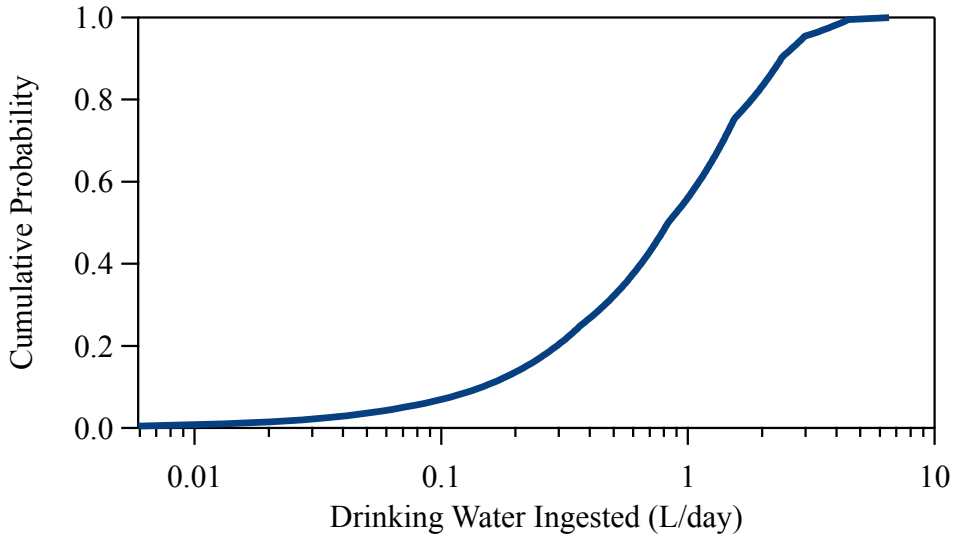


Figure 1.2: A CDF of the NHANES study results, which surveyed how much community drinking water was ingested per day (L/day) either through direct or indirect consumption.

In this study, the consumer-only estimates of combined direct and indirect water ingestion of community water for all ages were used to estimate drinking water ingestion (Figure 1.2). Consumer only intake rates represent the quantity of water consumed only by individuals who reported water intake during the survey period. Direct ingestion is defined as direct consumption of water as a beverage, while indirect ingestion includes water added during food or beverage preparation, such as the water used to make tea or soup, but does not include water intrinsic to purchased foods, such as the water in fruit.[27] Community water is water from public water distribution systems.[27]

1.5.2 Review of Daily Sodium Intake Recommendations

To confirm the dietary sodium recommendations set out in Cruz et.al., a review of various sources, including governmental agencies (FDA, NIH, CDC), the 2020 – 2025 Dietary Guidelines for Americans, Dietary Reference Intakes published by the NASEM, as well as national health organizations (AHA, AAFP) and governmental organizations (FDA, CDC,

NIH) was conducted (Table 1.1).

This review confirmed that 2300 mg/day is an appropriate recommendation for sodium intake for adults on a standard diet. Additionally, several organizations, such as the NIH, AHA, and AAFP, recommended individuals with conditions such as high blood pressure and kidney disease consume no more than 1500 mg/day of sodium. Therefore, this consumption rate was used to represent individuals on a ”low salt diet”.

Table 1.1: Summary of daily sodium intake recommendations from various governmental agencies and professional organizations.

Intake Recommendation	Chronic Disease	Organization	Year Recommended	Internal Reference
2300 mg/day	None	2020 - 2025 Dietary Guidelines for Americans[21]	2020	DRI for Sodium & Potassium, NASEM[12]
1500 mg/day (AI) 2300 mg/day (CDRR)	None	DRI for Sodium & Potassium, NASEM[12]	2019	
1500 - 2300 mg/day	None	AHA[3]	2021	
2300 mg/day	None	CDC[5]	2021	2020 - 2025 Dietary Guidelines for Americans[21]
2300 mg/day	None	AAFP[1]	2020	
2300 mg/day	None	U.S. FDA[28]	2022	2020 - 2025 Dietary Guidelines for Americans[21]
1500 mg/day	Chronic Kidney Disease	NIH[14]	2016	
1500 - 2300 mg/day	Hypertension	DASH Eating Plan, NIH[13]	2021	
1500 mg/day	High Blood Pressure	AAFP[1]	2020	

Chapter 2

Purpose Statement

The purpose of the following analyses is to facilitate further convergence among the stakeholders about whether salinization in the Occoquan Reservoir is a concern and to help facilitate discussions about an appropriate management threshold. Because a regulatory framework currently exists for addressing water quality impairments through a TMDL, analyses focused on possible human health effects.

These analyses aim to gain an understanding of 1) how the concentrations of sodium and chloride in the reservoir's tributaries and finished drinking water compare statewide, 2) how concentrations of sodium in finished drinking water have changed over time, 3) to understand what, if any, impact to human health may be from current and future sodium levels in the reservoir.

Chapter 3

Methods & Results

3.1 Water Quality of the Occoquan Reservoir Watershed

3.1.1 Current Status of Reservoir Tributaries

The Occoquan Reservoir is fed by two major rivers: Bull Run and the Occoquan River. The 2022 Integrated Report states that the Occoquan River is fully supporting the aquatic life designated use. However, Cub Run, which is a major tributary to Bull Run, is considered impaired for benthic macroinvertebrates, and is therefore not supporting aquatic life.[32] Additionally, the upper segment of Bull Run, beginning at the confluence of Cub Run and continuing downstream until the confluence with Popes Head Creek, is considered impaired for aquatic life.[32] The lower portion of Bull Run, from the confluence with Popes Head Creek to the confluence with the the Occoquan River, was not assessed.[32] (Figure 1.1).

These impairments indicate that these streams are eligible for TMDL development. A review of the most recently available TMDL priorities list, 2018 - 2020, shows that of the major tributaries in this watershed, only Cub Run is currently being prioritized for a TMDL. Additionally, three TMDLs are being developed for a tributary to Cub Run, Sand Branch, to address an aquatic life impairment. The stressor analysis for Sand Branch determined three probable pollutants were causing the impairment: total dissolved solids, total phospho-

rus, and sediment.[19] Focusing on total dissolved solids allows for flexibility in addressing multiple ions that were of concern, primarily sulfate, chloride, sodium, and potassium.[19]

3.1.2 Comparison of the Reservoir Tributaries to Statewide

DEQ provided sampling data from their Probabilistic Monitoring program. Sampling data was available from 2001 - 2020, with chloride and sodium sampling beginning monthly in 2005. This analysis looked at the most recent 15 years of data, divided into 3 five-year intervals (Table 3.1).

Table 3.1: Number of sodium and chloride samples and descriptive statistics for DEQ's Probablisitc Monitoring dataset per time period of interest.

		Sodium			Chloride		
Date Range	n	Median	Range	Std. Dev.	Median	Range	Std. Dev.
2006 - 2010	216	4.44	0.32 - 133	14.47	5.00	0.22 - 267	20.66
2011 - 2015	213	4.95	0.32 - 220	19.36	4.14	0.65 - 239	20.80
2016 - 2020	203	3.50	0.16 - 83	9.32	3.83	0.26 - 160	14.31

Empirical cumulative distribution functions (CDFs) were created for sodium concentrations, chloride concentrations, and the sodium:chloride molar ratio for each time period (Figure 3.1). Empirical CDFs were created using the following formula, taken from Aang and Tang, 2006[2]:

For a set of N observations x_1, x_2,..., x_N, arranged in increasing order, the mth value is plotted at the cumulative probability of

$$\frac{m}{N+1} \tag{3.1}$$

OWML monitors the water quality of the reservoir's major tributaries, Occoquan River and Bull Run, at ST45 and ST10, respectively (Figure 1.1). During 2006 - 2020, sampling at ST10 for sodium occurred approximately weekly (n = 709). However, sampling for chloride at ST10 did not begin until July 2019 (n = 81). Sampling at ST45 occurred approximately weekly for sodium and chloride during 2006 - 2020 (n = 823). Violin plots of monitoring data from OWML's stations ST10 and ST45, were created and used to compare these tributaries to statewide concentrations (Figure 3.1).

Over the time period analyzed, no major shifts in sodium concentrations statewide were detected. The CDFs for 2006 - 2010 and 2011 - 2015 cross, indicating they are very comparable (Figure 3.1). The CDF 2016 - 2020 is shifted slightly to the left, indicating a lower sodium concentration results in a greater cumulative probability (Figure 3.1). This indicates there may be a small increase with sodium concentration with time. However, the maximum value for 2016 - 2020 (83 mg/L) is much lower than 2011 - 2015 (220 mg/L) and 2006 - 2010 (133 mg/L) (Figure 3.1).

A comparison of the ST10 and ST45 sodium data to statewide values illuminate several important points. Firstly, median sodium concentrations at ST10 and ST45 are among the highest statewide ($91st$ and $98th$ percentiles when compared to 2016 - 2020 values). Additionally, ST45's concentration have a much greater range that ST10's and appear to frequently be above the $99th$ percentile (Figure 3.1).

Likewise, no major shifts in chloride concentrations statewide were seen over this time period. The CDFs of 2011 - 2015 and 2016 - 2020 cross, indicating they are very comparable. The CDF 2006 - 2010 shows a higher sodium concentration results in a greater cumulative probability, indicating there may have been a small increase in chloride concentration with time. However, the maximum value has again decreased with time, with the max for 2016 - 2020 (267 mg/L) much lower than 2011 - 2015 (239 mg/L) and 2006 - 2010 (160 mg/L).

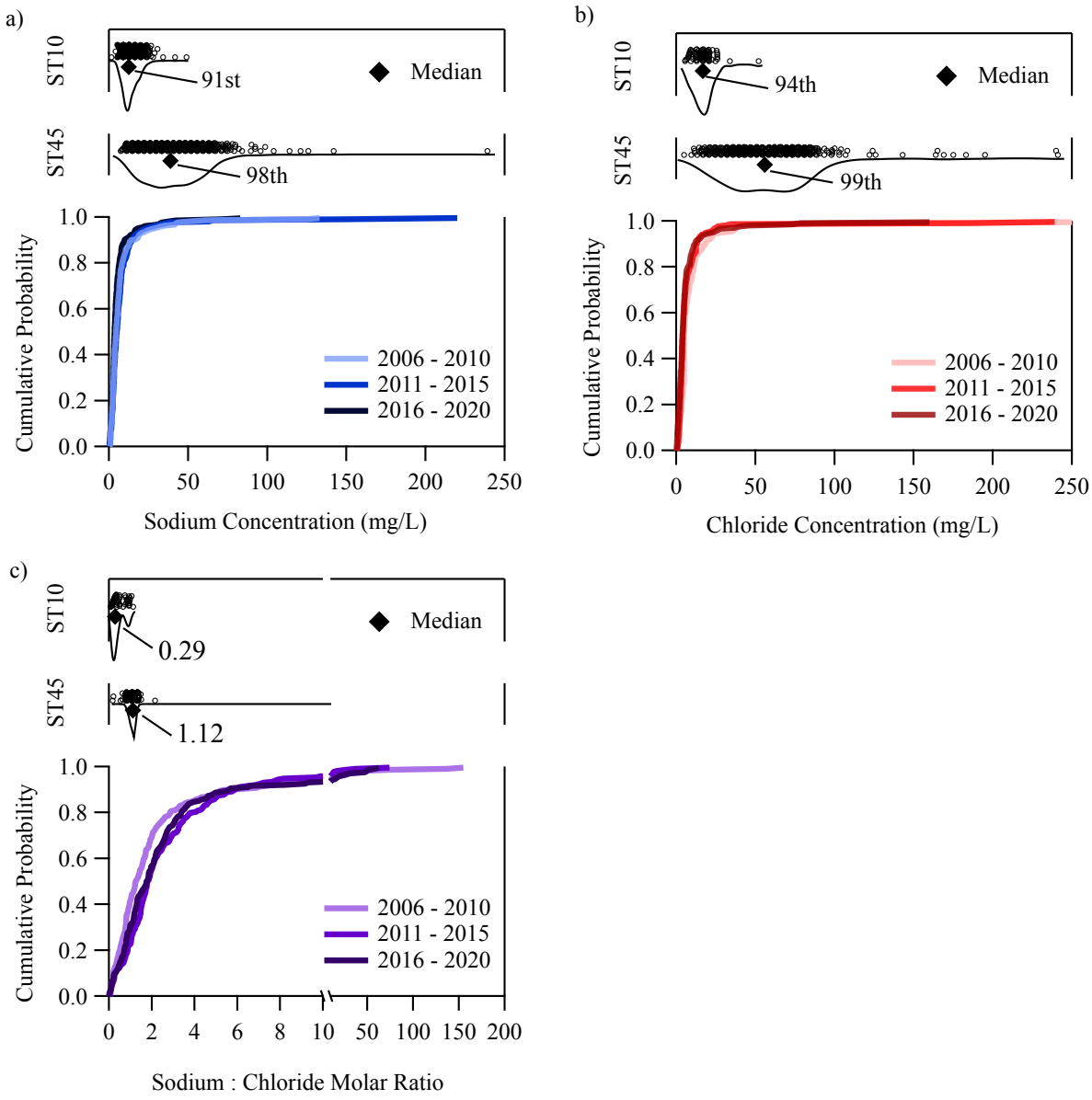


Figure 3.1: Comparison of violin plots of ST01 (Occoquan River) and ST45 (Bull Run) to empirical CDFs of statewide values for a) sodium, b) chloride, and c) sodium:chloride molar ratio. Percentiles shown on a) and b) are for how the median value for each monitoring station compare to the most recent time period, 2016 - 2020. The value of the median molar ratio is shown in c). The horizontal axis for c) has been split for clarity.

A comparison of the ST10 and ST45 chloride data to statewide values again show that median concentrations at ST10 and ST45 are among the highest statewide ($94st$ and $99th$ percentiles when compared to 2016 - 2020 values). By definition, this means that 50% of ST45's chloride concentrations are greater than the 99% of statewide values, likely making Bull Run one of saltiest streams in the Commonwealth (Figure 3.1).

A comparison of the sodium to chloride molar ratio was also conducted. Statewide, a wide range of ratios is seen (0.005 - 154). Interestingly, there is a significant difference between ST10 and ST45, with the medians being 0.29 and 1.12. This indicates that ST45 has a much higher proportion of sodium relative to chloride than ST10. ST45 is located almost immediately downstream of UOSA's discharge, therefore, this could be indicative of the discharge's influence on the tributary. These results are also consistent with previous investigations into sources of sodium in the Occoquan Reservoir, which have shown historically higher concentrations at ST45 than ST10.[4]

3.2 Sodium and Chloride in Drinking Water

Sodium and chloride concentrations in drinking water were provided by Fairfax Water, the study area's local water purveyor, from January 1981 – September 2022. Finished drinking water samples initially came from three drinking water treatment plants that used the Occoquan Reservoir as a raw intake source. These plants were eventually consolidated when the Frederick P. Griffith Water Treatment Plant came online in July 2006. Sampling occurred approximately monthly for sodium. Sampling for chloride began at twice a month for chloride and appeared to switch to monthly beginning in March 1995.

Three large gaps in the data were identified: January 1992 – December 1994, January 2000 – January 2002, July 2009 – August 2013 (Figure 3.3). Due to the consolidation of water treatment plants, variations in sample frequency, and time gaps in the sampling, the amount of samples for each decade varied from 270 (1980 – 1989) to 62 (2010 – 2019) for sodium and 1165 (1980 - 1989) to 62 (2010 - 2019) for chloride (Table 3.2).

OWML has a monitoring station, ST01, located at the dam at the mouth of the Occoquan Reservoir, making the monitoring station representative of Fairfax Water's raw drinking water intake (Figure 1.1). This monitoring station has been sampled approximately weekly for sodium since August 1996 and weekly for chloride since April 2002. Therefore, to account for the gaps in the drinking water dataset from Fairfax Water, an interpolation was done between the sodium and chloride concentrations at ST01 and the sodium and chloride concentrations of the finished drinking water.

3.2.1 Interpolation

The sampling date at Fairfax Water was subtracted from the sampling date at ST01 and the results sorted to determine days where sampling at ST01 occurred either on the same day as sampling by Fairfax Water (difference = 0) or a day before sampling by Fairfax Water (difference = 1 day) (Appendix B).

A delay of sampling by a day was of interest, as it may represent the residence time in the drinking water plant (i.e., the amount of time to produce finished drinking water from the raw water intake).[17]

Three linear regressions were conducted for the sodium concentrations: 1) samples taken on the same day ($n = 56, R^2 = 0.7093$), 2) Fairfax Water samples taken a day after an ST01 sample ($n = 132, R^2 = 0.7408$), 3) combination of both samples taken on the same day and one day apart ($n = 188, R^2 = 0.7212$).

The linear regression accounting for the residence time in the system had the highest R^2 value, indicating that a higher amount of the variance in the sodium concentration of the drinking water could be explained by the sodium concentration at ST01 the day before sampling at the drinking water plant took place. This process was repeated with the chloride data taken one day apart, which resulted in a linear regression with an R^2 of 0.90. (Figure 3.2).

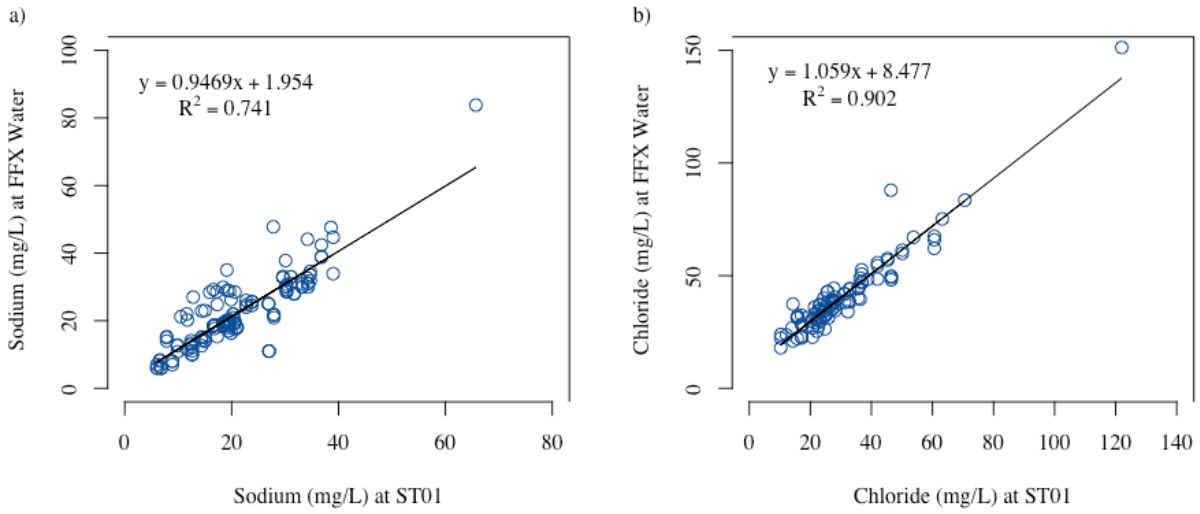


Figure 3.2: Regression analyses between a) sodium and b) chloride concentrations at OWML's monitoring station at the reservoir's dam, ST01, and Fairfax Water's finished drinking water from the Griffith Water Treatment Plant. Analyses take into account a one day residence time between the drinking water intake and finished water. Linear relationships and R^2 values are shown for each regression.

The sampling at ST01 took place at a much higher frequency than sampling by Fairfax Water. To avoid inundating the dataset with interpolated data, one sampling event per month was randomly selected for interpolation with the equation from the linear regresion (Appendix C). This resulted in two of the three major gaps in the sodium data being filled and one gap in the chloride data being filled 3.2. Overall, this resulted in a more consistent n-values across the decades when data from ST01 were available (Table 3.2).

Table 3.2: Number of sodium and chloride samples of finished water from Griffith Water Treatment Plant before and after interpolation.

	Sodium Interpolation		Chloride Interpolation	
Date Range	Before	After	Before	After
1980 - 1989	270	270	1165	1165
1990 - 1999	214	214	377	377
2000 - 2009	157	199	168	174
2010 - 2019	62	105	62	106
Total	703	788	1772	1822

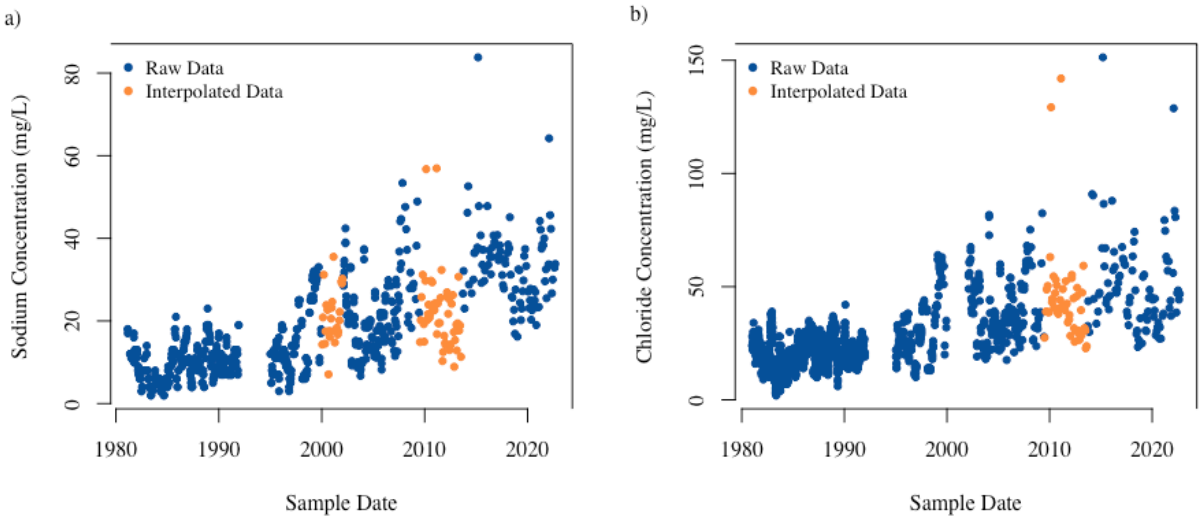


Figure 3.3: Interpolation results for Fairfax Water's a) sodium and b) chloride dataset.

3.3 Comparison of Statewide Drinking Water Concentrations

To understand how the sodium and chloride concentrations in this system compare to statewide values, the median sodium concentrations from Fairfax Water were compared to the median values for individual public water systems (PWS) statewide.

The Virginia Department of Health (VDH) provided sodium and chloride data for PWS across Virginia. Sodium data was available consecutively from 2002 - 2022. The data was divided into 4 five-year increments, beginning in 2003. The dataset was divided between systems that receive their raw drinking water from either surface water or purchase surface water (n = 151) and those that use groundwater as their source water (n = 1,179).

The population served by systems using surface water as a raw water source ranged from 50 to over 1 million customers, with a median of 5,785. 13 systems did not report the population they served and as such, were not included in this analysis.

For systems using groundwater as the raw water source, the population served ranged from 15 to over 50,000, with a median of 130. Of the 1,179 systems, three did not sample for sodium and 10 did not sample for chloride. There was no overlap between these sets.

Systems that used surface water for raw water tended to serve a larger population. Assuming that size of population served is a reasonable surrogate for size of the drinking water plant itself, it can be assumed that these systems are larger and therefore, likely have a higher level of treatment. Systems using groundwater for raw water are then typically smaller, which aligns with groundwater systems being more common in rural, less populated areas.

Table 3.3: Summary statistics of public water systems whose data were provided by the Virginia Department of Health. Surface water as a raw water source accounts for both systems that directly intake from a surface water and those that purchase surface water.

		Population Served		
Raw Water Source	No. of PWS	Minimum	Maximum	Median
Surface water	151	50	1,121,613	5,785
Groundwater	1,179	15	51,137	129

For each time period, an empirical CDF of the median values for each individual PWS was created using Equation 3.1. Several medians were calculated to be 0.0. Due to uncertainty in any changes to the limit of detection over time, a nominal limit of detection of 0.01 mg/L was assumed. Therefore, any value reported below 0.01 mg/L was treated as 0.01 mg/L.

This analysis was first completed for all PWS in Virginia that receive their raw drinking water from freshwater sources and then repeated for PWS that receive their raw drinking water from groundwater sources. The median sodium and chloride concentrations for the same time period was found from the interpolated dataset of Fairfax Water's sodium concentrations, described in Section 3.2.1.

These analyses show that the median sodium concentration from Fairfax Water almost doubled from 2003 - 2017, from 18 mg/L to 35 mg/L. The concentration then decreased from 35 mg/L during 2013 - 2017 to 30 mg/L 2018 - 2022. Whether this shows the concentration is leveling out or decreasing remains to be seen. Additionally, the median concentration of sodium has been above EPA's DWEL of 20 mg/L since 2008.

Conversations with stakeholders could help illuminate whether changes in the watershed over this time period could be responsible for these increased concentrations. One such change was the discharge of industrial wastewater from a semiconductor manufacturer to UOSA

begin in late 1996.[10] Previous investigations have determined that this manufacturer is responsible for 13.8% of the annual sodium load in UOSA's discharge.[4]

When compared to other drinking water utilities that receive their raw intake water from freshwater sources, Farifax Water's sodium levels have increased from the 75^{th} to 93^{rd} percentiles since 2003. Median chloride concentrations have consistently been among the highest in the state during the time period analyzed, ranging from the 95^{th} to 98^{th} percentile.

As such, this drinking water is already saltier than almost any other drinking water that receives it's raw water from freshwater sources. Any additional input of sodium or chloride would likely result in Fairfax Water having the saltiest drinking water in the state, which is important context for future decisions on what level of sodium is acceptable in the reservoir.

Further investigations of this dataset are needed to determine how Fairfax Water's drinking water compares to it's regional counterparts and how it compares to statewide to drinking water systems of a similar size.

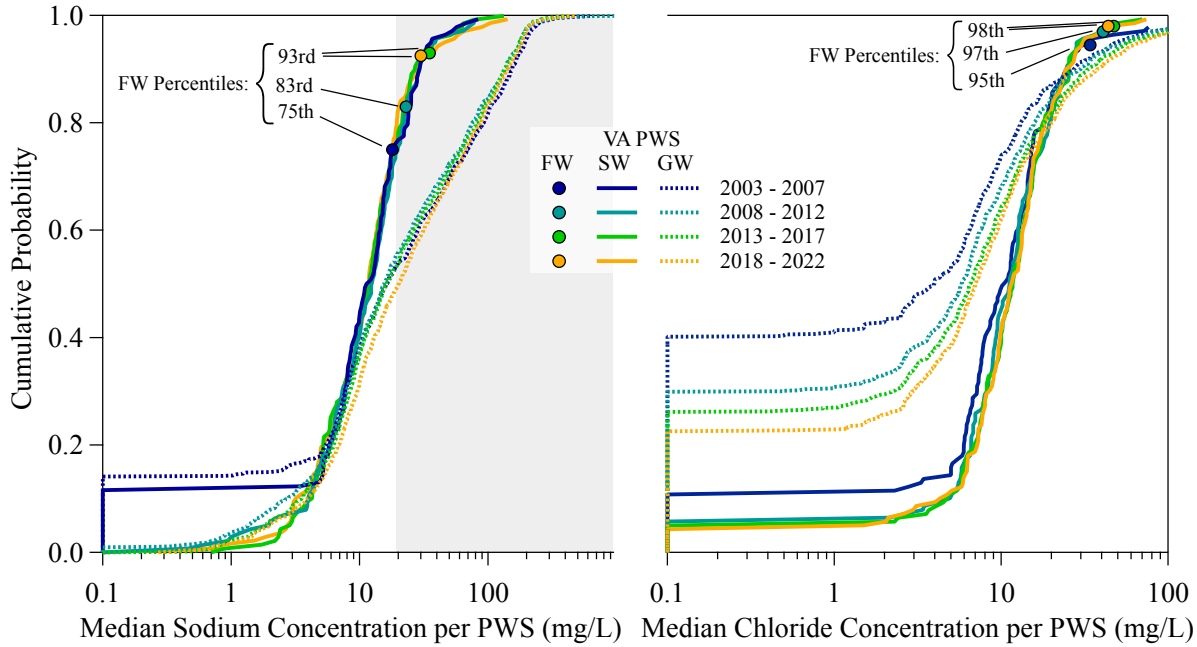


Figure 3.4: Comparison of sodium and chloride concentrations between Fairfax Water's (FW) with median concentrations of PWS statewide. PWS using surface water for raw water are indicated with solid lines, while those using groundwater for raw water are indicated with dashed lines. Fairfax Water's median concentration are marked on each CDF. Time periods are differentiated by color. On the right panel, concentrations above EPA's DWEL of 20 mg/L are shown in grey.

3.4 Trends of Reservoir Sodium Concentrations

Over time, sodium concentrations in drinking water have increased and become more varied (Figure 3.5). Over the past four decades, the median sodium concentration has increased: 9 mg/L in the 1980s, 11 mg/L in the 1990s, 19.6 mg/L in the 2000s, and 28.8 mg during the 2010s (Figure 3.5).

For each decade of sodium data, the normal distribution parameters of the ln-transformed data (mean (μ_{lnNa}), standard deviation (σ_{lnNa})) were found using statistical software (Appendix D). A linear regression was done on μ_{lnNa} and σ_{lnNa}; both showed strong linear relationships with time, with R^2 values of 0.9935 for μ_{lnNa} and 0.8767 for σ_{lnNa} (Figure 3.6).

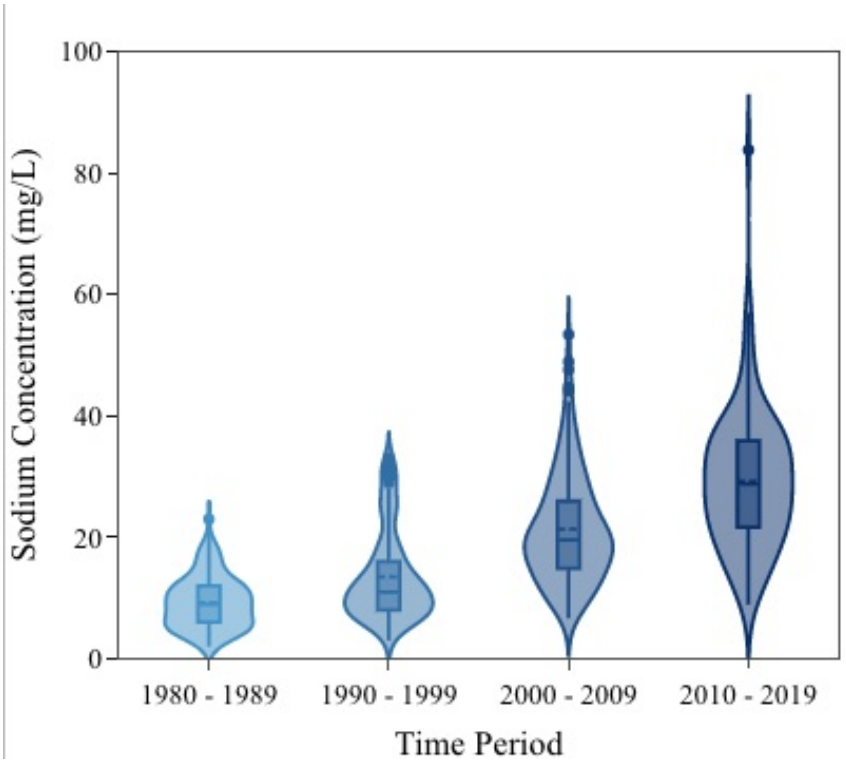


Figure 3.5: Violin and box plots of sodium concentration from Fairfax Water's drinking water over different decades (1980 - 2019).

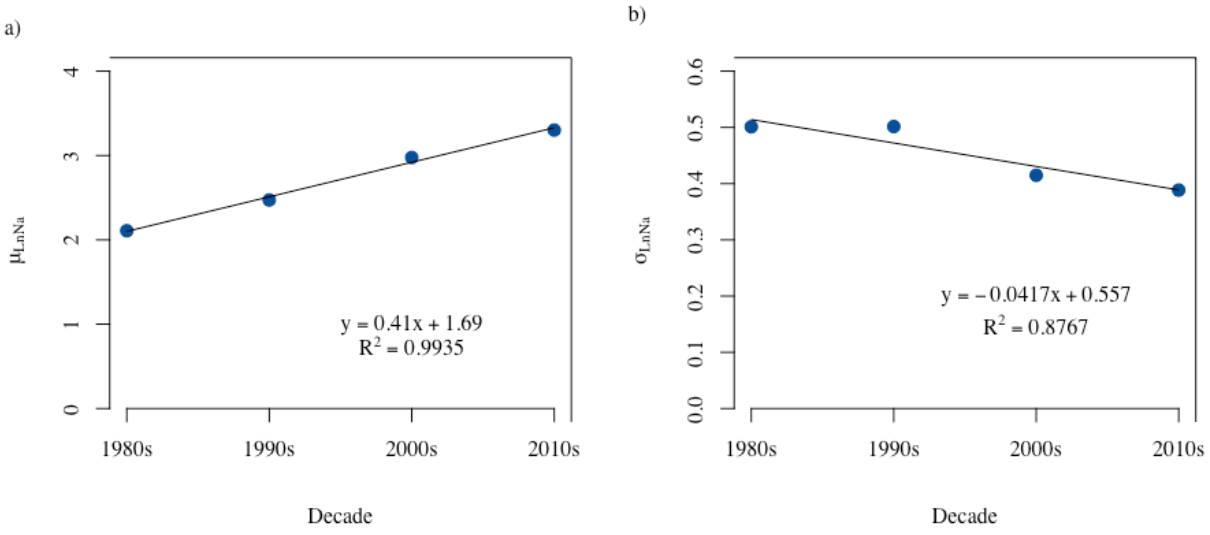


Figure 3.6: Regression analyses between a) μ_{lnNa} and b) σ_{lnNa} with time. Linear relationships and R^2 values are shown for each regression.

These regressions were used to determine what the μ_{lnNa} and σ_{lnNa} for future decades would be, if this trend were to continue. This is a simple linear extrapolation, which assumes sodium concentration increases continue on the current trajectory.

Estimates of the μ_{Na} and σ_{Na} were calculated from μ_{lnNa} and σ_{lnNa} using the following formulas for transforming the estimated log mean and log variance to an estimated arithmetic mean and arithmetic standard deviation[16]:

$$\mu_{Na} = e^{((\mu_{lnNa} + \frac{(\sigma^2_{lnNa})}{2}))} \tag{3.2}$$

$$\sigma_{Na} = \sqrt{(e(2\mu_{lnNa} + \sigma^2_{lnNa})(e^{\sigma^2_{lnNa}} - 1))} \tag{3.3}$$

The 10^{th} and 90^{th} percentiles for each decade were also calculated from the normal distributions of the ln-transformed data for each decade, using μ_{lnNa} and σ_{lnNa}, and then exponentiated into arithmetic values.

Should this trend continue, the concentration of sodium in drinking water would increase rapidly in the coming decades. Beginning with this decade, the 2020s, the 90^{th} percentile sodium concentration will exceed EPA's upper Taste Threshold of 60 mg/L. By the 2030s, the median concentration will exceed 60 mg/L, and by the 2040s, the majority of all drinking water will have sodium concentrations in excess of 60 mg/L.

Increasing sodium concentrations can alter the perception of potability, which may in turn alter the perception of indirect potable reuse as a viable solution to drinking water supply challenges.[8]

Table 3.4: Calculated and projected sodium mean (μ_{lnNa}, μ_{Na}), standard deviation (σ_{lnNa}, σ_{Na}), and the 10^{th} and 90^{th} percentiles for each decade. Distributions were assumed to be normal.

	Ln(Sodium)		Sodium (mg/L)			
Date Range	μ_{lnNa}	σ_{lnNa}	μ_{Na}	σ_{Na}	10^{th} %	90^{th} %
1980 - 1989	2.11	0.50	9.33	4.99	1.47	15.65
1990 - 1999	2.47	0.50	13.45	7.19	1.83	22.55
2000 - 2009	2.99	0.41	21.71	9.34	2.45	33.39
2010 - 2019	3.30	0.39	29.29	11.82	2.80	44.68
2020 - 2029	3.74	0.34	44.84	15.90	3.29	65.29
2030 - 2039	4.15	0.30	66.64	20.56	3.76	93.03
2040 - 2049	4.56	0.6	99.21	26.11	4.22	132.55
2050 - 2059	4.97	0.22	147.99	32.36	4.68	188.86

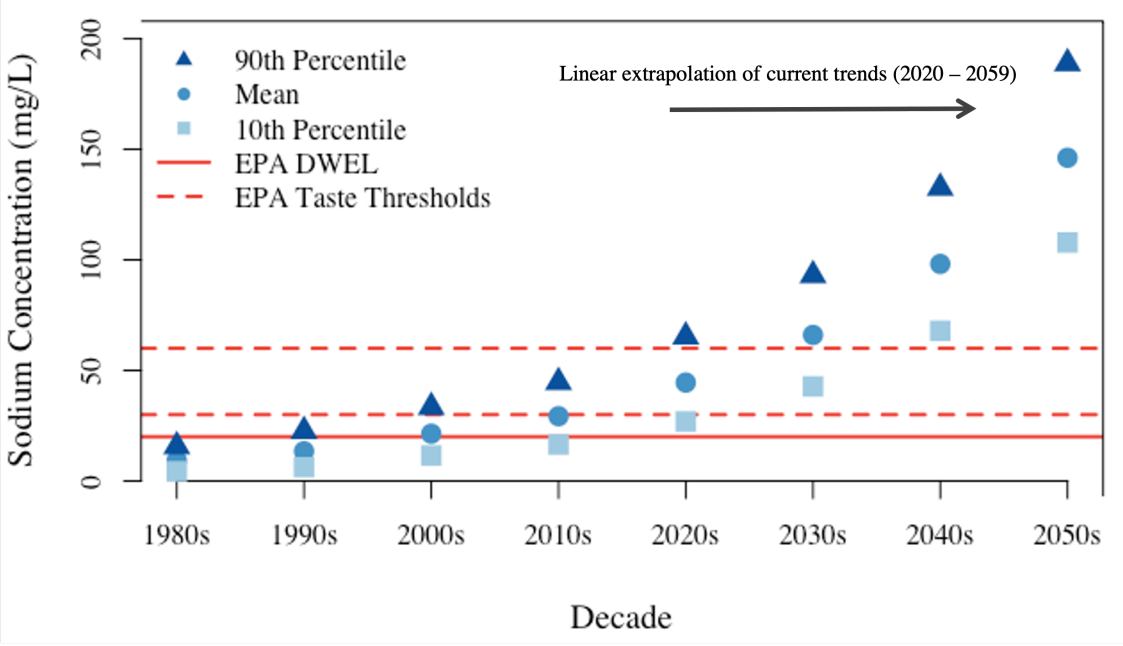


Figure 3.7: Trends of sodium concentration in Fairfax Water's finished drinking water (1980 - 2019). Historical decades (1980 - 2019) are based on Fairfax Water's finished drinking water and the interpolated data from ST01. Future concentrations (2020 - 2059) are based on a worst case scenario, where this trend continues and no interventions are taken. EPA's Drinking Water Equivalent Level (DWEL) of 20 mg/L and taste thresholds of 30 - 60 mg/L are delineated.

3.5 Percent of Daily Sodium Ingested from Drinking Water

To determine whether there are health implications from the sodium concentration in drinking water increasing over time, Monte Carlo simulations of the percent of daily sodium ingested in drinking water were conducted using the dosing equation formula presented in Cruz, et.al (Equation 1.1).

The fitted normal distributions for each decade of ln-transformed sodium data were used (Section 3.4, allowing a sodium concentration to be chosen at random. The volume of drinking water ingested was also selected at random from EPA's consumer-only estimates of combined direct and indirect water ingestion of community water for all ages (Figure 1.2). In total, 16 scenarios were evaluated, with two scenarios for each decade, one looking at a standard diet (TUL: 2300 mg/day) and one for a low salt diet (TUL: 1500 mg/day) (Appendix E). For each scenario, the simulations were ran for 100,000 realizations, the results were logarithmically transformed and histogrammed (Figure 3.8).

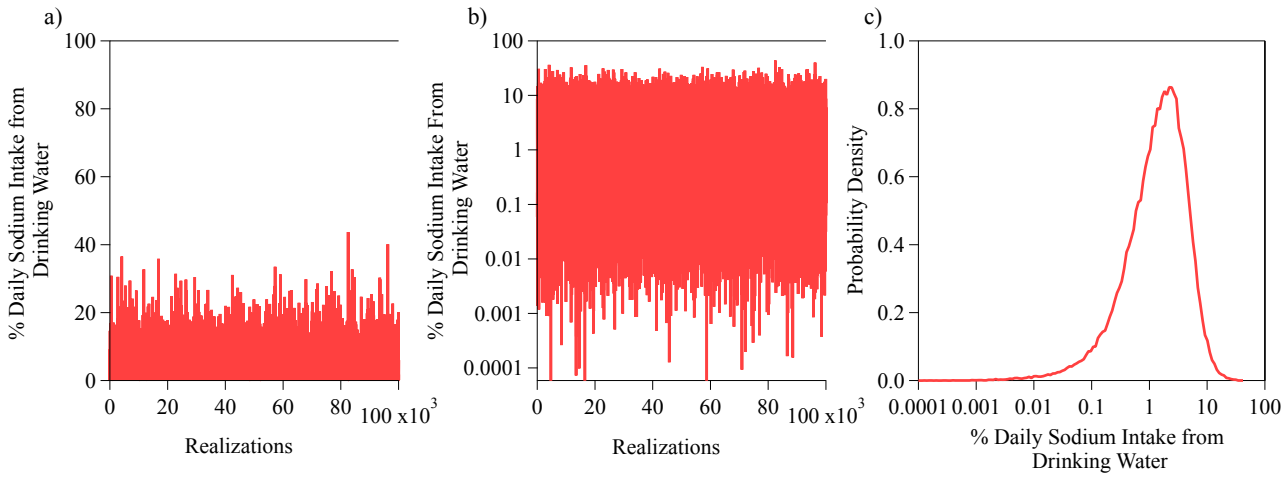


Figure 3.8: Example of the methods used for Monte Carlo simulations. Simulations were ran for 100,000 realizations (a). Results were log-transformed (b) and histogrammed, which was normalized to a probability distribution function (c).

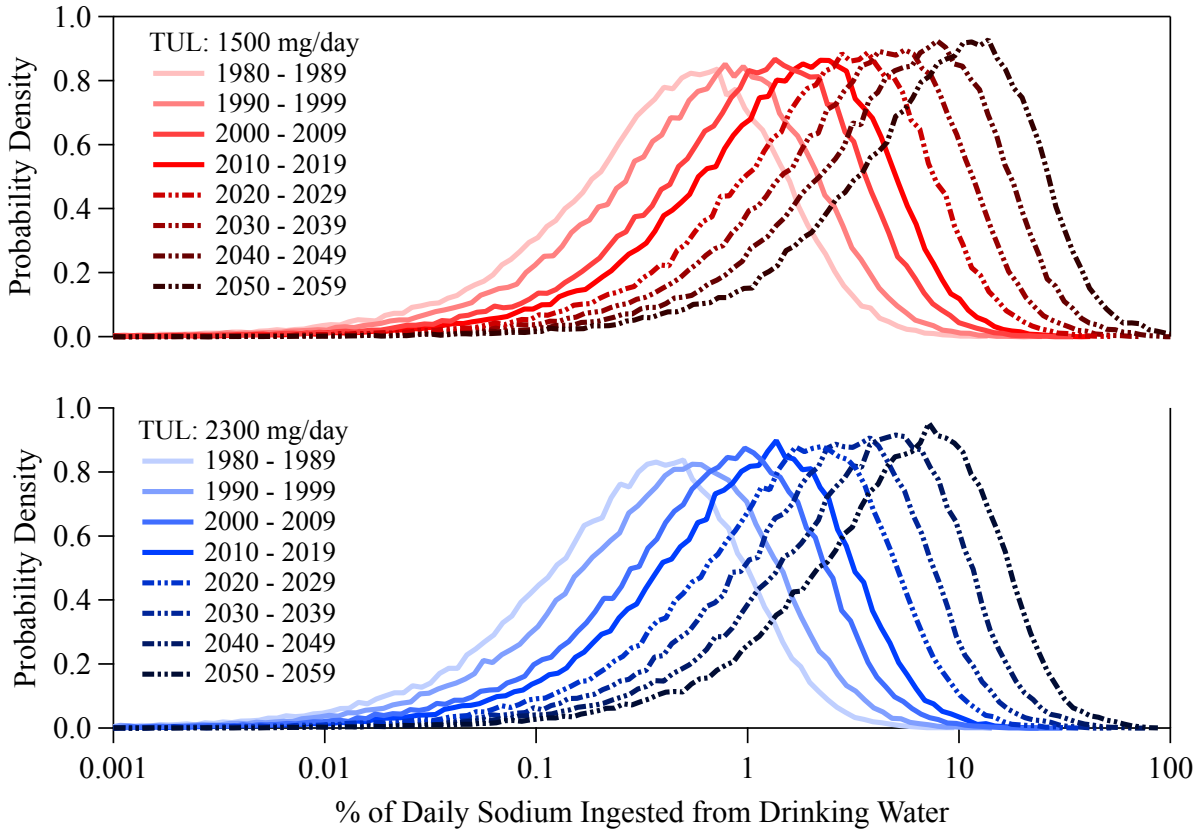


Figure 3.9: Results of the Monte Carlo simulations. Simulations were ran for each decade and for low sodium (TUL: 1500 mg/day) and standard diets (TUL: 2300 mg/day). Solid lines represent decades where the sodium concentration CDFs used in the simulation were based on historical data. Dashed lines represent decades where the sodium concentration CDFs used in the simulations were based on current trends continuing.

As expected, the Monte Carlo simulations show the intake of sodium from drinking water increasing over time, with those on a low-salt diet intaking a higher percentage than those on a standard diet. The results show a progressive increase between each consecutive decade. However, each decade shows a wide range of possible intakes. Because the drinking water distribution used was very broad, the amount of water ingested is a major driver in the final intake result.

To better understand how the sodium intake levels are changing with time, the percentage of the population intaking greater than or equal to 10% of their daily sodium from drinking water was calculated using the following formula:

$$P\left(x \geq 0.1\right) = 1 - \int_0^{0.1} p_x(x)dx \tag{3.4}$$

These results show several things. First, historically, the intake of sodium from drinking water is remarkably low. From the 1980s through the 2010s, less than 1.3% of the population on a low salt data received greater than or equal to 10% of their salt intake from drinking water. However, under this worst case scenario, this value would increase exponentially. By the end of the 2050s, this trend indicates that 23% of the population on a standard diet and 40% of the population on a low salt diet will be ingesting greater than 10% of their daily sodium from drinking water.

The timing of the increase is also significant, as this trend indicates the rapid increase will begin in the present decade. This reinforces the need for management actions to be undertaken now to prevent the system from reaching these levels.

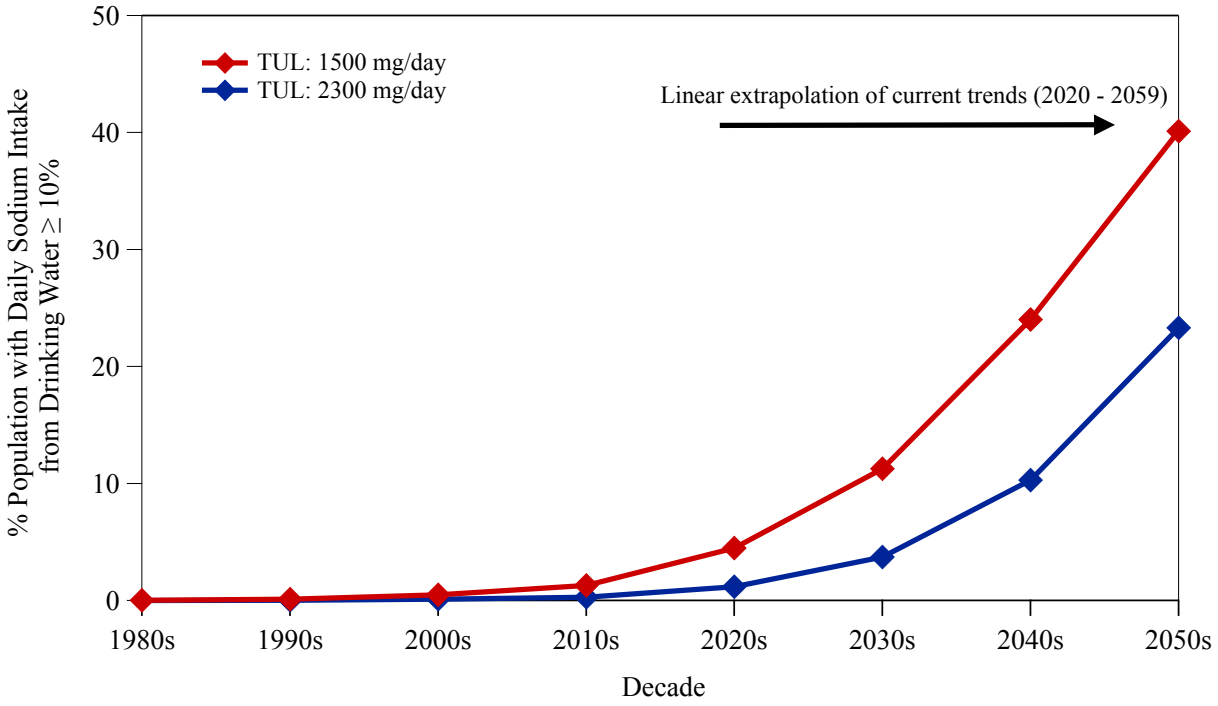


Figure 3.10: Percent of population ingesting greater than 10% of their daily sodium from drinking water over time. This analysis was completed for individuals on a low salt (TUL: 1500 mg/day) and standard diet (TUL: 2300 mg/day).

Chapter 4

Discussion & Conclusion

Current water regulations, such as the Virginia's WQS and VPDES regulations, look at multiple beneficial uses for a waterbody, such as aquatic life, recreation, and a public water supply (9VAC25-260-10).[33] The most stringent water quality standard is used when developing a VPDES permit limit, allowing for all beneficial uses to be protected (9VAC25-31).[34]

Similarly, the intention behind these different analyses was to frame the sodium concentrations in various ways to account for different concerns and perspectives, including: 1) how sodium levels have been increasing over time and what this trend could suggest for the coming decades if no management steps are taken, 2) what human health implications there may be from these increases, 3) how the concentrations in drinking water compare to other PWS statewide, and 4) how the sodium and chloride concentrations in the watershed compare statewide.

These analyses show that drinking water from this reservoir have much higher levels of sodium and chloride relative to other drinking water derived from freshwater sources. This results in the population ingesting increasingly higher levels of sodium from their drinking water, which depending on the volume of water a person typically drinks, could reach hazardous levels. The watershed also has some of the highest concentrations of sodium and chloride statewide, which effects the benthic invertebrates community and other aquatic life, as well as the water quality in the raw drinking water. Most alarming, these trends show

these levels increasing exponentially by the end of the decade, should no management actions be implemented.

These analyses show the need for proactive management of salinization in the reservoir. When choosing a management threshold, stakeholders should consider several options, including:

- Is there a historic sodium level that is protective of the reservoir's beneficial uses?
- When comparing to other PWSs, is there a percentile range that is an appropriate aim?
- Is there an acceptable percentage of sodium ingested from drinking water? In setting recommendations in the past, EPA has considered a dietary intake of 2.5 - 5% from drinking water as reasonable. Is this appropriate today?
- Should there be multiple thresholds, with different management actions triggered when crossed?

Considering the multiple beneficial uses of the reservoir will ensure the threshold is appropriate long term. Given the significant resources being deployed to address this issue, it is unlikely such an effort could be replicated again. As such, it is essential that any threshold(s) chosen be appropriate long term and is protective of the various beneficial uses of the watershed.

Such a framework is consistent with current water protection regulations, which many of the stakeholder have some level of familiarity with. If the stakeholders already have a level of comfort with or trust in the current regulatory system, they may be more accepting of threshold(s) developed in a similar manner. Using this framework would therefore be advantageous as researchers and stakeholders work towards defining what the importance of this reservoir is and how it should be protected.

Bibliography

[1] AAFP. Lifestyle changes to help lower your blood pressure. Available at: https://familydoctor.org/lifestyle-changes-to-help-lower-your-blood-pressure/. Last accessed 4 February 2023.

[2] Aang, A., and Tang, W. *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*, 2nd ed. 2006. ISBN: 978-0471720645.

[3] AHA. How much sodium should i eat per day? Available at: https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/sodium/how-much-sodium-should-i-eat-per-day. Last accessed 4 February 2023.

[4] Bhide, S., Grant, S., Parker, E., Rippy, M., Godrej, A., Kaushal, S., Prelewicz, G., Saji, N., Curtis, S., Vikesland, P., Maile-Moskowitz, A., Edwards, M., Lopez, K., Birkland, T., and Schenk, T. Addressing the contribution of indirect potable reuse to inland freshwater salinization. *Nature Sustainability 4*, 8 (2021), 699–707.

[5] CDC. Sodium. Available at: https://www.cdc.gov/heartdisease/sodium.htm. Last accessed 4 February 2023.

[6] CDC. Water and healthier drinks. Available at: https://www.cdc.gov/healthyweight/healthy_eating/water-and-healthier-drinks.html. Last accessed 4 February 2023.

[7] Cruz, Y., Rossi, M., and Goldsmith, S. Impacts of road deicing application on sodium and chloride concentration in philadelphia region drinking water. *GeoHealth 6*, 2 (2022).

[8] Dietrich, A., and Burlingame, G. Critical review and rethinking of usepa secondary standards for maintaining organoleptic quality of drinking water. *Environmental Science & Technology 49*, 2 (2015), 708–720.

[9] Grant, S., Rippy, M., Birkland, T., Schenk, T., Rowles, K., Misra, S., Aminpour, P., Vikesland, P., Bergland, E., Gomez-Velez, J., Hotchkiss, E., Perez, G., Zhang, H., Armstrong, K., Bhide, S., Krauss, L., Maas, C., Mendoza, K., Shipman, C., Zhang, Y., and Zhong, Y. Can common pool resource theory catalyze stakeholder-driven solutions to the freshwater salinization theory? *Environmental Science and Technology 56*, 19 (2022), 13517–12527.

[10] Grizzard, T. Dominion semiconductor project. official memorandum, 1996. Director, Occoquan Watershed Monitoring Lab.

[11] Kaushal, S., Likens, G., Pace, M., Utz, R., Haq, S., and ad Melissa Grese, J. G. Freshwater salinization syndrome on a continental scale. *Proceedings of the National Academy of Sciences 115*, 4 (2018).

[12] NASEM. *Dietary Reference Intakes for sodium and potassium.* National Academies Press, Washington, DC, 2019.

[13] NIH. Dash eating plan. Available at: https://www.nhlbi.nih.gov/education/dash-eating-plan. Last accessed 4 February 2023.

[14] NIH. Sodium: tips for people with chronic kidney disease (ckd). Available at: https://www.niddk.nih.gov/health-information/kidney-disease/chronic-kidney-disease-ckd/eating-nutrition, 2011. Last accessed 4 February 2023.

[15] Poeper, K., Min, T., Jones, N., Weiss, S., Greene, A., Mohin, H., Parks, J., and Edwards, M. Impact of road salt on drinking water quality and infrastructure corrosion in private wells. *Environmental Science and Technology 52*, 24 (2018), 14078–14087.

[16] Rippy, M. Random variables and probability distribution, 2021. Lecture notes in Data Analysis for Environmental Science and Engineering.

[17] Saji, N. Private Communication. Water Resources Engineer, Fairfax Water.

[18] Schuler, M., Cañedo-Argüelles, M., Hintz, W., Dyack, B., Birk, S., and Relyea, R. Regulations are needed to protect freshwater ecosystems from salinization. *Philosophical Transactions of the Royal Society B: Biological Sciences 374*, 1764 (2019), 20180019.

[19] Sivers, S. Private Communication, November 2022. Water Permits, Planning, and Monitoring Manager, Northern Regional Office, DEQ.

[20] Stets, E., Sprague, L., Oelsner, G., Johnson, H., Murphy, J., Ryberg, Kand Vecchia, A., Zuellig, R. E., Falcone, J., and Riskin, M. Landscape drivers of dynamic change in water quality. *Environmental Science and Technology 54*, 7 (2020), 4336–4343.

[21] USDA, and USDHHS. *Dietary Guidelines for Americans, 2020 - 2025*, 9th ed. 2020.

[22] USEPA. National primary drinking water regulations. Available at: https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations. Last accessed 5 February 2023.

[23] USEPA. National secondary drinking water regulations. Available at: https://www.epa.gov/sdwa/secondary-drinking-water-standards-guidance-nuisance-chemicals. Last accessed 5 February 2023.

[24] USEPA. Contaminant candidate list regulatory determination support document for sodium. Tech. Rep. EPA 815-R-03-15, U.S. Environmental Protection Agency, July 2003.

[25] USEPA. Drinking water advisory: Consumer acceptability advice and health effect analysis on sodium. Tech. Rep. EPA 822-R-03-006, U.S. Environmental Protection Agency, February 2003.

[26] USEPA. Exposure factors handbook, chapter 1: Introduction. Tech. Rep. EPA/600/R-09/052F, U.S. Environmental Protection Agency, September 2011.

[27] USEPA. Exposure factors handbook, chapter 3: Ingestion of water and other select liquids. Tech. Rep. EPA/600/R-18/259F, U.S. Environmental Protection Agency, February 2019.

[28] USFDA. Sodium in your diet. Available at: https://www.fda.gov/food/nutrition-education-resources-materials/sodium-your-diet. Last accessed 4 February 2023.

[29] USGAO. Freshwater: Supply concerns continue, and uncertainties complicate planning. Tech. Rep. GAO-14-430, United States Government Accountability Office, May 2014.

[30] VaDEQ. Probabilistic monitoring. Available at: www.deq.virginia.gov/water/water-quality/monitoring/probabilistic-monitoring. Last accessed 4 February 2023.

[31] VaDEQ. Tmdl development. Available at: https://www.deq.virginia.gov/water/water-quality/tmdl-development. Last accessed 27 February 2023.

[32] VaDEQ. 2022 305(b)/303(d) water quality assessment integrated report. Available at: https://www.deq.virginia.gov/water/water-quality/assessments/integrated-report, 2022.

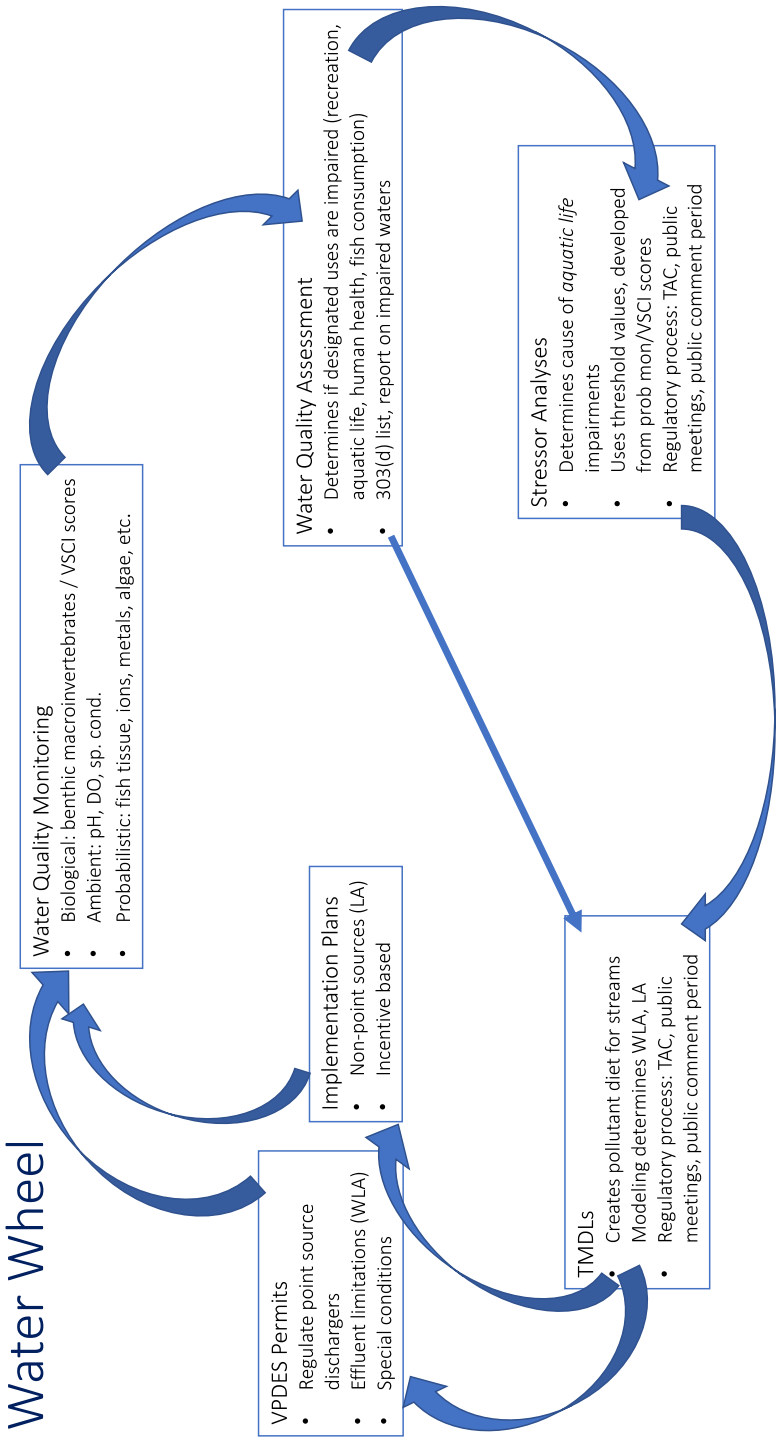
[33] Water quality standards. *Virginia Administrative Code*, 9VAC25-260 (2019).

[34] Virginia pollution discharge elimination system (vpdes) permit regulation. *Virginia Administrative Code*, 9VAC25-31 (2022).

Appendices

Appendix A

Virginia's Regulatory Water Wheel


Water Wheel
Water Quality Monitoring
• Biological: benthic macroinvertebrates / VSCI scores
• Ambient: pH, DO, sp. cond.
• Probabilistic: fish tissue, ions, metals, algae, etc.
Water Quality Assessment
• Determines if designated uses are impaired (recreation, aquatic life, human health, fish consumption)
• 303(d) list, report on impaired waters
Stressor Analyses
• Determines cause of *aquatic life* impairments
• Uses threshold values, developed from prob mon/VSCI scores
• Regulatory process: TAC, public meetings, public comment period
Implementation Plans
• Non-point sources (LA)
• Incentive based
TMDLs
• Creates pollutant diet for streams
Modeling determines WLA, LA
• Regulatory process: TAC, public meetings, public comment period
VPDES Permits
• Regulate point source dischargers
• Effluent limitations (WLA)
• Special conditions

Appendix B

Code for Date Matching & Linear Regressions (R)

```
#Thresholds for the Occoquan Reservoir

###PART1: Interpolation of ST01 sodium data to fill in the gaps of the FFX
water dataset

#Step 1: Compare dates between ST01 and FFX water

#clear global environment
rm(list=ls())

#upload data
library(readxl)

ST01_Na <- as.data.frame(read_excel("~/Documents/Virginia Tech/Research/
Sodium Ingestion/FFX_ST01_Analysis.xlsx",
                    sheet = "ST01"))
ST01_Na$DATE <- as.POSIXct(ST01_Na$DATE, format = "%m/%d/%y", tz ="EST")
#assigns date format
ST01_Na$ID <- 1
ST01_Na = ST01_Na[c("ID","DATE","NA_ST01")] #reorders columns


FFX_Na <- as.data.frame(read_excel("~/Documents/Virginia Tech/Research/Sodium
Ingestion/FFX_ST01_Analysis.xlsx",
                   sheet = "FFX"))
FFX_Na$DATE <- as.POSIXct(FFX_Na$DATE, format ="%m/%d/%y", tz = "EST")
#assigns date format
FFX_Na$ID <- 1
FFX_Na = FFX_Na[c("ID","DATE","NA_FFX")] #reorders columns

#filtering to match FFX sample date to closest ST01 sample date
library(data.table)

dt1 <- setDT(FFX_Na)
dt2 <- setDT(ST01_Na)

setkey(dt2, ID, DATE)[, datematch:= DATE]

DATA_filtered_Na = dt2[dt1, roll = 'nearest'] #wrap longer dataset around
shorter dataset
setDF(DATA_filtered_Na)

View(DATA_filtered_Na)

####
#find samples within one date of each other

#subtract to find difference
DATA_filtered_Na$difference = DATA_filtered_Na$DATE -
DATA_filtered_Na$datematch

#sort matrix by difference
DATA_filtered_Na[order(DATA_filtered_Na[,6],decreasing = FALSE),]
```

```
#1 day = 86400 seconds
#if difference = 86400, FFX water sample was taken one day after the ST01
samples
#if difference = 0, samples taken on same day

#Step 2: linear regression of STO1 and FFX water data with 1 day difference

#select data from DATA_filtered
ST01_regress_Na = DATA_filtered_Na[DATA_filtered_Na$difference == 86400, 3]

FFX_regress_Na = DATA_filtered_Na[DATA_filtered_Na$difference == 86400, 5]

#Linear regression between ST01_regress & FFX_regress

#visualize
plot_Na =
  plot(ST01_regress_Na, FFX_regress_Na,
     col = "blue", pch = 19,
     xlab = 'Sodium (mg/L) at ST01',
     ylab = 'Sodium (mg/L) at FFX Water' )


#interested in how the  concentration of sodium in drinking water (FFX)
(dependent variable, y-axis)
#relates to the concentration at ST01(independent variable, x-axis)

#fit a line to the relationship between FFX and ST01 using Nelder-Mead
Simplex Algorithem

#create function to minimize

#SE - sum of residuals squared
#BETA[1] - slope
#BETA[2] - y intercept

SE = function (BETA) {sum((FFX_regress_Na -
                             (BETA[1]*ST01_regress_Na + BETA[2]))^2)}

#find B1 and B0 that minimize function using the Nelder-Mead Simplex
Algorithm

library(pracma)

STRUCT = fminsearch(SE, c(100, 0.5))

B_Na = STRUCT$xmin

#plot best fit line

Y_Na = B_Na[1]*(ST01_regress_Na) + B_Na[2]

lines(ST01_regress_Na, Y_Na, col = "black")

#B values
```

```
#Beta[1] = 0.9649 (slope)
#Beta[2] = 1.9540 (intercept)
#trendline: y = 0.9469x+1.9540

#find R^2 value
R_Na = cor(cbind(ST01_regress_Na, FFX_regress_Na))
RR_Na = R^2
RR_Na
#RR is 0.740811

text(19, 80, expression(y == 0.9469*x + 1.9540)) #add trendline to plot

text(15, 77, expression(RR^2 == 0.741)) #add R^2 to plot

#create matrix of ST01 sample date, ST01 data (x values), interpolated FFX
data (y values)
INTERP_Na = cbind.data.frame(ST01_Na$DATE, ST01_Na$NA_ST01,
ST01_Na$NA_ST01*0.9649+1.9504)

#export to Excel
library(openxlsx)

write.xlsx(DATA_filtered_Na, "Desktop")
write.xlsx(INTERP_Na, "Desktop")

####PART 2
#Recognized ST01 had several sample events per month, whereas FFX sampled
monthly
#In a Juptyer Notebook, wrote code to randomly select ST01 sample event per
month during date gaps (2000 - 2002, July 2009 - August 2013)
#Upload results

FFX_interp_Na <- read_excel("~/Documents/Virginia Tech/Research/Sodium
Ingestion/Python_Output_ST01_update.xlsx")

par(family = "serif")

plot2 =
  plot(FFX_Na$DATE, FFX_Na$NA_FFX,
     col = "blue",
     pch = 19,
     frame = FALSE,
     xlab = 'Sample Date',
     ylab = 'Sodium (mg/L)' )

  points(FFX_interp_Na$`Sample Date`, FFX_interp_Na$Sodium,
     col = "orange", pch = 19)

  legend("topleft",
         legend = c("Raw Data", "Interpolated Data"),
         col = c("blue","orange"),
         pch = c(16, 16),
         cex = 0.9,
         bty = 'n'
```

```
        )
########
#plot mean, 90th, and 10th percentiles for each decade for SODIUM
#MU_LN and SIGMA_LN for each decade are from Mathematica

DECADE = cbind(1:8)

#qnorm assumes normal distribution, therefore using MU_LN and SIGMA_LN to
calculate 90_LN & 10_LN
  #and then exponentiate the results to get percentiles in mg/L

MU_LN = cbind(2.10817, 2.4730, 2.97664, 3.30171, 3.7358, 4.1442, 4.5526,
4.9610)
SIGMA_LN = cbind(0.501145, 0.501565, 0.414817, 0.388495, 0.3457, 0.3033,
0.2609, 0.2185)

NINETIETH_LN = matrix('', nrow = 1, ncol = 8)

#calculate 90th percentile for normal distribution from MU_LN and SIGMA_LN
for(i in 1:8){
  NINETIETH_LN <- qnorm(0.9, MU_LN, SIGMA_LN)
}

#exponeniate to convert to mg/L
NINETIETH = exp(NINETIETH_LN)

#repeat for 10th percentile
TENTH_LN = matrix('', nrow = 1, ncol = 8)

for(i in 1:8){
  TENTH_LN <- qnorm(0.1, MU_LN, SIGMA_LN)
}

TENTH = exp(TENTH_LN)

#########
###compare with violin plots

#Q1
Q1_LN = matrix('', nrow = 1, ncol = 8)

for(i in 1:8){
  Q1_LN <- qnorm(0.25, MU_LN, SIGMA_LN)
}

Q1 = exp(Q1_LN)
##########

#calculate MU from MU_LN
#formula from Meg's notes
MU = exp(MU_LN+((SIGMA_LN)^2)/2)

#plot the results
```

```
#set up list of names to be plotted instead of decade numbers (1, 2, 3,...)
DECADE_names = cbind('1980s', '1990s', '2000s', '2010s', '2020s', '2030s',
'2040s', '2050s')
#DECADE_names2 = cbind('1980 - 1989', '1990 - 1999', '2000 - 2009', '2010 -
2019', '2020 - 2029', '2030 - 2039', '2040 - 2049', '2050 - 2059')

par(family = "serif", mfrow=c(1,1))

plot3 =

  plot(DECADE, MU,
       #log = "y",
       col = "#4292C6", pch = 16,
       cex.lab = 1.2, cex.axis = 1, cex = 1.6,
       #frame = TRUE,
       bty = "7",
       xaxt = 'n', #turn off automatic x axis
       xlab = 'Decade',
       ylab = 'Sodium Concentration (mg/L)',
       ylim = c(1, 200),
       #add EPA thresholds behind the data
       panel.first = c(
         abline(h = 20, col = "#EF3B2C", lwd = 2),
         abline(h = 30, col = "#EF3B2C", lwd = 2, lty = 2),
         abline(h = 60, col = "#EF3B2C", lwd = 2, lty = 2)))

axis(1, at = DECADE,  labels = DECADE_names, pos = 0, cex.axis = 1) #new and
improved x axis
#text(srt = 35, adj = 0.965) ## Adjust the labels

points(DECADE, NINETIETH,
       col = "#08519C", pch = 17, cex = 1.6)

points(DECADE, TENTH,
       col = '#9ECAE1', pch = 15, cex = 1.6)

legend("topleft",
       legend = c("90th Percentile","Mean","10th Percentile", "EPA DWEL",
"EPA Taste Thresholds"),
       col = c("#08519C","#4292C6",'#9ECAE1', "#EF3B2C", "#EF3B2C"),
       pch = c(17, 16, 15, NA, NA),
       lty = c(NA, NA, NA, 1, 2),
       lwd = c(NA, NA, NA, 2, 2),
       cex = 1,
       inset = c(0.01, 0.01),
       bty = 'n')

######################
#repeat interp with chloride data

#Chloride Interpolation

###PART1: Interpolation of ST01 sodium data to fill in the gaps of the FFX
water dataset
```

```
#Step 1: Compare dates between ST01 and FFX water

ST01_Cl <- as.data.frame(read_excel("~/Documents/Virginia Tech/Research/
Sodium Ingestion/Chloride_Analysis.xlsx",
                                    sheet = "ST01_CL"))
ST01_Cl$DATE <- as.POSIXct(ST01_Cl$DATE, format = "%m/%d/%y", tz ="EST")
#assigns date format
ST01_Cl$ID <- 1
ST01_Cl = ST01_Cl[c("ID","DATE","ST01_CL")] #reorders columns


FFX_Cl <- as.data.frame(read_excel("~/Documents/Virginia Tech/Research/Sodium
Ingestion/Chloride_Analysis.xlsx",
                                   sheet = "FFX_CL"))
FFX_Cl$DATE <- as.POSIXct(FFX_Cl$DATE, format ="%m/%d/%y", tz = "EST")
#assigns date format
FFX_Cl$ID <- 1
FFX_Cl = FFX_Cl[c("ID","DATE","FFX_CL")] #reorders columns

#filtering to match FFX sample date to closest ST01 sample date

dt1 <- setDT(FFX_Cl)
dt2 <- setDT(ST01_Cl)

setkey(dt2, ID, DATE)[, datematch:= DATE]

DATA_filtered_Cl = dt2[dt1, roll = 'nearest'] #wrap longer dataset around
shorter dataset
setDF(DATA_filtered_Cl)

View(DATA_filtered)

####
#find samples within one date of each other

#subtract to find difference
DATA_filtered_Cl$difference = DATA_filtered_Cl$DATE -
DATA_filtered_Cl$datematch

#sort matrix by difference
DATA_filtered_Cl[order(DATA_filtered_Cl[,6],decreasing = FALSE),]

#1 day = 86400 seconds
#if difference = 86400, FFX water sample was taken one day after the ST01
samples
#if difference = 0, samples taken on same day

#Step 2: linear regression of STO1 and FFX water data with 1 day difference

#select data from DATA_filtered
ST01_regress_Cl = DATA_filtered_Cl[DATA_filtered_Cl$difference == 86400, 3]

FFX_regress_Cl = DATA_filtered_Cl[DATA_filtered_Cl$difference == 86400, 5]
```

```
#Linear regression between ST01_regress & FFX_regress

#visualize
plot_Cl =
  plot(ST01_regress_Cl, FFX_regress_Cl,
       col = "blue", pch = 19,
       xlab = 'Chloride (mg/L) at ST01',
       ylab = 'Chloride (mg/L) at FFX Water' )

#interested in how the  concentration of chloride in drinking water (FFX)
(dependent variable, y-axis)
#relates to the concentration at ST01 (independent variable, x-axis)

#fit a line to the relationship between FFX and ST01 using Nelder-Mead
Simplex Algorithem

#create function to minimize

#SE - sum of residuals squared
#BETA[1] - slope
#BETA[2] - y intercept

SE = function (BETA) {sum((FFX_regress_Cl -
                             (BETA[1]*ST01_regress_Cl + BETA[2]))^2)}

#find B1 and B0 that minimize function using the Nelder-Mead Simplex
Algorithm

library(pracma)

STRUCT_Cl = fminsearch(SE, c(100, 0.5))

B_Cl = STRUCT_Cl$xmin

#plot best fit line

Y_Cl = B_Cl[1]*(ST01_regress_Cl) + B_Cl[2]

lines(ST01_regress_Cl, Y_Cl, col = "black")

#B values
#Beta[1] = 1.059317 (slope)
#Beta[2] = 8.477419 (intercept)
#trendline: y = 1.059317x+8.477419

#find R^2 value
R_Cl = cor(cbind(ST01_regress_Cl, FFX_regress_Cl))
RR_Cl = R^2
RR_Cl
#RR is 0.9021103

text(30, 140, expression(y == 1.059*x + 8.477)) #add trendline to plot
```

```
text(24, 135, expression(R^2 == 0.902)) #add R^2 to plot


#create matrix of ST01 sample date, ST01 data (x values), interpolated FFX
data (y values)
INTERP_Cl = cbind.data.frame(ST01_Cl$DATE, ST01_Cl$ST01_CL,
ST01_Cl$ST01_CL*1.059317+8.477419)

#export to Excel
library(openxlsx)

write.xlsx(DATA_filtered, "Desktop")
write.xlsx(INTERP, "Desktop")

####PART 2
#Recognized ST01 had several sample events per month, whereas FFX sampled
monthly
#In a Juptyer Notebook, wrote code to randomly select ST01 sample event per
month during date gaps (July 2009 - August 2013)
#Would be more seamless to do this all in R, but ¯\_(ツ)_/¯
#Upload results

FFX_interp_Cl <- read_excel("~/Documents/Virginia Tech/Research/Sodium
Ingestion/Python_Output_2021.01.22.xlsx")

par(family = "serif")

plot5 =
  plot(FFX_Cl$DATE, FFX_Cl$FFX_CL,
       col = "blue",
       pch = 19,
       bty = "7",
       xlab = 'Sample Date',
       ylab = 'Chloride (mg/L)' )

points(FFX_interp_Cl$Date, FFX_interp_Cl$Conc_Interp,
       col = "orange", pch = 19)

legend("topleft",
       legend = c("Raw Data", "Interpolated Data"),
       col = c("blue","orange"),
       pch = c(16, 16),
       cex = 0.9,
       bty = 'n'
)

###############################################3
#combine plots & make pretty for thesis

#pick colors
library("RColorBrewer")
brewer.pal.info
display.brewer.pal(9, 'Blues')
```

```
brewer.pal(n = 9, 'Blues') #use to find CSS codes for 'Blues' color palette

display.brewer.pal(9, 'Oranges')
brewer.pal(n = 9, 'Oranges') #use to find CSS codes for 'Oranges' color
palette

############################
#linear regressions of sodium & chloride between FFX Water & ST01
par(mfrow=c(1,2)) #nrows, ncols

plot_Na =
  plot(ST01_regress_Na, FFX_regress_Na,
       col = "#08519C", pch = 1, cex = 1.25,
       ylim = c(0, 100),
       xlim = c(0, 80),
       bty = "7",
       xlab = 'Sodium (mg/L) at ST01',
       ylab = 'Sodium (mg/L) at FFX Water' )

  lines(ST01_regress_Na, Y_Na, col = "black")
  text(19, 90, expression(y == 0.9469*x + 1.9540)) #add trendline to plot
  text(19, 84, expression(R^2 == 0.741)) #add R^2 to plot
  text(-20, 110, 'a)', xpd = NA)

plot_Cl =
  plot(ST01_regress_Cl, FFX_regress_Cl,
       col = "#08519C", pch = 1, cex = 1.25,
       ylim = c(0, 150),
       xlim = c(0, 140),
       bty = "7",
       xlab = 'Chloride (mg/L) at ST01',
       ylab = 'Chloride (mg/L) at FFX Water' )

  lines(ST01_regress_Cl, Y_Cl, col = "black")

  text(33, 140, expression(y == 1.059*x + 8.477)) #add trendline to plot
  text(33, 130, expression(R^2 == 0.902)) #add R^2 to plot
  text(-35, 165, 'b)', xpd = NA)

############################################################
#interpolations
par(mfrow=c(1,2)) #nrows, ncols

plot2 =
  plot(FFX_Na$DATE, FFX_Na$NA_FFX,
       col = "#08519C",
       pch = 19, cex = 0.7,
       bty = "7",
       xlab = 'Sample Date',
       ylab = 'Sodium Concentration (mg/L)' )

  points(FFX_interp_Na$`Sample Date`, FFX_interp_Na$Sodium,
       col = "#FD8D3C", pch = 19, cex = 0.7)
```

```
  text(-5, 93, 'a)', xpd = NA)

  legend("topleft",
       legend = c("Raw Data", "Interpolated Data"),
       col = c("#08519C","#FD8D3C"),
       pch = c(16, 16),
       cex = 0.9,
       bty = 'n'
)

plot5 =
  plot(FFX_Cl$DATE, FFX_Cl$FFX_CL,
       col = "#08519C",
       pch = 19, cex = 0.7,
       bty = "7",
       xlab = 'Sample Date',
       ylab = 'Chloride Concentration (mg/L)' )

  points(FFX_interp_Cl$Date, FFX_interp_Cl$Conc_Interp,
       col = "#FD8D3C", pch = 19, cex = 0.7)

  text(-5, 170, 'b)', xpd = NA)

  legend("topleft",
       legend = c("Raw Data", "Interpolated Data"),
       col = c("#08519C","#FD8D3C"),
       pch = c(16, 16),
       cex = 0.9,
       bty = 'n'
)
#####################################################################
###pretty plots for mu and sigma regressions

#mu
#visualize only decades with data (not projections)
plot_mu =
    plot(DECADE[1:4], MU_LN[1:4],
         cex.lab = 1, cex.axis = 1, cex = 1.25,
         col = '#08519C', pch = 19,
         bty = "7",
         xaxt = 'n', #turn off automatic x axis
         ylim = c(0, 4),
         xlim = c(1, 4),
         xlab = 'Decade',
         ylab = expression(μ[LnNa]))

axis(1, at = DECADE[1:4],  labels = DECADE_names[1:4], pos = 0, cex.axis = 1)
#new and improved x axis

#interested in how mu_LnNa (dependent variable, y-axis)
#relates to the time (independent variable, x-axis)

#fit a line to the relationship between mu_LnNa and decades using Nelder-Mead
Simplex Algorithem
```

```
#create function to minimize

#SE - sum of residuals squared
#BETA[1] - slope
#BETA[2] - y intercept

SE = function (BETA) {sum((MU_LN[1:4] -
                                (BETA[1]*DECADE[1:4] + BETA[2]))^2)}

#find B1 and B0 that minimize function using the Nelder-Mead Simplex
Algorithm

library(pracma)

  STRUCT_muLN = fminsearch(SE, c(100, 0.5))
  B_muLN = STRUCT_muLN$xmin

  #plot best fit line

  Y_muLN = B_muLN[1]*(DECADE[1:4]) + B_muLN[2]

  lines(DECADE[1:4], Y_muLN, col = "black")

  #B values
  #Beta[1] = 0.41 (slope)
  #Beta[2] = 1.69 (intercept)
  #trendline: y = 0.41x+1.69

  #find R^2 value
  R_muLN = cor(cbind(DECADE[1:4], MU_LN[1:4]))
  RR_muLN = R_muLN^2
  RR_muLN
  #RR is 0.99935

  text(3, 1, expression(y == 0.41*x + 1.69)) #add trendline to plot
  text(3, 0.75, expression(R^2 == 0.9935)) #add R^2 to plot
  text(0.2, 4.5, 'a)', xpd = NA)

#sigma
#visualize only decades with data (not projections)
plot_sigma =
  plot(DECADE[1:4], SIGMA_LN[1:4],
       cex.lab = 1, cex.axis = 1, cex = 1.25,
       col = "#08519C", pch = 19,
       bty = "7",
       xaxt = 'n', #turn off automatic x axis
       ylim = c(0, 0.6),
       xlim = c(1, 4),
       xlab = 'Decade',
       ylab = expression(σ[LnNa]) )

axis(1, at = DECADE[1:4],  labels = DECADE_names[1:4], pos = 0, cex.axis = 1)
#new and improved x axis
```

```
#interested in how mu_LnNa (dependent variable, y-axis)
#relates to the time (independent variable, x-axis)

#fit a line to the relationship between mu_LnNa and decades using Nelder-Mead
Simplex Algorithem

#create function to minimize

#SE - sum of residuals squared
#BETA[1] - slope
#BETA[2] - y intercept

  SE = function (BETA) {sum((SIGMA_LN[1:4] -
                               (BETA[1]*DECADE[1:4] + BETA[2]))^2)}

  #find B1 and B0 that minimize function using the Nelder-Mead Simplex
Algorithm

  library(pracma)

  STRUCT_sigmaLN = fminsearch(SE, c(100, 0.5))

  B_sigmaLN = STRUCT_sigmaLN$xmin

  #plot best fit line

  Y_sigmaLN = B_sigmaLN[1]*(DECADE[1:4]) + B_sigmaLN[2]

  lines(DECADE[1:4], Y_sigmaLN, col = "black")

  #B values
  #Beta[1] = -0.0417 (slope)
  #Beta[2] = 0.5557 (intercept)
  #trendline: y = -0.0417x+0.5557

  #find R^2 value
  R_sigmaLN = cor(cbind(DECADE[1:4], SIGMA_LN[1:4]))
  RR_sigmaLN = R_sigmaLN^2
  RR_sigmaLN
  #RR is 0.8767

  text(3, 0.15, expression(y == -0.0417*x + 0.557)) #add trendline to plot
  text(3, 0.1, expression(R^2 == 0.8767)) #add R^2 to plot
  text(0.2, 0.7, 'b)', xpd = NA)
```

Appendix C

Code for Random Date Selection (Python)

```
#randomly selecting sample of sodium concentration per month from list of data
#dataset: [date1, concentration1]

import pandas as pd
import random
from datetime import date
from calendar import monthrange

from google.colab import files


#upload Excel spreadsheet through files tab on lefthand side
#copy file path

df = pd.read_excel('/content/Python_Upload_2023.22.01.xlsx')
print(df) #print to check upload

df_list = df.values.tolist()
print(df_list) #print to check

#reconfigure date format
date_set = set()
for item in df_list:
    year, month, day = item[0].strftime('%Y-%m-%d').split('-')
    #print(f'year: {year}')
    #print(f'month: {month}')

    if int(month) != 10: #if month does not equal 10, to avoid turning 10 to 1
      month = f'{month}'.replace('0', '') #curated string of months to convert 01 to string, and then remove the
    date_set.add(f'{year}-{month}')

date_list = list(date_set) #create list

date_list = sorted(date_list) #sort list
print(date_list)
```

```
           Date  ST01_conc  Interp_conc
0    2009-07-06       17.9    27.439193
1    2009-07-13       18.3    27.862920
2    2009-07-20       18.9    28.498510
3    2009-07-27       19.8    29.451896
4    2009-08-03       20.9    30.617144
..          ...        ...          ...
218  2013-07-22       10.3    19.388384
219  2013-07-28       11.7    20.871428
220  2013-08-01       16.0    25.426491
221  2013-08-12       14.4    23.731584
222  2013-08-23       19.1    28.710374

[223 rows x 3 columns]
[[Timestamp('2009-07-06 00:00:00'), 17.9, 27.4391933], [Timestamp('2009-07-13 00:00:00'), 18.3, 27.8629201]
['2009-10', '2009-11', '2009-12', '2009-7', '2009-8', '2009-9', '2010-1', '2010-10', '2010-11', '2010-12',
```

```
po = []

#for loop to create the start and end date for each month in data series
for date_item in date_list:
    data = []
    year, month = date_item.split('-')
    month = int(month)
```

```
    year = int(year)
    end = monthrange(year, month)[1] #determines how many days in a month
    start_date = date(year, month, 1)
    end_date = date(year, month, end)

#select random date and corresponding concentration for each month
    for item in df_list:
        year, month, day = item[0].strftime('%Y-%m-%d').split('-')
        month = int(month)
        year = int(year)
        day = int(day)
        record_date = date(year, month, day)
        if record_date >= start_date and record_date <= end_date:
            data.append(item)

#append results to one dataset
    chosen_item = random.choice(data)
    po.append(chosen_item)

#generate dataframe from list and write to xlsx
df = pd.DataFrame(po, columns = ['date', 'concentration', 'interp'])

#convert to .xlsx and export
df.to_excel('output.xlsx', index=False, header=False)

files.download('output.xlsx')
```

Appendix D

Code for Distribution Fitting (Mathematica)

CDFs of FFX Finished Drinking Water & ST01 Calcs

Import data of sodium (mg/L) from Excel

In[ ]:=
```
all = Import["/Users/caitlinshipman/Documents/Virginia Tech/Research/Sodium
    Ingestion/Mathematica_Upload.xlsx", {"Dataset", 1}, "HeaderLines" → 1]
```

Out[ ]=

1980_1989	1990_1999	2000_2009	2010_2019
18.0	16.0	14.3009	29.7345
18.0	16.0	20.8602	56.7433
17.0	17.0	31.1814	20.1849
12.0	10.0	14.4938	21.6318
12.0	9.0	17.4841	23.9469
12.0	8.0	23.8504	20.7637
12.0	7.0	22.6929	21.6318
11.0	9.0	7.06638	29.7345
12.0	12.0	17.5805	29.3486
11.0	11.0	16.1336	22.5964
11.0	11.0	20.5708	22.9823
11.0	7.0	23.1752	22.5964
10.0	8.0	24.6221	24.3327
16.0	8.0	35.5221	56.9362
17.0	8.0	17.677	16.8088
10.0	7.0	14.7832	24.2363
13.0	8.0	17.8699	19.4133
10.0	8.0	18.3522	19.5097
13.0	8.0	16.7124	25.5867
13.0	8.0	17.4841	32.3389

rows 1–20 of **793**

Take data from table and place into arrays by decade, removing blanks

In[ ]:=
```
firstdecade = Normal[all[[All, "1980_1989"]]]
```

Out[ ]=
```
{18., 18., 17., 12., 12., 12., 12., 11., 12., 11., 11., 11., 10., 16., 17., 10., 13.,
 10., 13., 13., 13., 11., 11., 12., 13., 13., 12., 9., 8., 9., 7., 7., 7., 6., 7.,
 6., 7., 6., 6., 3., 3., 4., 4., 4., 8., 8., 8., 11., 11., 12., 14., 12., 8., 8.,
 8., 8., 6., 6., 6., 4., 4., 2., 2., 2., 3., 3., 3., 4., 4., 4., 6., 6., 6., 7.,
 8., 5., 5., 5., 6., 6., 6., 5., 5., 5., 6., 6., 6., 4., 3., 5., 4., 5., 5., 5.,
 5., 5., 2., 2., 2., 5., 5., 5., 6., 6., 5., 5., 5., 5., 4., 4., 4., 8., 8., 9.,
 5., 5., 5., 7., 7., 7., 9., 9., 9., 11., 11., 12., 12., 13., 13., 10., 10., 10.,
 13., 12., 12., 17., 18., 18., 16., 21., 12., 6., 6., 6., 4., 4., 8., 8., 12., 12.,
 12., 11., 11., 11., 9., 9., 9., 11., 10., 10., 13., 13., 13., 14., 15., 15., 15.,
 15., 16., 16., 16., 18., 17., 17., 10., 10., 11., 10., 10., 10., 8., 7., 7., 6.,
 5., 4., 5., 5., 4., 4., 4., 8., 8., 9., 10., 10., 10., 10., 11., 10., 10., 10.,
 10., 10., 10., 14., 13., 13., 8., 8., 8., 8., 8., 10., 10., 9., 10., 12., 12.,
 8., 8., 5., 5., 5., 7., 7., 8., 4., 4., 13., 11., 15., 15., 15., 19., 18., 19.,
 19., 23., 18., 19., 17., 15., 16., 15., 10., 9., 9., 12., 11., 11., 5., 6., 6.,
 7., 8., 7., 7., 12., 11., 12., 13., 14., 13., 9., 8., 8., 9., 8., 9., , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , }
```

```
In[•]:= firstdecade = {18.`, 18.`, 17.`, 12.`, 12.`, 12.`, 12.`, 11.`, 12.`, 11.`, 11.`, 11.`,
    10.`, 16.`, 17.`, 10.`, 13.`, 10.`, 13.`, 13.`, 13.`, 11.`, 11.`, 12.`, 13.`,
    13.`, 12.`, 9.`, 8.`, 9.`, 7.`, 7.`, 7.`, 6.`, 7.`, 6.`, 7.`, 6.`, 6.`, 3.`,
    3.`, 4.`, 4.`, 4.`, 8.`, 8.`, 8.`, 11.`, 11.`, 12.`, 14.`, 12.`, 8.`, 8.`,
    8.`, 8.`, 6.`, 6.`, 6.`, 4.`, 4.`, 2.`, 2.`, 2.`, 3.`, 3.`, 3.`, 4.`, 4.`, 4.`,
    6.`, 6.`, 6.`, 7.`, 8.`, 5.`, 5.`, 5.`, 6.`, 6.`, 6.`, 5.`, 5.`, 5.`, 6.`, 6.`,
    6.`, 4.`, 3.`, 5.`, 4.`, 5.`, 5.`, 5.`, 5.`, 5.`, 2.`, 2.`, 2.`, 5.`, 5.`, 5.`,
    6.`, 6.`, 5.`, 5.`, 5.`, 5.`, 4.`, 4.`, 4.`, 8.`, 8.`, 9.`, 5.`, 5.`, 5.`, 7.`,
    7.`, 7.`, 9.`, 9.`, 9.`, 11.`, 11.`, 12.`, 12.`, 13.`, 13.`, 10.`, 10.`, 10.`,
    13.`, 12.`, 12.`, 17.`, 18.`, 18.`, 16.`, 21.`, 12.`, 6.`, 6.`, 6.`, 4.`,
    4.`, 8.`, 8.`, 12.`, 12.`, 12.`, 11.`, 11.`, 11.`, 9.`, 9.`, 9.`, 11.`, 10.`,
    10.`, 13.`, 13.`, 13.`, 14.`, 15.`, 15.`, 15.`, 15.`, 16.`, 16.`, 16.`, 18.`,
    17.`, 17.`, 10.`, 10.`, 11.`, 10.`, 10.`, 10.`, 8.`, 7.`, 7.`, 6.`, 5.`, 4.`,
    5.`, 5.`, 4.`, 4.`, 4.`, 8.`, 8.`, 9.`, 10.`, 10.`, 10.`, 10.`, 11.`, 10.`,
    10.`, 10.`, 10.`, 10.`, 10.`, 14.`, 13.`, 13.`, 8.`, 8.`, 8.`, 8.`, 8.`, 10.`,
    10.`, 9.`, 10.`, 12.`, 12.`, 8.`, 8.`, 5.`, 5.`, 5.`, 7.`, 7.`, 8.`, 4.`, 4.`,
    13.`, 11.`, 15.`, 15.`, 15.`, 19.`, 18.`, 19.`, 19.`, 23.`, 18.`, 19.`, 17.`,
    15.`, 16.`, 15.`, 10.`, 9.`, 9.`, 12.`, 11.`, 11.`, 5.`, 6.`, 6.`, 7.`, 8.`,
    7.`, 7.`, 12.`, 11.`, 12.`, 13.`, 14.`, 13.`, 9.`, 8.`, 8.`, 9.`, 8.`, 9.`};
```

```
In[]:= seconddecade = Normal[all[[All, "1990_1999"]]]

Out[]= {16., 16., 17., 10., 9., 8., 7., 9., 12., 11., 11., 7., 8., 8., 8., 7., 8., 8., 8., 8.,
  11., 10., 10., 16., 15., 14., 7., 8., 8., 7., 7., 7., 8., 7., 8., 7., 11., 10., 7.,
  7., 7., 9., 9., 9., 10., 10., 10., 13., 13., 12., 10., 10., 9., 13., 12., 12., 7.,
  8., 19., 19., 12., 12., 5., 5., 8., 9., 8., 7., 7., 7., 13., 13., 12., 13., 12., 11.,
  11., 11., 14., 15., 15., 15., 14., 15., 7., 7., 7., 3., 6., 16., 11., 11., 11., 12.,
  11., 12., 11., 11., 11., 18., 19., 18., 18., 18., 18., 13., 13., 13., 10., 11.,
  11., 12., 11., 11., 9., 9., 9., 9., 9., 10., 8., 8., 9., 4., 3., 6., 6., 6., 5., 6.,
  6., 8., 8., 7., 14., 14., 14., 12., 12., 12., 12., 12., 14., 14., 14., 14., 14.,
  15., 16., 16., 16., 18., 18., 18., 25., 24., 25., 25., 25., 25., 20., 19., 20.,
  6., 6., 6., 7., 7., 7., 7., 6., 6., 11., 11., 11., 8., 7., 8., 10., 10., 10., 22.,
  22., 22., 25., 25., 26., 26., 26., 30., 30., 31., 25., 26., 25., 32., 32., 32.,
  28., 28., 28., 30., 29., 30., 31., 30., 31., 33., 31., 33., 11., 10., 18., 17., ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
  , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , }

In[]:= seconddecade = {16.`, 16.`, 17.`, 10.`, 9.`, 8.`, 7.`, 9.`, 12.`, 11.`, 11.`, 7.`,
    8.`, 8.`, 8.`, 7.`, 8.`, 8.`, 8.`, 8.`, 11.`, 10.`, 10.`, 16.`, 15.`, 14.`,
    7.`, 8.`, 8.`, 7.`, 7.`, 7.`, 8.`, 7.`, 8.`, 7.`, 11.`, 10.`, 7.`, 7.`, 7.`,
    9.`, 9.`, 9.`, 10.`, 10.`, 10.`, 13.`, 13.`, 12.`, 10.`, 10.`, 9.`, 13.`,
    12.`, 12.`, 7.`, 8.`, 19.`, 19.`, 12.`, 12.`, 5.`, 5.`, 8.`, 9.`, 8.`, 7.`,
    7.`, 7.`, 13.`, 13.`, 12.`, 13.`, 12.`, 11.`, 11.`, 11.`, 14.`, 15.`, 15.`,
    15.`, 14.`, 15.`, 7.`, 7.`, 7.`, 3.`, 6.`, 16.`, 11.`, 11.`, 11.`, 12.`,
    11.`, 12.`, 11.`, 11.`, 11.`, 18.`, 19.`, 18.`, 18.`, 18.`, 18.`, 13.`, 13.`,
    13.`, 10.`, 11.`, 11.`, 12.`, 11.`, 11.`, 9.`, 9.`, 9.`, 9.`, 9.`, 10.`, 8.`,
    8.`, 9.`, 4.`, 3.`, 6.`, 6.`, 6.`, 5.`, 6.`, 6.`, 8.`, 8.`, 7.`, 14.`, 14.`,
    14.`, 12.`, 12.`, 12.`, 12.`, 12.`, 14.`, 14.`, 14.`, 14.`, 14.`, 15.`, 16.`,
    16.`, 16.`, 18.`, 18.`, 18.`, 25.`, 24.`, 25.`, 25.`, 25.`, 25.`, 20.`, 19.`,
    20.`, 6.`, 6.`, 6.`, 7.`, 7.`, 7.`, 7.`, 6.`, 6.`, 11.`, 11.`, 11.`, 8.`,
    7.`, 8.`, 10.`, 10.`, 10.`, 22.`, 22.`, 22.`, 25.`, 25.`, 26.`, 26.`, 26.`,
    30.`, 30.`, 31.`, 25.`, 26.`, 25.`, 32.`, 32.`, 32.`, 28.`, 28.`, 28.`, 30.`,
    29.`, 30.`, 31.`, 30.`, 31.`, 33.`, 31.`, 33.`, 11.`, 10.`, 18.`, 17.`};

In[]:= thirddecade = Normal[all[[All, "2000_2009"]]]
```

```
{14.30088`, 20.86016`, 31.18138`, 14.4938`, 17.48406`, 23.85042`, 22.6929`,
 7.06638`, 17.58052`, 16.13362`, 20.57078`, 23.1752`, 24.6221`, 35.52208`,
 17.67698`, 14.78318`, 17.8699`, 18.3522`, 16.71238`, 17.48406`, 20.08848`,
 22.2106`, 29.54156`, 29.05926`, 30.21678`, 31.18138`, 30.7`, 28.5`, 29.1`,
 35.42562`, 34.6`, 31.8`, 33.6`, 37.45128`, 39.`, 38.8`, 42.4`, 24.81502`,
 24.`, 24.7`, 26.`, 23.7`, 21.4`, 24.4`, 16.9053`, 18.73804`, 21.82476`, 20.`,
 19.4`, 19.2`, 21.9`, 21.6`, 20.9`, 27.0336`, 33.1`, 32.8`, 32.8`, 27.41944`,
 29.63802`, 10.3`, 9.2`, 10.`, 25.49024`, 25.7`, 24.5`, 25.5`, 21.`, 19.`,
 20.6`, 15.2`, 14.2`, 15.1`, 9.8`, 8.7`, 10.`, 14.7`, 14.`, 15.`, 8.1`, 6.7`,
 8.4`, 10.3`, 9.8`, 10.3`, 10.2`, 9.7`, 10.1`, 37.1`, 34.9`, 37.4`, 19.4`,
 18.4`, 19.4`, 12.8`, 10.8`, 12.6`, 14.1`, 14.`, 14.3`, 18.5`, 18.5`, 18.4`,
 20.1`, 18.6`, 20.1`, 20.2`, 19.6`, 20.2`, 14.1`, 12.9`, 13.3`, 16.2`, 13.2`,
 16.`, 12.4`, 11.4`, 12.2`, 27.1`, 25.6`, 26.6`, 12.1`, 11.1`, 12.1`, 17.7`,
 17.7`, 17.7`, 15.3`, 14.`, 15.`, 18.7`, 17.7`, 18.8`, 21.9`, 20.9`, 22.`,
 10.9`, 8.2`, 9.7`, 18.`, 17.`, 18.`, 21.8`, 21.1`, 21.9`, 16.2`, 14.9`, 16.2`,
 25.9`, 24.8`, 26.2`, 28.4`, 29.9`, 18.7`, 28.5`, 17.4`, 24.8`, 15.3`, 24.6`,
 14.6`, 22.1`, 12.8`, 20.2`, 11.5`, 30.`, 19.5`, 28.8`, 18.6`, 22.9`, 12.6`,
 26.2`, 16.3`, 36.8`, 24.7`, 44.1`, 32.6`, 44.7`, 33.9`, 53.4`, 47.6`, 42.2`,
 37.2`, 19.6`, 21.2`, 27.9`, 31.7`, 18.3`, 28.6`, 27.9`, 29.`, 25.3`, 38.2`,
 48.9`, 22.`, 14.87964`, 20.18494`, 25.77962`, 31.18138`, 19.22034`, 14.9761`}
```

In[•]:=
```
thirddecade = {14.30088`, 20.86016`, 31.18138`, 14.4938`, 17.48406`, 23.85042`,
   22.6929`, 7.06638`, 17.58052`, 16.13362`, 20.57078`, 23.1752`, 24.6221`,
   35.52208`, 17.67698`, 14.78318`, 17.8699`, 18.3522`, 16.71238`, 17.48406`,
   20.08848`, 22.2106`, 29.54156`, 29.05926`, 30.21678`, 31.18138`, 30.7`, 28.5`,
   29.1`, 35.42562`, 34.6`, 31.8`, 33.6`, 37.45128`, 39.`, 38.8`, 42.4`, 24.81502`,
   24.`, 24.7`, 26.`, 23.7`, 21.4`, 24.4`, 16.9053`, 18.73804`, 21.82476`, 20.`,
   19.4`, 19.2`, 21.9`, 21.6`, 20.9`, 27.0336`, 33.1`, 32.8`, 32.8`, 27.41944`,
   29.63802`, 10.3`, 9.2`, 10.`, 25.49024`, 25.7`, 24.5`, 25.5`, 21.`, 19.`,
   20.6`, 15.2`, 14.2`, 15.1`, 9.8`, 8.7`, 10.`, 14.7`, 14.`, 15.`, 8.1`, 6.7`,
   8.4`, 10.3`, 9.8`, 10.3`, 10.2`, 9.7`, 10.1`, 37.1`, 34.9`, 37.4`, 19.4`, 18.4`,
   19.4`, 12.8`, 10.8`, 12.6`, 14.1`, 14.`, 14.3`, 18.5`, 18.5`, 18.4`, 20.1`,
   18.6`, 20.1`, 20.2`, 19.6`, 20.2`, 14.1`, 12.9`, 13.3`, 16.2`, 13.2`, 16.`,
   12.4`, 11.4`, 12.2`, 27.1`, 25.6`, 26.6`, 12.1`, 11.1`, 12.1`, 17.7`, 17.7`,
   17.7`, 15.3`, 14.`, 15.`, 18.7`, 17.7`, 18.8`, 21.9`, 20.9`, 22.`, 10.9`,
   8.2`, 9.7`, 18.`, 17.`, 18.`, 21.8`, 21.1`, 21.9`, 16.2`, 14.9`, 16.2`, 25.9`,
   24.8`, 26.2`, 28.4`, 29.9`, 18.7`, 28.5`, 17.4`, 24.8`, 15.3`, 24.6`, 14.6`,
   22.1`, 12.8`, 20.2`, 11.5`, 30.`, 19.5`, 28.8`, 18.6`, 22.9`, 12.6`, 26.2`,
   16.3`, 36.8`, 24.7`, 44.1`, 32.6`, 44.7`, 33.9`, 53.4`, 47.6`, 42.2`, 37.2`,
   19.6`, 21.2`, 27.9`, 31.7`, 18.3`, 28.6`, 27.9`, 29.`, 25.3`, 38.2`, 48.9`,
   22.`, 14.87964`, 20.18494`, 25.77962`, 31.18138`, 19.22034`, 14.9761`};
```

```
In[ ]:= fourthdecade = Normal[all[[All, "2010_2019"]]]

Out[ ]= {29.7345, 56.7433, 20.1849, 21.6318, 23.9469, 20.7637, 21.6318, 29.7345, 29.3486,
 22.5964, 22.9823, 22.5964, 24.3327, 56.9362, 16.8088, 24.2363, 19.4133, 19.5097,
 25.5867, 32.3389, 10.2881, 12.4681, 16.423, 14.6867, 14.2044, 25.8761, 26.9371,
 21.7283, 25.2973, 12.9504, 14.3009, 24.1398, 26.2619, 17.2911, 8.90877, 15.3619,
 19.1239, 12.854, 19.4133, 30.6991, 17.677, 18.931, 11.3106, 26.6, 32.1, 23.,
 46.2, 52.6, 28.8, 26.7, 30., 36.4, 36.6, 37.8, 83.8, 47.8, 37.4, 40.7, 29.2,
 33.6, 37.2, 29.7, 32.2, 47.8, 33.6, 37.1, 35.4, 31.3, 35.5, 37.9, 40.7, 39.2,
 38.6, 40.8, 35.7, 34.7, 38.3, 36.8, 31.1, 28.7, 27.8, 35., 33.9, 36.5, 38.5,
 37.6, 45.1, 31.2, 25.9, 25.6, 19.9, 22.7, 16.9, 27.6, 16.2, 32.5, 32.3, 24.4,
 20.2, 23., 27.2, 29.8, 35.7, 37.4, 33.3, , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , ,
 , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , , }

In[ ]:= fourthdecade = {29.73448`, 56.74328`, 20.18494`, 21.63184`, 23.94688`, 20.7637`,
   21.63184`, 29.73448`, 29.34864`, 22.59644`, 22.98228`, 22.59644`, 24.33272`,
   56.9362`, 16.80884`, 24.23626`, 19.41326`, 19.50972`, 25.5867`, 32.3389`,
   10.288144`, 12.46814`, 16.423`, 14.68672`, 14.20442`, 25.87608`, 26.93714`,
   21.7283`, 25.29732`, 12.95044`, 14.30088`, 24.1398`, 26.26192`, 17.29114`,
   8.908766`, 15.36194`, 19.12388`, 12.85398`, 19.41326`, 30.69908`, 17.67698`,
   18.93096`, 11.31062`, 26.6`, 32.1`, 23.`, 46.2`, 52.6`, 28.8`, 26.7`,
   30.`, 36.4`, 36.6`, 37.8`, 83.8`, 47.8`, 37.4`, 40.7`, 29.2`, 33.6`, 37.2`,
   29.7`, 32.2`, 47.8`, 33.6`, 37.1`, 35.4`, 31.3`, 35.5`, 37.9`, 40.7`, 39.2`,
   38.6`, 40.8`, 35.7`, 34.7`, 38.3`, 36.8`, 31.1`, 28.7`, 27.8`, 35.`, 33.9`,
   36.5`, 38.5`, 37.6`, 45.1`, 31.2`, 25.9`, 25.6`, 19.9`, 22.7`, 16.9`, 27.6`,
   16.2`, 32.5`, 32.3`, 24.4`, 20.2`, 23.`, 27.2`, 29.8`, 35.7`, 37.4`, 33.3`};
```

Log (ln) transform each decade

In[ ]:=
```
lnfirstdecade = Log[firstdecade];
lnseconddecade = Log[seconddecade];
lnthirddecade = Log[thirddecade];
lnfourthdecade = Log[fourthdecade];
```

Find Distribution Parameters for Each Decade

1980 - 1989, First Decade

In[ ]:=
```
DistLnFirstDecade = FindDistribution[lnfirstdecade, 5, All]
```

Out[ ]=

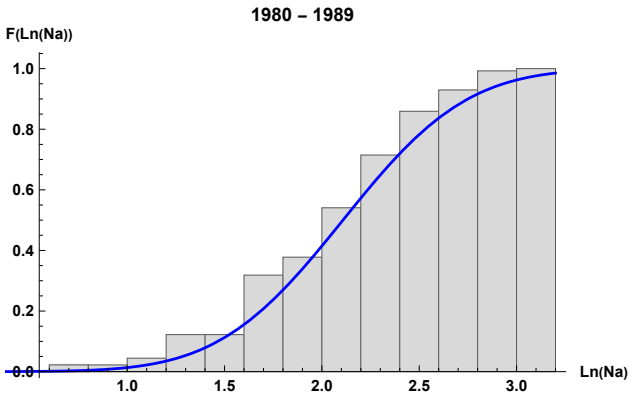
	PearsonChiSq	CramerVonMis	BIC	AIC	HQIC	LogLikelihood	
MixtureDis	0.737263	0.995961	−1.46489	−1.41477	−1.45554	−0.676734	{
WeibullDis	0.224188	0.453789	−1.45553	−1.44265	−1.45319	−0.713835	{
NormalDist	0.0330203	0.16131	−1.50134	−1.48847	−1.49901	−0.736744	{
MixtureDis	0.103543	0.225232	−1.49308	−1.44296	−1.48373	−0.690828	{
StudentTDi	0.0318189	0.188815	−1.51535	−1.49612	−1.51184	−0.736783	{

In[ ]:=
```
FindDistributionParameters[lnfirstdecade, NormalDistribution[μ, σ]]
```

Out[ ]=
```
{μ → 2.10817, σ → 0.501145}
```

In[ ]:=
```
Show[Histogram[lnfirstdecade, Automatic, "CDF",
  AxesLabel → {"Ln(Na)", "F(Ln(Na))"}, ChartStyle → LightGray],
 Plot[CDF[NormalDistribution[2.1081734838925383, 0.5011454143217744], x],
  {x, 0, 3.2}, PlotStyle → Blue],
 PlotLabel → "1980 - 1989", LabelStyle → Directive[Black, Bold]]
```

Out[ ]=



1990 - 1999, Second Decade

In[ ]:=
```
DistLogSecondDecade = FindDistribution[lnseconddecade, 5, All]
```

Out[ ]=

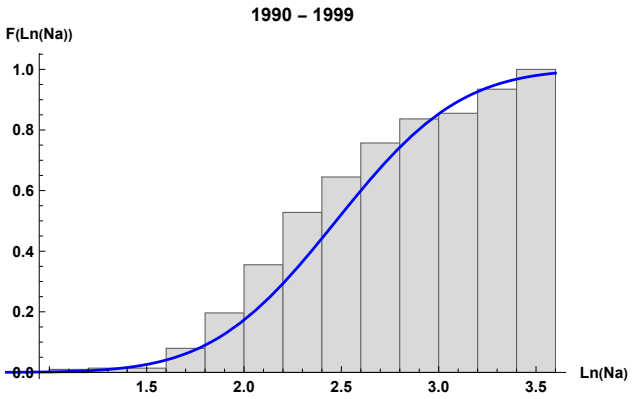
	PearsonChiSq	CramerVonMis	BIC	AIC	HQIC	LogLikelihood	
MixtureDis	0.562533	0.943326	−1.47646	−1.42262	−1.47019	−0.672286	{
NormalDist	0.00135855	0.220418	−1.52188	−1.50786	−1.52031	−0.744452	{
GammaDistr	0.0864557	0.679288	−1.51718	−1.50317	−1.51561	−0.742104	{
LogNormalD	0.0160428	0.759528	−1.53963	−1.52562	−1.53806	−0.753329	{
StudentTDi	0.00135855	0.22138	−1.53857	−1.51768	−1.53622	−0.744556	{

In[ ]:=
```
FindDistributionParameters[lnseconddecade, NormalDistribution[μ, σ]]
```

Out[ ]=
```
{μ → 2.473, σ → 0.501565}
```

In[ ]:=
```
Show[Histogram[lnseconddecade, Automatic, "CDF",
  AxesLabel → {"Ln(Na)", "F(Ln(Na))"}, ChartStyle → LightGray],
 Plot[CDF[NormalDistribution[2.472996350467118`, 0.5015645342393554`], x],
  {x, 0, 3.6}, PlotStyle → Blue],
 PlotLabel → "1990 - 1999", LabelStyle → Directive[Black, Bold]]
```

Out[ ]=



2000 - 2009, Third Decade

In[ ]:=
```
DistLogThirdDecade = FindDistribution[lnthirddecade, 5, All]
```

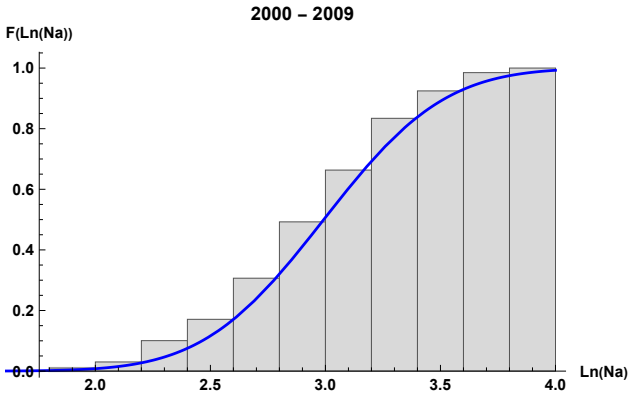
Out[ ]=

	PearsonChiSq	CramerVonMis	BIC	AIC	HQIC	LogLikelihood	
NormalDist	0.991886	0.971572	−1.10953	−1.09521	−1.1083	−0.537401	{
WeibullDis	0.997502	0.998863	−1.11207	−1.09074	−1.11022	−0.529987	{
WeibullDis	0.836214	0.816273	−1.12233	−1.10801	−1.1211	−0.543803	{
GammaDistr	0.405029	0.533306	−1.1335	−1.11918	−1.13227	−0.549386	{
StudentTDi	0.991886	0.971175	−1.12753	−1.10621	−1.12569	−0.53772	{

```
In[ ]:= FindDistributionParameters[lnthirddecade, NormalDistribution[μ, σ]]
Out[ ]= {μ → 2.9926, σ → 0.412277}

In[ ]:= Show[Histogram[lnthirddecade, Automatic, "CDF",
          AxesLabel → {"Ln(Na)", "F(Ln(Na))"}, ChartStyle → LightGray],
         Plot[CDF[NormalDistribution[2.992598032968942`, 0.4122772663893793`], x],
          {x, 0, 4}, PlotStyle → Blue],
         PlotLabel → "2000 - 2009", LabelStyle → Directive[Black, Bold]]
```

Out[ ]=



2010 - 2019, fourth decade

```
In[ ]:= DistLogFourthDecade = FindDistribution[lnfourthdecade, 5, All]
```

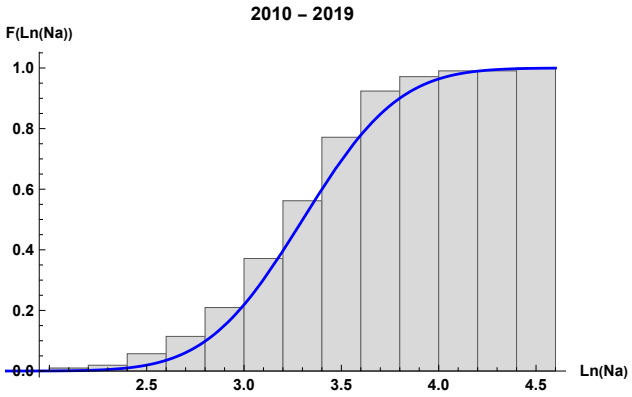
Out[ ]=

	PearsonChiSq	CramerVonMis	BIC	AIC	HQIC	LogLikelihood	
NormalDist	0.301926	0.40341	−1.04256	−1.02813	−1.0475	−0.49446	{
LogisticDi	0.31806	0.635922	−1.02613	−1.01171	−1.03107	−0.486246	{
WeibullDis	0.286351	0.4319	−1.0665	−1.04544	−1.07391	−0.493019	{
StudentTDi	0.256889	0.635941	−1.054	−1.03295	−1.06141	−0.486771	{
WeibullDis	0.125262	0.352057	−1.07208	−1.05765	−1.07702	−0.509218	{

```
In[ ]:= FindDistributionParameters[lnfourthdecade, NormalDistribution[μ, σ]]
Out[ ]= {μ → 3.30171, σ → 0.388495}
```

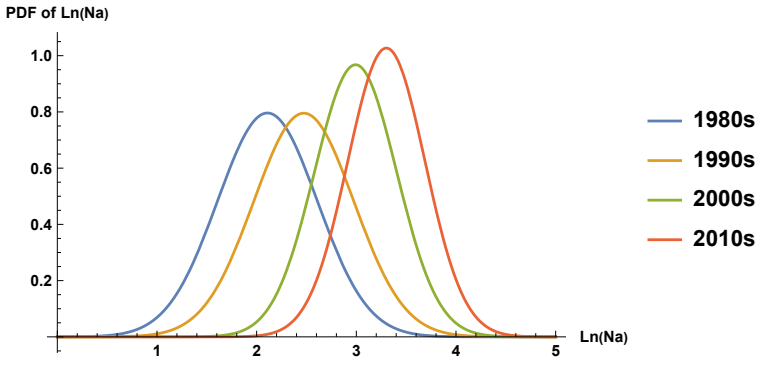
In[ ]:=
```
Show[Histogram[lnfourthdecade, Automatic, "CDF",
  AxesLabel → {"Ln(Na)", "F(Ln(Na))"}, ChartStyle → {LightGray}],
 Plot[CDF[NormalDistribution[3.3017050787526845`, 0.38849457586916114`], x],
  {x, 0, 4.6}, PlotStyle → Blue],
 PlotLabel → "2010 - 2019", LabelStyle → Directive[Black, Bold]]
```

Out[ ]=



In[ ]:=
```
Plot[{PDF[NormalDistribution[2.1081734838925383, 0.5011454143217744], x],
  PDF[NormalDistribution[2.472996350467118`, 0.5015645342393554`], x],
  PDF[NormalDistribution[2.992598032968942`, 0.4122772663893793`], x],
  PDF[NormalDistribution[3.3017050787526845`, 0.38849457586916114`], x]},
 {x, 0, 5},
 AxesLabel → {"Ln(Na)", "PDF of Ln(Na)"},
 LabelStyle → Directive[Black, Bold],
 PlotLegends → {"1980s", "1990s", "2000s", "2010s"}]
```

Out[ ]=



Appendix E

Code for Monte Carlo Simulations (Igor64)

```
Function Salt_Distribution_Gen(TUL, ingest_inv_CDF, output, ln_mu_sodium,
ln_sigma_sodium)
        Variable TUL, ln_mu_sodium, ln_sigma_sodium
        WAVE ingest_inv_CDF, output
        Variable numsims, rnd1, rnd2, i, sodium_inv_CDF

        //determine how many simulations we're going to run
        numsims=numpnts(output)

        //initialize counter
        i=0
        //initialize do-loop

        Do
        //draw the percentile values for ingestion volume (rnd1) and sodium concentration
(rnd2)
        rnd1=abs(enoise(1))
        rnd2=abs(enoise(1))

        sodium_inv_CDF = StatsInvNormalCDF(rnd2, ln_mu_sodium, ln_sigma_sodium)

        //compute the % of daily TUL provided by community drinking water
        output[i]=(ingest_inv_CDF(rnd1)*(e^(sodium_inv_CDF))/(TUL*1000))*100

        //move on to the next realization
        i+=1
        While (i<numsims)
end

/////
```