Biofilm Detection through the use of Factor Analysis and Principal Component Analysis

By

Richard Gallant

A Thesis Submitted to the Faculty of

The College of Engineering and Computer Science

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, FL

August 2019

Copyright 2019 by Richard Gallant

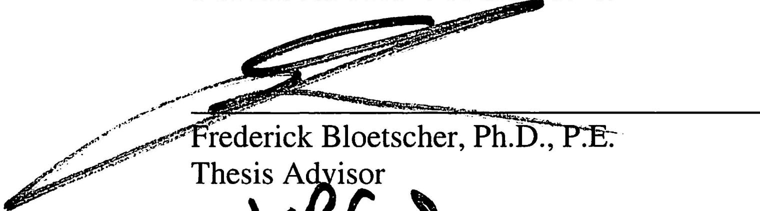
Biofilm Detection through the use of Factor Analysis and Principal Component Analysis

by

Richard Gallant

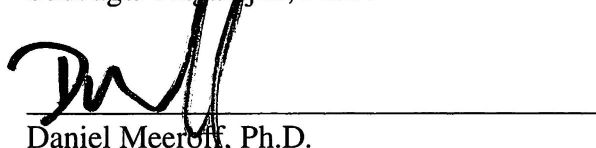
This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Frederick Bloetscher, Department of Engineering and Computer Science, and has been approved by all members of the supervisory committee. It was submitted to the faculty of the College of Engineering and Computer Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:



Frederick Bloetscher, Ph.D., P.E.
Thesis Advisor

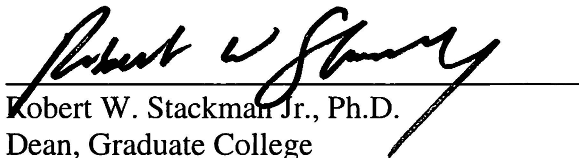
Sudhagar Nagarajan, Ph.D.



Daniel Meeroff, Ph.D.

Yan Yong, Ph.D.
Chair, Department of Civil, Environmental, And Geomatics Engineering

Stella Batalama, Ph.D.
Dean, College of Engineering and Computer Science



Robert W. Stackman Jr., Ph.D.
Dean, Graduate College

July 17, 2019
Date

Acknowledgements

The author wishes to express sincere gratitude to his committee members for all of their guidance and support, and special thanks to my advisor for his persistence, patience, and encouragement during the typing of this manuscript. The author is grateful to the participating water utilities for providing the data and ability to conduct the research needed for this manuscript. Last but not least, the author wishes to thank those persons who aided in the support of this manuscript.

Abstract

Author: Richard Gallant

Title: Biofilm Detection through the use of Factor Analysis and Principal Component Analysis

Institution: Florida Atlantic University

Thesis Advisor: Dr. Frederick Bloetscher

Degree: Master of Science

Year: 2019

Safe drinking water is paramount to a healthy society. Close to a hundred contaminants are regulated by the government. Utilities are using chloramines to disinfect water to reduce harmful byproducts that may present themselves with the use of chlorine alone. Using chlorine and ammonia to disinfect, ammonia oxidizing bacteria can present themselves in an unsuspecting utilities distribution network.

There are precursors that are used to possibly see an issue, but there is no true test to predict these anomalies. Biofilms and nitrification are not really regulated. Their introduction into a distribution system can lead to issues with regulated substances.

This thesis uses factor analysis with principal component analysis to predict the existence of and provide a tool for utilities to detect biofilms. Using sampling data, this analysis was conducted using the XLStat®. Results from this analysis show that the existence of biofilms in potable water distribution systems is possible.

Dedication

This manuscript is dedicated to the loving memory of my mother. Through her entire life she was always there for me. She always told me I could complete any task that I set my sights on. I also dedicate this work to my patient, and loving wife, Kirsten, who has always been my rock and has endured the years of schooling and research to complete this work. I further dedicate this work to my children, Jonathan, Mark, and Jadelyn. They too have missed many moments of our lives through schooling and research, but have also been supportive in this project.

Biofilm Detection through the use of Factor Analysis and Principal Component Analysis

List of Tables ix

List of Figures xi

List of Equations xiv

Introduction 1

Effort to Address Public Health – Regulations 4

How biofilms form 7

Nitrification in the Water Distribution System 12

How Bacteria Get into the Distribution System 18

Detection of Biofilms/Nitrification 19

Biofilm/ Nitrification Control Strategies 23

Hypothesis 29

Methodology 30

Data Collection 30

Statistical Analysis 32

Descriptive Statistics 33

Correlation Matrix 33

Regression Model 34

PCA and FA Analysis 34

Linear Regression 48

Development of Model Data 50

Results and Discussion 52

Discriminant Analysis 73

Linear Regression 75

Data Modification 80

Conclusions: 93

Appendices 95

References 103

List of Tables

Table 1 Summary Statistics for Variables Analyzed 53

Table 2 Correlation among Variables 55

Table 3 Bartlett's Sphericity Test 56

Table 4 Eigenvectors from PCA Analysis 57

Table 5 Factor Loading 57

Table 6 Factor Loading Correlations 58

Table 7 Factor loading percentages for Factors developed from PCA analysis. 65

Table 8 Factor Loadings after rotation.. 66

Table 9 Factor loadings after varimax rotation showing improvement in the explanation of variables as a result of the rotation. 66

Table 10 percent of variability for each factor D1-D12. 66

Table 11 Contribution of the variable to each varimax factor 67

Table 12 Table of Centroids 73

Table 13 Confusion matrix for the estimation sample:.. 75

Table 14 Removed Variables.. 75

Table 15 Goodness of Fit Statistics 76

Table 16 Model Parameters 76

Table 17 Summary Statistics 2013-2014 Data Only... 82

Table 18 Correlation Matrix for 2013-2014 Data... 82

Table 19 Components of Factors from the Scree Plot 83

Table 20 Factor Loading 84

Table 21 Correlations between Variables within Factors 84

Table 22 Contribution Percent of Each Variable to Factor.. 85

Table 23 Correlations between Variables and Factors after Varimax Rotation 85

Table 24 Contribution of Variables in percent after Varimax Rotation 86

Table 25 Centroids for the Plots for Nitrification 87

Table 26 Confusion Matrix for the Estimation Sample 88

Table 27 Standardized Coefficients 89

List of Figures

Figure 1 Scree Plot of PCA Factor Contribution to Variance Among Factors....................... 56

Figure 2 Factor diagram showing grouped correlations among variables constituting F1 and F2.. 59

Figure 3 Factor diagram showing grouped correlations among variables constituting F1 and F3.. 60

Figure 4 Factor diagram showing grouped correlations among variables constituting F1 and F4.. 61

Figure 5 Factor diagram showing grouped correlations among variables constituting F3 and F2.. 62

Figure 6 Factor diagram showing grouped correlations among variables constituting F4 and F2.. 63

Figure 7 Factor diagram showing grouped correlations among variables constituting F3 and F4.. 64

Figure 8 Factor diagram showing grouped correlations among variables constituting D1 and D2. 67

Figure 9 Factor diagram showing grouped correlations among variables constituting D1 and D3 68

Figure 10 Factor diagram showing grouped correlations among variables constituting D1 and D4 68

Figure 11 Factor diagram showing grouped correlations among variables constituting D4 and D2. 69

Figure 12 Factor diagram showing grouped correlations among variables constituting D3 and D2. 69

Figure 13 Factor diagram showing grouped correlations among variables constituting D3 and D4 70

Figure 14 Varimax biplot showing grouped correlations among variables constituting D1 and D2 70

Figure 15 Varimax biplot showing grouped correlations among variables constituting D1 and D3 71

Figure 16 Varimax biplot showing grouped correlations among variables constituting D1 and D4 71

Figure 17 Varimax biplot showing grouped correlations among variables constituting D4 and D2 72

Figure 18 Varimax biplot showing grouped correlations among variables constituting D13 and D2 72

Figure 19 Varimax biplot showing grouped correlations among variables constituting D3 and D4 73

Figure 20 DA analysis showing centroids of the two populations 74

Figure 21 DA showing the centroids and the populations– 74

Figure 22 Nitrification Present 77

Figure 23 Differentiation of nitrification and non-nitrification results. 79

Figure 24 Standardized residuals show separation of t nitrification and non-nitrification samples 79
Figure 25 Prediction of Nitrification 80
Figure 26 Scree Plot Revision - 2013-2014 Data Only 83
Figure 27 Centroids for the Plots for Nitrification 87
Figure 28 Universe for Observations for Nitrification 88
Figure 29 Coefficients for Linear Regression Mode 90
Figure 30 Prediction of Results Comparing Nitrification to Areas Without 90
Figure 31 Prediction of the Model 91
Figure 32 Results of Linear Regression Equation when Applied to 2015-2017 Data 92
Figure 33 System 8 North 95
Figure 34 All Utility Sampling Locations 96
Figure 35 System 3 Sampling Locations 97
Figure 36 System 2 Sampling Locations 97
Figure 37 System 3 South Sampling Locations 98
Figure 38 System 8 Sampling Locations 99
Figure 39 System 8 East Sampling Locations 100
Figure 40 System 8 North Sampling Locations 101
Figure 41 System 9 Sampling Locations 101
Figure 42 System 8 West Sampling Locations 102

List of Equations

Equation 1 Composition of Nitrosomonas (Liu, Taylor, Randall, & Dietz, 2005)........... 13

Equation 2 Composition of Nitrobacter (Liu, Taylor, Randall, & Dietz, 2005)............... 14

Equation 3 Chlorine Conversion from Nitrite (Liu, Taylor, Randall, & Dietz, 2005) 15

Equation 4 Chloramine Conversion to Nitrite (Liu, Taylor, Randall, & Dietz, 2005) 15

Equation 5 First Step in Converting AOB to Nitrite... 15

Equation 6 Second Step Converting from Nitrite to Nitrate.. 15

Equation 7 Variance of Linear Composite... 41

Equation 8 Constraint to Weight Factor 41

Equation 9 Factor Analysis Model Equation... 46

Equation 10 Condition Index 49

Introduction

Access to adequate, safe potable water supplies is a critical requirement for a stable society. That is why governments have attempted to manage this critical resource using public water works utilities (Bloetscher, Infrastructure Management, 2019). For over 150 years, improvements have been made to waterworks systems in an effort to reduce waterborne disease outbreaks. For much of the developed world, Western Europe, Canada and the US, these efforts have been relatively successful as waterborne disease outbreaks are uncommon. This is a very different scenario that the ancient world when water was more difficult to acquire.

In the beginning people had to use buckets, pots and other vessels to carry water from the local well or river in their homes. The use of water was minimalized due to the effort that took place to get water from one point to another. The earliest water distribution systems were in the Minoan cites of Crete, dating back to 1400BC. Some of those pipes are still in service today (Walski, 2006). Other cites, like Ephesus and Perge located in Asia Minor, had pipe networks made of clay that are over 2000 years old to move water for irrigation and other uses to villages (Walski, 2006). The first screw pumps used to irrigation and removal of water from ships were used in the first century BC, invented by Archimedes of Syracuse (Walski, 2006).

The most extensive water distribution system of ancient times was built by the Romans. They built the first aqueduct around 312 BC and continued to build them to convey water around the Roman Empire. The Romans were also the first civilization to use

lead pressure pipes (Walski, 2006). During the dark ages technology for water distribution ceased and the existing systems were allowed to decay (Mays, 1999). Systems in those times were far worse than during the Roman Rule.

It was not until the thirteenth century that water distribution technology began again when a 5.5-kilometer lead pipe was installed from Tybourne Brook to London in the United Kingdom (Walski, 2006). Over time other lines were installed, but the same problem existed as before. The water lines brought water to a centralized location in the town and residents had to use buckets to bring water to their homes.

By the mid eighteenth century, the city of London had installed over fifty kilometers of water mains of various construction (Sanks, 2005). These mains included lead, wood, and cast-iron pipe. Many of the smaller diameter pipe was bored wood logs that were hollowed out to carry water. They had handmade bell and sockets that would be used in conjunction with lead to seal the water main (Sanks, 2005). For larger diameter water mains wood staves, similar to whisky barrels, were used. The wood would be assembled dry and banded in that state. Then water would be introduced, causing the wood to swell and seal. Many of these systems were used into the nineteenth century. This system worked rather well for the period. Even today occasionally a news report would come out of a water utility finding wood water mains that are still in service and over one hundred years old.

The first cast iron water main was installed in Philadelphia, Pennsylvania in 1817 (Walski, 2006). Originally cast-iron pipes were made vertically in pits. Over time the wood water main was replaced with cast iron pipe. This pipe became more favorable during the 1800’s (Sanks, 2005). Eventually technological advances allowed cast iron to be developed

by the use of centrifuges (Walski, 2006). Another style of iron pipe was also introduced during that time period, wrought iron. The city of San Francisco in California installed over 125 kilometers of wrought iron pipe in 1892 (Sanks, 2005).

Over time and with advances in technology, pipes that were joined with poured lead soon became roll on joints. A roll-on joint consisted of rubber compression gasket that would be installed as the bell and socket were inserted (Cast Iron Soil Pipe Institute, 2006). Those joints eventually became the push on joints used today (Walski, 2006). Besides the push on joints seen today, mechanical, welded, flanges, ball and socket, and restrained jointing methods are used to hold mains together. Cement and mortar lined pipes began to be used in the 1920's in Charleston, South Carolina (Walski, 2006). The cement reduced the contact between the water and the iron, reducing water color complaints. By the 1940's, the use of cement mortar lined pipes in cast iron pipes became standard practice to inhibit internal corrosion (Walski, 2006).

The next advancement to come in water distribution was the advent of ductile iron pipe. First used in 1948, ductile iron pipes were used to replace the old cast iron and wood pipes of years before after it was discovered that the tensile strength of ductile iron was greater than that of cast iron (Sanks, 2005).

Although archaeologist discovered its use in clay pots dating back 10,000 years; the use of asbestos took off at the turn of the century (Mesotehlioma Justic Network, 2019). Most asbestos was sourced from Quebec, Canada for use in the North America. Asbestos is mined from metamorphic rocks. Asbestos is a generic name that was given to the six naturally occurring fibers that makeup the product. Those fibers are Chrysotile, Amosite, Crocidolite, Anthophyllite, Tremolite, and Actinolite (Asbestos.comLLC, 2019). After the

World War 2, the use of asbestos in water increased. Asbestos pipe was sold under the generic brand name of Transite from the 1950s through 1980. Asbestos fell out of favor in the 1970's after it was discovered that exposure to the fibers caused cancer (Mesotehlioma Justic Network, 2019). As most of these pipes are entering the end of their service life, they are being replaced with newer materials.

The most widely used pipe today is made from polyvinyl chloride or PVC. PVC pipe was originally developed in the 1930's in Germany (Walski, 2006). PVC was introduced in the United States in the 1950's. As standards developed for compatibility with ductile and cast-iron pipe in the 1970's, PVC saw a major increase in use for piping purposes (Walski, 2006). In 1976 the American Water Works Association developed the first standard for C-900 PVC (Uni-Bell PVC Pipe Association, 2012). High density Polyethylene is also gaining favor with respect to directionally drilled water mains.

Effort to Address Public Health – Regulations

To protect public health in the US, Congress instructed the US Environmental Protection Agency (USEPA) to establish National Primary Drinking Water Regulations (NPDWRs) that set enforceable water quality standards for drinking water contaminants (US Envrionmental Protection Agency , 2002). These enforceable standards created "maximum contaminant levels" (MCLs), which are the maximum allowable concentration of a contaminant in drinking water. Currently in the US there are over 90 regulated contaminants in drinking waters, including seven microorganisms. These include the pathogenic microorganisms *Cryptosporidium, Giardia lamblia, Legionella* and enteric viruses, plus several indicators of microbial risk and treatment system effectiveness,

including heterotrophic plate count, total coliforms, and turbidity (Bloetscher & Plummer, Evaluating the Significance of Certian Pharmaceuticals and Emerging Pathogens in Raw Water Supplies, 2011).

USEPA has also established National Secondary Drinking Water Regulations (NSDWRs) that set non-mandatory water quality standards for 15 additional contaminants. These contaminants are ***not*** considered to present a risk to human health, so there is no associated enforcement. Instead, secondary MCLs are established as guidelines to assist public water systems in managing their drinking water for aesthetic considerations, such as taste, color, and odor (US Envrionmental Protection Agency, 2002) (Walski, 2006). Those standards determined that the water quality parameters had to be met at the tap and not the source (Walski, 2006). Backflow prevention, lead and copper rules, and monitoring of disinfectants were only some of the items that came about due to the Safe Water Drinking Act.

It was always assumed that the water one put in the distribution system at the plant would maintain its quality throughout the distribution system. Yet despite these efforts, periodic waterborne illnesses are reported. The Centers for Disease Control (CDC) and USEPA have maintained a collaborative surveillance system for collecting and reporting waterborne disease outbreaks since 1971 (Bloetscher & Plummer, Evaluating the Significance of Certian Pharmaceuticals and Emerging Pathogens in Raw Water Supplies, 2011). For the time spanning from 1997 through 2006, 137 waterborne disease outbreaks were reported to the CDC, with a total of 8,498 illnesses and 17 deaths (Lee & Kim, 2003). Of the outbreaks with a known cause (101), 17 were attributed to chemical or toxin poisoning and 84 to pathogens. Nonspecific bacteria were among the most commonly

implicated pathogens. The highest number of outbreaks where the culprit was known were due to *Legionella, Giardia, Campylobacter,* norovirus and *E. coli* O157:H7 with 78% of outbreaks attributed to groundwater systems (Bloetscher & Plummer, Evaluating the Significance of Certian Pharmaceuticals and Emerging Pathogens in Raw Water Supplies, 2011). The source of the bacteria at the tap was not identified. However, the delivery of contaminated water to the tap can indicate either a treatment issue or an issue in the distribution system, so at least a portion of these issues are likely to have occurred as a result of impacts from the distribution system, given the amount of testing at water plants (noting that many of the groundwater systems may do nothing more that chlorinate the system – Alamosa, CO did not chlorinate when they had their issues in 2008). Water distribution issues are more challenging to find, especially if they relate to biofilms. The Centers for Disease Control and Prevention (CDC) identified biofilms as the source for 65% of human bacterial infections from community water supply associated outbreaks (US Envrionmental Protection Agency, 2002) reported epidemiological evidence that children may incur multiple digestive problems per person per year that are attributable to pathogens that developed in the water transmission network after disinfection.

Note that only coliforms are routinely analyzed for in drinking water, as mandated under the Total Coliform Rule (TCR) and the Ground Water Rule of the Safe Drinking Water Act (SDWA), while many of these opportunistic pathogens are not. In addition, the primary ingredient that causes nitrification is excess ammonia in a distribution system. Currently there is no regulations on the amount of ammonia that can be contained in a water supply. While the United States does not have a standard, the European community has established a maximum contaminant limit of 0.5 mg/L (American Water Works

Association, 2013). They have also established a guide limit of 0.05 mg/L (American Water Works Association, 2013). Even though there is no federal guidance on the level of ammonia in water in the near future; it is beneficial for utilities to reduce the amount of excess ammonia in the water supply (Lansey & Boulos, 2005).

How biofilms form

Microorganisms are the most widely distributed life forms on the planet and are known to inhabit and thrive in the presence of moisture and nutrients, both of which exist in water distribution networks (Videla, 1996) (Bloetscher F. , Bullock, Fergen, Witt, & Fries, 2002). Biofilms are persistent because they are complex aggregates of microorganisms embedded in a highly hydrated extracellular matrix that shows structural heterogeneity resulting from a diverse and complex microcosm (Videla, 1996) (Bloetscher F. , Bullock, Fergen, Witt, & Fries, 2002), (Rayburn, Blaha, Anderson, & Rotert, 2011) (Ruda, 1997; Ruda, 1997), (Trussell, 1999) (Eboigbodin, Serh, & Biggs, 2008). During biofilm growth, microorganisms excrete a matrix of extracellular polymeric substances (EPSs), which lead to the formation of a slime layer that connects cells and anchors them to the surface and to each other (Bloetscher F. , Bullock, Fergen, Witt, & Fries, 2002). In fact, fifty to ninety percent of the biofilm matrix is extra-cellular matrix (Eboigbodin, Serh, & Biggs, 2008), with the remaining amount consisting of microbial cells. These remaining cells are considered to be freely living cells that possess the ability to perform physiological functions and have the ability to adapt to certain environmental changes (Eboigbodin, Serh, & Biggs, 2008) (Schrantz, Pressman, & Wahman, 2013).

Biofilms in water distribution systems are often observed as an accumulation attached to the inner wall of a water distribution line. From the microbial perspective, the extra-cellular matrix of a biofilm provides an ideal habitat given that they included a source of nutrients, oxygen stratification, resistance to velocity currents, encapsulation for resistance to disinfectants, and protection from grazers and biocides (Videla, 1996). From the utility perspective, the undesirable accumulation of biofilms with actively growing slime layers can become a considerable issue for water distribution systems, particularly with regard to maintenance issues such as iron pipe damage (corrosion and tubercles) and economic consequences (early pipe failure/replacement and more frequent flushing of clean water), as well as public health issues such as proliferation of pathogens.

Mature biofilms in drinking water distribution systems can also be a highly diverse potential source of human pathogens. A wide range of primary and opportunistic pathogens that cause disease in individuals with underlying conditions that may facilitate infection, have demonstrated the ability to survive and thrive in biofilms (Bloetscher *et al.,* 2010). Bloetscher et al (2010 noted that "Primary pathogens, opportunistic pathogens and indicator organisms including *Clostridium* (Emde, Smith, & Fancey, 1992), *E. coli* (Emde, Smith, & Fancey, 1992); (Geldreich, 1996); (Satory & Holmes, 1997), *Enterobacter* (LeChevallier, Lowry, & Lee, 1990); (Emde, Smith, & Fancey, 1992); (Geldreich, 1996); (Satory & Holmes, 1997); (Lee & Kim, 2003), *Legionella* (Murga, et al., 2001), *Pseudomonas* (Bloetscher F. , Bullock, Fergen, Witt, & Fries, 2002); (LeChevallier, Lowry, & Lee, 1990); (Emde, Smith, & Fancey, 1992); (Geldreich, 1996); (Norton & LeChevallier, 2000); (Lee & Kim, 2003), (Bloetscher & Plummer, Evaluating the Significance of Certian Pharmaceuticals and Emerging Pathogens in Raw Water Supplies,

2011), and *Staphylococcus* (Geldreich, 1996) (Lee & Kim, 2003), among others have been reported in biofilms collected from water distribution networks." Other pathogenic microorganisms have also been isolated from biofilms including viruses, fungi, yeast, protozoa such as amoebae and ciliates, diatoms and other algae, invertebrates, and microbial toxins (Eboigbodin, Serh, & Biggs, 2008), (Bachmann & Edyyean, 2006).

Videla, (1996) reported that the steps in the life cycle of a biofilm include attachment, slime formation, growth, and detachment or sloughing. The first microorganisms to attach are called "*pioneers*," which are most commonly facultative anaerobes. Vanhaeecke, et al., (1990) conducted a study of permanent adhesion rates (a precursor to biofilm formation) between electro-polished surface and rougher surfaces (400, 320, and 120 grit surfaces) at various pH values. While the electro-polished steel adhesion rate was up to 100 times slower for a biofilm than the 120 grit surfaces, in all cases, permanent adhesion occurred in a matter of minutes (Vanhaeecke, et al., 1990). Key requirements for biofilm development include an active microbial community and interaction with pipe materials (Videla, 1996). As the biofilm continues to coat the pipe surface, acid-formers can reduce the pH near the pipe wall and accelerate corrosion (Videla, 1996). A depleted oxygen layer forms near the wall and an anaerobic environment in which sulfate-reducing bacteria (SRB) proliferate because the transport of oxygen into the anaerobic layer of the biofilm is limited by the biological activity in the upper layers (Bloetscher F. , Bullock, Fergen, Witt, & Fries, 2002). Anaerobic cause larger issues (Miller, Friedman, Koci, & Moore, 2013) - there are cases where the extracellular polymeric substances are gram positive as in staphylococci (Donlan, 2002). The ESP provides the opportunity to a wide variety of bacteria to join the matrix. Biofilms are an

aggregation of bacterial cells that have attached to a surface and produced extracellular polymeric slime that form microscopic layers or sheets on the inside of water mains and other locations. (Heydorn, et al., 2002) (Miller, Friedman, Koci, & Moore, 2013) (Hwang, 2010) (Rayburn, Blaha, Anderson, & Rotert, 2011). Bacteria can cause corrosion directly by their metabolic processes, forming specific chemical species such as ammonia, hydrogen sulfide, dissolved sulfate, ferric or manganic chlorides (Dillion, 1995).

The heterotrophic biomass typically found in a biofilm supports the synergistic effects of different growth rates for different layers of the biofilm matrix,, mixed metabolisms, different preferences for nutrient processing, Different means to consume oxygen, and a high surface area to volume ratio, which allow the biofilm to thrive even within the relatively hostile conditions of a pipe environment (Dillion, 1995). Aerobic bacteria present in the biofilm can also corrode metals directly via oxidation (Videla, 1996), which attracts other species. The appearance of SRB is indicative of mature biofilm growth.

As the biofilm matures, it grows thicker from the diverse community of microorganisms attracted to the biofilm and from the accumulation of particles that stick to the ESP matrix. At some point the thickness of the biofilm will create a situation where the “weight” of the biofilm exceeds its adhesive capacity. In such cases, the bulk fluid velocity will shear off the portions of the biofilm closest tot eh water – the matrix closest to the pipe will remain in place. The action is called sloughing, and releases bacteria into the water where it can travel to another location to establish a new matrix, re-attach and create new sources of taste, odor, and color issues in the distribution network. Sloughing can also occur if the system is disturbed by changes in velocity of the water or water hammer.

Once mature colonies are established, the effects of microbiologically induced corrosion (MIC) are often seen, but also the reduction in chlorine residual, color, tastes and other issues can develop. Sulfur can be reduced by anaerobic bacteria to release hydrogen sulfide, which can significantly increase the susceptibility of the pipe to pitting. The issue is of particular concern in water distribution systems constructed with metallic materials, given the MIC, if not addressed, can result in pipe damage/failure, premature aging/replacement, clogging, and increased maintenance requirements. Proper corrosion control has also been shown to increase disinfection effectiveness on biofilms in iron pipes (LeChevallier, Lowry, & Lee, 1990).

Bloetscher F. , Bullock, Fergen, Witt, & Fries, (2002) noted that "the key factors that affect biofilm development in distribution networks include: environmental factors (pH, temperature, dissolved oxygen, etc.), water quality (nutrients, inorganics, dissolved organic carbon, etc.), pipe materials, system hydraulic regime (stagnant conditions), corrosion control measures, presence of a disinfectant residual, age of pipe, and sediment accumulation. Most of the data on the factors that influence biofilm development are based upon changes in total viable counts (e.g., heterotrophic plate count) or on changes in the growth of specific microorganisms (e.g., total coliforms)." Although a number of comprehensive review articles have been published (Geldreich, 1996); (Batte, Mathieu, Laurent, & Prevost, 2003); (Bachmann & Edyyean, 2006); (Geldreich, 1996), the interaction among these factors is complex and variable, making predictions difficult. Examples include:

- Speller (1951) reported that iron bacteria can cause damage to pipes due to the deposition of iron compounds, resulting in clogging or tuberculation of

pipes and red water Videla (19960 noted that MIC induced tubercles can impede the penetration of biocides and corrosion inhibitors).

- Sulfate-reducing bacteria may accelerate tubercle growth in the inner area of the tubercle where conditions are anaerobic.
- Hydrogen sulfide production reduces pH as well as causing corrosion in conjunction with chloride ions (Dillion, 1995).
- *Gallionella* has been shown to be dependent on a symbiotic relationship with sulfate-reducing bacteria (Videla, 1996).
- The slime forming *Pseudomonads* are frequently found near tubercles (Videla, 1996).

Anaerobic conditions will form under deposits, in crevices and under the influence of BOD or COD, independently of dissolved oxygen content, typically an indicator of biofilm development.

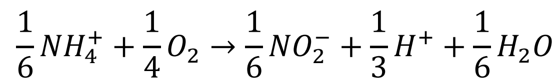
Nitrification in the Water Distribution System

The process of nitrification involves ammonia. Ammonia occurs naturally in some groundwater supplies, and groundwater can become contaminated with nitrogen as agriculture runoff percolates into aquifers. In many places, ammonia is common and can be well above 1 mg/L in some groundwater. Ammonia is also deliberately added to a chloraminated water supply. In drinking water, ammonia is often added where maintenance of a free chlorine residual is difficult due to water temperature, large retention tie of the water in the distribution system, presence of organic constituents, and local preferences. The process is called chloramination. Chloramination is commonly used for secondary

disinfection purposes to control microbial growth in finished water because it does not form trihalomethanes and it forms a residual that is longer lasting than chlorine. This is particularly at issue in water systems that have warm water (>60°F) and organics (disinfection byproduct precursors). In these systems, the concentration of free ammonia present in the finished water will be a function of the chlorine to ammonia-N (Cl_2:NH_3-N) ratio. Free ammonia is almost completely eliminated when a 5:1 weight ratio of Cl_2:NH_3-N is used (Kirkmeyer, Richards, & Smith, 1994) (Kirkmeyer, Friedman, Martel, Noran, & Smith, 2001).

Ammonia-oxidizing bacteria are encouraged when chloramination processes are used (Skadsen, 1993) (Wahman, Maestre, & Speitel Jr., 2016). Ammonia is required for nitrifying bacteria (Skadsen, 1993). As the disinfectant is released during bacteria reduction in a distribution system, the remaining ammonia is left behind. The process of nitrification is a two-step process (Liu, Taylor, Randall, & Dietz, 2005). Ammonia-oxidizing bacteria then work to convert that free ammonia into a nitrate (NO_3^-). Nitrosomonas is the bacteria that causes ammonia to convert to nitrate (Skadsen, 1993) (American Water Works Association, 2013). Its chemical change is shown in *Equation 1*.

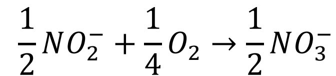
Equation 1 Composition of Nitrosomonas (Liu, Taylor, Randall, & Dietz, 2005)



$$\frac{1}{6}NH_4^+ + \frac{1}{4}O_2 \rightarrow \frac{1}{6}NO_2^- + \frac{1}{3}H^+ + \frac{1}{6}H_2O$$

If left untreated this process continues to degrade and converts those nitrates to nitrite (NO_2^-) (Pintar, Anderson, Smith, & Huck, 2005). The conversion of nitrite to nitrate is also a chemical change called Nitrobacter and is shown in *Equation 2* (Skadsen, 1993). Nitrobacter is the sole nitrite oxidizer that is responsible for converting nitrites to nitrates (Skadsen, 1993) (American Water Works Association, 2013).

Equation 2 Composition of Nitrobacter (Liu, Taylor, Randall, & Dietz, 2005)

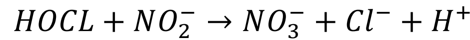


$$\frac{1}{2}NO_2^- + \frac{1}{4}O_2 \rightarrow \frac{1}{2}NO_3^-$$

The initial conversion of ammonia from NH_3 to a nitrate is cause for concern and is indicative of partial nitrification (Pintar, Anderson, Smith, & Huck, 2005). These reactions are complicated but are simplified in *Equation 3* and *Equation 4*. This process leads to increased disinfectant reduction at an accelerated rate. The conversion from free ammonias presence in the distribution system to a nitrate can easily be missed. The absorption of disinfectant is minimal during initial stages of conversion (Pintar, Anderson, Smith, & Huck, 2005) (US Envrionmental Protection Agency , 2002). During this initial stage the residual may be lower but still within the range most would consider normal. Some systems have been found to maintain a hardy residual during conversion from free ammonia to nitrate and even into nitrite conversion (Pintar, Anderson, Smith, & Huck, 2005). The deterioration of chlorine residual rapidly increases the longer the conversion of ammonia to nitrates is allowed to continue (Liu, Taylor, Randall, & Dietz, 2005). Studies have shown that even in the presence of 1.2 to 1.5 mg/L monochloramine residuals ammonia-oxidizing bacteria is capable of growth (Pintar, Anderson, Smith, & Huck, 2005). Even with a chloramine residual as high as 5 mg/L, the pipes have still shown signs of nitrification (Liu, Taylor, Randall, & Dietz, 2005). (Skadsen, 1993) conducted a study that found a monochloramine residual of 8 mg/L was not effective in reversing the nitrification that had occurred in the Ann Arbor Water plant in 1993. Additional research has also found that ammonia oxidizing bacteria become resistant to monochloramine disinfection then free chlorine disinfection (Skadsen, 1993). *Equation 3* and *Equation 4* show the chemical

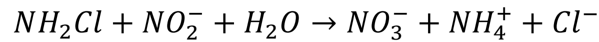
transformation of a 5 mg/L chlorine demand that is exerted on a 1 mg/L nitrite sample (Liu, Taylor, Randall, & Dietz, 2005) (Skadsen, 1993).

Equation 3 Chlorine Conversion from Nitrite (Liu, Taylor, Randall, & Dietz, 2005)



$$HOCL + NO_2^- \rightarrow NO_3^- + Cl^- + H^+$$

Equation 4 Chloramine Conversion to Nitrite (Liu, Taylor, Randall, & Dietz, 2005)



$$NH_2Cl + NO_2^- + H_2O \rightarrow NO_3^- + NH_4^+ + Cl^-$$

According to USEPA (2002), the microbial process for nitrification involves two sequential steps: 1) the oxidation of nitrogen compounds (primarily ammonia), to nitrite, and 2) the oxidation of nitrite to nitrate. The nitrification process is primarily accomplished by two groups of autotrophic nitrifying bacteria that can build organic molecules using energy obtained from inorganic sources, in this case ammonia or nitrite. In the first step of nitrification, ammonia-oxidizing bacteria oxidize ammonia to nitrite according to Equation 5.

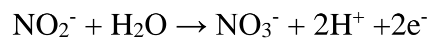
Equation 5 First Step in Converting AOB to Nitrite

$NH_3 + O_2 \rightarrow NO_2^- + 3H^+ + 2e^-$

Nitrosomonas is the most frequently identified genus associated with this step, although other genera, including *Nitrosococcus*, and *Nitrosospira*. Some subgenera, *Nitrosolobus* and *Nitrosovibrio*, can also autotrophically oxidize ammonia (Watson, Valos, & Waterbury, 1981).

In the second step of the process, nitrite-oxidizing bacteria oxidize nitrite to nitrate according to Equation 6 (US Envrionmental Protection Agency, 2002).

Equation 6 Second Step Converting from Nitrite to Nitrate



$NO_2^- + H_2O \rightarrow NO_3^- + 2H^+ +2e^-$

Woolschlager et al. (2001) describe five mechanisms of ammonia release. Nitrite and nitrate are produced during nitrification through ammonia utilization by nitrifying bacteria. According to Equation 5 and Equation 6, for every mole of ammonia-N, a 1-mole equivalent of nitrite-N is produced. Subsequently, for every mole of nitrite-N produced, a 1- mole equivalent of nitrate-N is produced (US Envrionmental Protection Agency, 2002). According to (Valentine, Ozekin, & Vikesland, 1998), the overall net stoichiometries can be used to examine the relationship between chloramine decay and ammonia production.

Nitrobacter is the most frequently identified genus associated with this second step, although other genera, including *Nitrospina*, *Nitrococcus*, and *Nitrospira* can also autotrophically oxidize nitrite (Watson, Valos, & Waterbury, 1981). Various groups of heterotrophic bacteria and fungi can also carry out nitrification, although at a slower rate than autotrophic organisms (Verstraete & Alexander, 1973); (Watson, Valos, & Waterbury, 1981). Speciation of nitrifying bacteria in drinking water systems (Wolfe, Leiu, Izaguirre, & Means, 1990) suggest that the number of heterotrophic nitrifiers in drinking water systems may be negligible compared to autotrophic nitrifiers.

Water pH value is an important factor in nitrification activity for two reasons. First, a reduction of total alkalinity may accompany nitrification because 5 moles of protons are released (consuming alkalinity) in the conversion of ammonia to nitrate. A model developed by (Gujer & Jenkins, 1974) indicates that 8.64 mg/L of bicarbonate (HCO_3^-) will be utilized for each mg/L of ammonia-nitrogen oxidized. While consuming alkalinity does not impose a direct public health impact, it does lower the buffering capacity, which can impact pH stability and corrosivity of the water toward lead and copper (Snoeyink and Jenkins, 1980).

Secondly, nitrifying bacteria are very sensitive to pH. Nitrifying bacteria are obligate aerobic organisms commonly found in terrestrial and aquatic environments (Holt, Todd, Delanoue, & Colbourne, 1995) (Watson, Valos, & Waterbury, 1981). Their growth rates are controlled by substrate (ammonia-nitrogen) concentration, temperature, pH, sunlight, oxygen concentration, and microbial community composition (U.S. Environmental Protection Agency, 2002). *Nitrosomonas* has an optimal pH between approximately 7.0 and 8.0, and the optimum pH range for *Nitrobacter* is approximately 7.5 to 8.0. Some utilities have reported that an increase in pH (to greater than 9) can be used to reduce the occurrence of nitrification (Skadsen, 1993). According to (Wilczak, 2001), pH appears to be the most important factor controlling the rate of chloramine auto-decomposition. In Florida, (Cates & Lavinder, 1999) noted that raising pH reduced nitrification– a pH above 8.7 limits nitrification activity.

Under the Safe Drinking Water Act (SDWA), primary MCLs have been established for nitrite-N, nitrate-N, and the sum of nitrite-N plus nitrate-N. The MCLs are 1 mg/L for nitrite-N, 10 mg/L for nitrate-N, and 10 mg/L for nitrite + nitrate (as N). The current nitrite and nitrate standards are measured at the point of entry to the distribution system so any subsequent elevated nitrite/nitrate levels resulting from nitrification within the distribution system are not identified by compliance monitoring.

Of interest is that the groundwater rule permits reduced monitoring if 4-logs of virus removal can be demonstrated, but this only applies when water sources have no ammonia in them (US Envrionmental Protection Agency, 1989). Regulatory agencies encourage compliance with the Groundwater Rule, but due to the presence of ammonia in raw water, some regulatory agencies have argued that compliance is possible if they first

“chlorinate the ammonia out of the water,” then create a free chlorine residual, then add ammonia back in to create chloramines (monochloramine). The problem with this solution is that a) it does not comply with the federal rule and b) the nitrogen nutrient does not disappear – it is just reacted with chlorine. That mean more nitrogen as a nutrient that can become bioavailable in transit.

How Bacteria Get into the Distribution System

There are a limited number of methods in which bacteria can be introduced into a water system. The first way is through the bypass of the treatment process (Besner, Gauthier, Servais, & Camper, 2002) (Miller, Friedman, Koci, & Moore, 2013). Ineffective or out of compliance treatment processes can allow coliforms and bacteria to become introduced into a distribution stream. The second method is through cross connection or introduction post treatment (Besner, Gauthier, Servais, & Camper, 2002). Cross connections can occur through various methods. One such method is a careless homeowner leaving a well connected to their house plumbing and their well pump overcoming the distribution system pressure. Another method to introduce microorganisms is through the use of vented tanks (Trussell, 1999) (Miller, Friedman, Koci, & Moore, 2013).

The use of vented tanks allows levels to rise and fall without extra equipment but those vents can also allow airborne bacteria to be introduced into the treated water supply (Hiltebrand & Theiss, 2002). These bacteria can become attached to the concrete walls of the tanks and allow bio-growth to be introduced to the warm, dark, and moist surfaces above the water line (Hiltebrand & Theiss, 2002). Think of all of the airborne pollen during the spring and then think about how much of that makes it past the animal screens that exist at water plant storage tank farms. The use of disinfectants alone is not satisfactory in

removal from tanks. Requirements that tanks be chemically cleaned to penetrate the mineralization and slime barrier (Miller, Friedman, Koci, & Moore, 2013).

A third is through repair of leaks or during negative pressure events (Besner, Gauthier, Servais, & Camper, 2002). These events occur when there is a small leak in a water main and the pressure inside the water main drops below that of the pressure exerted on the outside of the pipe (Miller, Friedman, Koci, & Moore, 2013). This can occur during rapid opening of fire hydrants, pump cycling, rapid opening or closing of valves, or loss of power (Besner, Gauthier, Servais, & Camper, 2002). This pressure loss due to breaks or repairs can allow bacteria to enter the water main. These different pathways are generally indistinguishable (Besner, Gauthier, Servais, & Camper, 2002) due to the small amount of water that is tested in the distribution stream. While repairs are made, contamination of the distribution pipe can occur (Cartier, et al., 2009) if the materials used for the repair are not effectively disinfected.

Detection of Biofilms/Nitrification

Currently, researchers lack techniques for effective detection, diagnosis, and control of biofilms. Low pH, low hardness, high chlorides, high sulfates, and a ratio of chlorides to bicarbonate above 0.3, all indicate a greater potential for corrosion and biofouling (Geldreich, 1996). One effective diagnostic tool involves the heterotrophic plate count (HPC) test. Samples that yield more than 500 CFU/100 mL and have chlorine residuals less than 0.2 mg/L typically indicate stagnant water or conditions that promote biofilm growth (Geldreich, 1996). Since HPC is non-selective, biofilm material can be collected from a pipe and cultured to isolate specific organisms. For instance, persistent

coliform levels may indicate extensive biofilm shedding (Crozes & Cushing, 2000). Biofilms can become a point source of coliforms, leading to TCR violations. Therefore, mechanisms for controlling biofilms may be of benefit to reducing coliform levels as well as other opportunistic pathogens.

According to Wilczak et al. (1996), nitrification is often indirectly identified by detection via monitoring the following (US Envrionmental Protection Agency, 2002):

- **Low residual disinfectant.** Low levels of residual disinfectant can allow bacteria in the biofilm to multiply. When residual disinfectant levels drop below the normal baseline, nitrification may soon follow. So, monitoring and mapping levels of residual disinfectant proves to be a quick and inexpensive tool to pinpoint affected areas and focus mitigation efforts.
- **Nitrite and nitrate levels in distribution.** Spikes in nitrite and/or nitrate levels can signal when action is needed. Systems are required to sample nitrate and nitrite quarterly in the distribution system. More frequent monitoring may allow early warning of potential problems.
- **Ammonia levels.** Ammonia is food for nitrifying bacteria. So, if ammonia levels are decreasing in at least part of the distribution system, nitrification could be the cause. Once a baseline ammonia levels is determined, frequent monitoring may be able to detect decreasing ammonia levels in localized areas of the distribution network. If the ammonia levels in the system are less than the ammonia levels of the water leaving the treatment plant, or less than the system's baseline ammonia levels, nitrification may be occurring.

- **Decrease in dissolved oxygen.** Decrease in dissolved oxygen may indicate bacteria are consuming oxygen via the nitrification process.
- **Drop in pH and alkalinity.** Sequential nitrification generates protons, which reduce the pH and consume alkalinity, lowering the buffering capacity of the water.
- **Temperature change.** Increase in water temperature facilitates bacterial growth. Water temperature is also been found to be a contributing factor in the formation and growth of biofilms (Zhang, Semmens, Schuler, & Hozalski, 2002). Temperature can also control the efficiency of disinfectants used to control biofilm growth. Studies show that an eighteen Fahrenheit degree drop is temperature will decrease biofilm growth activity by at least fifty percent (Eboigbodin, Serh, & Biggs, 2008) (Besner, et al., 2001). In studies in homes, the researchers discovered that the HPC count was 250 times higher in the hot water lines of apartment buildings then in the cold-water lines (Eboigbodin, Serh, & Biggs, 2008). Measurable effects on biofilm growth were shown even on small increases of temperature a little as four degrees (Ollos, Huck, & Slawson, 2003). This is definitive evidence that temperature is a major factor in the growth of biofilms. While temperature effects the speed in which biofilms grow, it has been shown that the presence of biofilms is not temperature dependent (Lin, Vidic, Stout, & Yu, 1998).
- **HPC counts.** Frequently, but not always, systems that have nitrification occurring may also have increases in heterotrophic plate counts (HPC), coliform-positive test results, or both.

If these issues occur, nitrification is suggested, and a more active control strategy must be put in place. In addition, operations personnel may monitor water quality for low hardness, high chlorides, high sulfates and if the ratio of chlorides to bicarbonates are above 0.3 (U.S. Environmental Protection Agency, 2002). Detection can be accomplished by using agar test to look for heterotrophic bacteria (U.S. Environmental Protection Agency, 2002). This was discovered by the US EPA as they were attempting to detect total coliforms. They found that the biofilms interfered with the total coliform testing when a large biofilm that had a portion slough off a pipe in the distribution system, creating an unusual heterotopic bacteria (US Envrionmental Protection Agency , 2002). While this is not a fool proof test to find biofilms in a distribution system, it is simpler than trying to obtain a sample directly from the pipe.

A method that was used during this process was to system “A” during the collection of samples. System “A” had known chlorine residual issues. The water plant superintendent had a feeling that biofilms existed in a specific area of the system. Due to the low chlorine issues, presumably caused by the presence of biofilms; a series of fire hydrants had been running. Some of these hydrants had been running for many months. When it came time to gather samples, those fire hydrants were able to be used. A long cotton swab was inserted deep into the fire hydrant during low flow. A sample of the biofilm was removed and sent to the lab for testing. Table 1 shows the results of this testing.

Water age is an emerging indicator of potential biofilm and other bacteriological growth (Mercer, 2011). Discoveries have been made showing that the older a treated water source is, the disinfectant residual degrades, the formation of disinfection byproducts begin to form, and customer based taste and odor complaints rise (Mercer, 2011). The key to

reducing the age of water is to keep it moving in the distribution system. This can be difficult in areas where the system is overbuilt or a seasonal area that has multiple months of lowered residency.

Biofilm/ Nitrification Control Strategies

Biofilms are difficult to remove from pipelines once established, and therefore they must be controlled and monitored. Many different methods have been used to control biofilms; however, in most circumstances, biofilm control requires the use of a variety of tools, and the relative effectiveness is typically site-specific. In general, biofilms cannot be removed, they can only be managed or controlled once initiated. They can be managed by removing organic matter and nutrients during water treatment, inactivation of microorganisms via disinfectants and maintenance of appropriate disinfectant residuals, and proper distribution system maintenance practices (i.e. flushing, avoiding stagnant conditions, minimizing the corrosion of iron pipe surfaces, and managing contamination from external sources). Typical methods for controlling growth in distribution systems include, the use of free chlorine, unidirectional flushing, standard hydrant flushing, and the use of mechanical pigs in the distribution systems (Trussell, 1999).

For many years, the use of sodium hypochlorite was believed to be the best method for eliminating biofilms and other harmful bacteria (American Water Works Association, 2013). It has been realized that the use of free chlorine in a distribution system alone will not solely reduce the presence of bacteria in a water system (LeChevallier, Lowry, & Lee, 1990). In multiple testing scenarios, it was observed that free and mono chlorine reacts differently to biofilms depending on the makeup of the distribution system. In the Los Angeles Aqueduct, it was found that 1 to 2 mg/L of free chlorine was not enough to control

bacterial growth (LeChevallier, Lowry, & Lee, 1990). The aqueduct had been recently cleaned with a 15 to 20 mg/L solution of chlorine as the system was having issues with biofilm fouling at the reverse osmosis membranes (LeChevallier, Lowry, & Lee, 1990). It was soon discovered that the biofilms returned quickly. This is because the microorganisms found in biofilms are more resistant to biocides because biofilms provide a barrier to mass transfer (Bloetscher et al 2010). Once a biofilm is established, it may take a high level of chlorine residual (> 0.2-1.0 mg/L) to reduce microbial levels appreciably (LeChevallier, Lowry, & Lee, 1990). Fass, et al., (2003) suggested that it may take 1-2 months of continuous chlorination to eradicate biofilm-associated biomass and desorbable organics at a dose of 3.5 mg/L Cl_2 with a residual of 0.1-0.2 mg/L Cl_2. Superchlorination is a solution in problematic areas, but isolation from the rest of the network can be problematic.

Chloramines can be an effective alternative because they are capable of deeper penetration into the biofilm, but they are less reactive (LeChevallier, Lowry, & Lee, 1990) (Bachmann & Edyyean, 2006). (LeChevallier, Lowry, & Lee, 1990), concluded that the use of monochloramines below a residual of 2 mg/L was effective in controlling additional biofilm growth on PVC, copper, or galvanized pipe systems (LeChevallier, Lowry, & Lee, 1990) (Ollos, Huck, & Slawson, 2003). They further found that the use of free chlorine in iron pipes was ineffective, even at high doses of 3 to 4 mg/L was ineffective in the control of biofilms. The observation was that a monochloramine level above 2 mg/L was effective in controlling biofilms in iron pipes (LeChevallier, Lowry, & Lee, 1990). Another study showed that even with the presence of chloramines, microorganisms are still present (Pintar, Anderson, Smith, & Huck, 2005). The reason found was these microorganisms are becoming resistant to chlorine (Eboigbodin, Serh, & Biggs, 2008).

As a result, where the potential for biofilms exist, a common control strategy employed by utilities is the maintenance of high distribution system disinfectant residuals (greater than 2 mg/L) with periodic breakpoint chlorination. Utilities can use booster chlorination in the distribution system to increase disinfectant residuals (Woolschlager, Rittmann, Piriou, & Schwartz, 2001) (Valentine, Ozekin, & Vikesland, 1998). However, blending of chlorinated and chloraminated water has been shown to cause unintended breakpoint chlorination, increases in DBP levels, or decreases in disinfectant residuals (Mahmood, Pimblett, Grace, & Utne, 1999) (Muylwyk, Smith, & MacDonald, 1999). Nitrification mitigation techniques such as breakpoint chlorination or temporarily switching from chloramines to free chlorine can result in increased levels of regulated DBPs.

Limiting the nutrient that causes the growth of heterotrophic bacteria is another method for reducing the advancement of biofilms in distribution systems (Trussell, 1999). This is usually the biodegradable fraction of dissolved organic carbon (Cartier, et al., 2009). The presence of phosphorus has shown to limit that growth (Besner, Gauthier, Servais, & Camper, 2002). By limiting the assimilable organic carbon and the biodegradable dissolved organic carbon, bacterial growth is limited (Besner, Gauthier, Servais, & Camper, 2002).

At lime softening plants, increasing lime dosage at the water plant to insure that finished water enters the distribution system at a pH above 8.7 and preferably around 9 is recommended. Since pH is aa secondary standard, a higher pH does not carry an MCL violation concern. A higher pH addresses the issue with nitrification since nitrifying bacteria are most active at a pH between 8.1 and 8.6.

LeChevallier et al (1990) showed that the addition of polyphosphate saw a significant decrease of a hundred percent (Ollos, Huck, & Slawson, 2003). The study also showed that the addition of zinc orthophosphate reduced biofilms viability thirty four percent, and the adjustment of the pH and alkalinity decreased biofilm viability by tenfold (Ollos, Huck, & Slawson, 2003). The addition of a corrosion inhibiter has also shown to reduce coliform bacteria in Seymour, Indiana as well (LeChevallier, Lowry, & Lee, 1990). The use of corrosion inhibitors is good but the design must take care not to over treat. If too much zinc orthophosphate is placed into a distribution system can cause a layer of zinc hydroxide to form on the pipe. This will result in a free chlorine demand with the formation of a bacteria laden zinc floc that forms (LeChevallier, Lowry, & Lee, 1990).

Design of the distribution system is of importance when trying to control biofilms. From a hydrodynamic perspective, transmission networks should be designed and operated to minimize sediment accumulation, microbial adhesion and regrowth. During stagnant flow conditions or during periods of low demand, conditions will be favorable for biofilm establishment. It is precisely in these areas of the system, where nutrients can accumulate, ferrous materials provide fresh surfaces for colonization, and disinfectant residuals can become rapidly depleted.

Routine systematic flushing (annually or semi-annually) is a primary component of proper distribution system maintenance, although it does not prevent re-colonization (Walker & Morales, 1997). The two systematic flushing methods for removal using high velocity flow is conventional and unidirectional flushing programs (Martin & Ries, 2014). Conventional flushing is used when manpower is limited or valves are not able to be isolated (Hiltebrand & Theiss, 2002). Conventional flushing is employed as a maintenance

method or as an emergency action following a customer complaint (Martin & Ries, 2014). In Ann Arbor, Michigan, their flushing program was minimalized until 1989 (Skadsen, 1993).

The use of high velocity water via unidirectional flushing is another method to remove biofilms. The two critical factors when determining the effectiveness of these methods is as follows. The flow velocity must be a minimum of 2 m/sec or greater (Hiltebrand & Theiss, 2002). The objective is to push any residual bacteria that may remain in the distribution system downstream so care must be taken to prevent infection of any a part of the system that currently does not have issues.

Another method for dispersal is the removal by force. This can be caused by a pig in the line scrubbing the outer wall of by increasing the velocity to the point of detachment. The rate of velocity is determined by the thickness of the biofilm. The thicker biofilms tend to sluff off, where the thinner layers tend to erode away (Rayburn, Blaha, Anderson, & Rotert, 2011). It is when thicker layers are sloughed off that the person conducting the removal operations must be careful. Thicker layers tend to be more nutrient rich and may expose dormant nutrients of sluff off something and send it downstream. In a water system in New Haven, Connecticut, the plant manager employed the use of mechanical cleaning through pigging to remove biofilms. The use of these methods only made their problems worse. They are now using higher than normal volumes of chlorine to control biofilms and they have commenced on a more rigorous flushing program. Along with the biofilm regrowth, an increase in coliform levels have been experienced. The cause is attributed to the release of the biofilms during unidirectional flushing and increased free chlorination periods. (LeChevallier, Lowry, & Lee, 1990).

In addition, dead zones (resulting in excessive hydraulic residence times) can be eliminated by valve exercising, and eliminating excess storage, while low flow areas can be eliminated by line resizing (Crozes & Cushing, 2000) or routing flow to fire hydrants. Eboigbodin, Serh, & Biggs, (2008) noted that temperature and velocity also play an important role in biofilm formation. Flushing and/or pigging (use of a water-propelled device) along with valve turning at regular intervals are typically used in practice. Specific subsections of a distribution system can be isolated and flushed with high velocities to discourage biofilm growth. According to (Schrempp, Goold, Kwan, & Sarai, 1994), mechanically cleaning pipelines; draining and cleaning reservoirs; and dead-end, unidirectional and continuous flushing; were not sufficient to control nitrification at one mid-western utility.

The removal of Legionella in biofilms can be done using Copper Silver ionization (Lin, Vidic, Stout, & Yu, 1998). This method has been shown to inhibit the amoeba growth of Legionella in hospital distribution systems (Lin, Vidic, Stout, & Yu, 1998). The concentration levels have to be kept low to no cause any other water borne issues. The manufacture of the ionization equipment recommends a dose between 0.2 and 0.4 mg/L for copper and 0.02 and 0.04 mg/L for silver. These levels are well below the maximum contaminant level for these metals (US Envrionmental Protection Agency , 2002) (Lin, Vidic, Stout, & Yu, 1998).

Niquette *et al*. (2000) demonstrated that proper pipe material selection is important in biofilm control. They noted that PVC and polyethylene had the least potential to form biofilms, although polyethylene may promote the growth of *Legionella* more than PVC (Van Der Kooij & Veenendaal, 2001). Cloete, (2003) reported similar findings,

demonstrating that galvanized iron had the least resistance to biofilm formation among the materials they tested. Asbestos concrete and cast iron showed similar biofilm formation potential. They found PVC to have more resistance than the other materials (Cloete, 2003). Thus, changing pipe materials is another control option.

The best method is to not let them form in the first place. Removing organic matter and nutrients during initial treatment will discourage the growth of biofilms. Proper distribution system maintenance is key to keeping the system clean. The objective is to keep the water in the distribution fresh and not to allow it to stagnate. Proper installation of ductile iron pipe and verifying it isn't knocked around during construction are also key to minimizing biofilms. Effective use of disinfectants and proper dosage to eliminate or reduce the amount of free ammonia that may end up in the distribution system are also important.

Hypothesis

Based on gathering information on water quality, pipe materials, age and distribution system characteristics, the hypothesis is that a tool can be developed to predict the likely locations for nitrification from biofilms will develop. The means to accomplish this tool will be a regression equation based on PCA analysis of water distribution systems with known biofilm/nitrification issues and comparing this data to systems without. The inputs to the regression process will include water quality data from 3 systems, data from water distribution models and filed data.

Methodology

Data Collection

Identifying locations where biofilms are likely to occur is something many utilities could utilize. The methodology for collecting the copious data needed for this analysis is to translate raw data and data files into an ArcGIS format (ESRI). A geographical information system (GIS) is a computer system designed to capture, store, manipulate, analyze, manage, and present all types of spatial or geographical data. The benefit of using GIS for this project is that a GIS system allows us to create data layers these layers permit a query to be performed which will allow investigators to answer specific questions.

Many utilities map their piping systems in geographic information system platforms like ArcGIS®. By doing this, that utility can later find those assets with something as simple as a cellular phone. The modeling software integrates into the ArcMap platform and the two become one. To develop a successful model many steps, need to be taken. The first step is to gather the appropriate data about the system one wishes to study. This data can include structural features, water quality parameters, and operational hydraulic data (Besner, et al., 2001). Important structural features to gather include the pipe diameter, pipe length, Hazen-Williams coefficients and elevations (Besner, et al., 2001) (Elton, Brammer, & Tansley, 1995). Pipe age is also an important characteristic to obtain (Deininger, Clark, Hess, & Bernstam, 1992). Water quality parameters that are important to obtain are the pH, temperature, free and combined chlorine residuals, color, turbidity, water age and any HPC counts that may be present (Besner, et al., 2001) (Elton, Brammer,

& Tansley, 1995). It is important to gather this information of an extended time period as to eliminate any outliers and anomalies from the source data.

The next step in this process is to set all of these parameters into the model. Completing this step is key to providing accurate predictive results. The basic program that almost all water modeling software is based in EPANET (American Water Works Association, 2012) (Deininger, Clark, Hess, & Bernstam, 1992). This program was written for the free use of the public. Possessing the ability to calibrate the model is key to successful use of the model. One must remember that modeling is a tool and not the end all of water quality assurance. The model only uses what is put into it. If the parameters change in the source water, and the parameters of the model are not; then the model will provide bad data. Obtaining quality calibration data is important to the use of computer modeling. Setting up demand nodes and pressure sensing equipment in about thirty to forty percent of the system is imperative to obtaining quality data (Elton, Brammer, & Tansley, 1995).

Visualization of the model is a good tool to provide the ability to observe a trend. One can observe the degradation of water quality and be able to identify the best locations to prevent any catastrophic failures in the distribution system. The majority of modeling software can also provide graphs and charts of trends that are emerging. Past data can be graphed and used to predict future events. Low flow areas can also be easily identified. Solutions can be visualized to give field staff methods to eliminate these problems in the field in real time.

The model can also be used to predict actions that would occur due to specific events. The effect of a water main break on the rest of the system can be simulated. The effects of that main break on pressures in the surrounding water mains can be modeled. The effects of the main break on water quality can be predicted as well. If a neighborhood reduces water usage due to seasonal use, what would happen to the water quality. This information can be used to develop flushing strategies and such to reduce future breakdown of quality water. These predictions can also be used to determine actions at the water treatment plant. If changes are being proposed, their effect can be applied into the model and the resultant effect can be observed in the distribution system before any changes are made (Elton, Brammer, & Tansley, 1995).

Another useful feature of water modeling is the prediction of the results of a well-planned flushing program. The model can develop good flow paths that would be the most efficient for the objective sought (American Water Works Association, 2012). A utility could also use the model to develop a map for constituent tracking. If a foreign agent enters the distribution system, a well configured model can determine when a contaminated water source will arrive in an area (American Water Works Association, 2012). A plan can be developed to mitigate further damage. The model can also be used to figure out where a contaminant entered the system by tracking the flow back to the source.

Statistical Analysis

To analyze the results from the master spreadsheet statistical analysis will be done using Excel® and XLStat ®. Excel® was chosen for its wide acceptance among professionals in scientific scholars first for its flexibility and accuracy in performing

statistical analysis. XLStat® was chosen as it was developed as an add-on tool for Excel spreadsheets.

Descriptive Statistics

Descriptive statistics were performed first on the final data from the master spreadsheet that combined data from all four sources. Before descriptive statistics can be applied to the data, it first must be modified so that there are no zeros within the data set. Any zeros left in the data set will cause some statistical test to either not run or give it incorrect outcomes.

The information and manufacturing economic sectors were removed first from the master spreadsheet because of the lack of data. The next step in cleaning up the data was to remove any utility that presented a zero in any field. The justification behind this was that the PCA and FA processes (discussed shortly) will not run without complete data.

Correlation Matrix

A correlation matrix was used to test dependency among economic activity variables. This test aims to provide insight as to which variables will have the largest impact in be the best contributors in formulating an equation to predict water use based upon economic activity. R values gathered from this help determine which sectors are good predictors.

Regression Model

The regression model was performed in order to estimate the relationship among variables. This technique is used for modeling equations and analyzing several variables at once when the focus is on a relationship between a dependent variable and multiple independent variables. In this case the dependent variable is the total water use and the independent variables are the economic activity. It will be assumed that the data is normally distributed.

PCA and FA Analysis

The factor analysis method dates from Spearman (1904) and continues to develop. Various factor analysis methods are used by XLStat® to reveal the possible existence of underlying factors which give an overview of the information contained in a very large number of measured variables. Factor analysis is the study of using a number of statistical techniques with the goal of representing a large data set with a reduced number of common factors (Kim & Mueller, 1978) (Stedmon & Bro, 2008). To complete this study the first item to determine is to examine the interrelationships of the provided variables. The association of the provided variables must be determined and inspected. The data used has to relate to each other in some manner or the results will be misleading or outright wrong. Once the correlation is established the factor analytic approach could be used to use the measured correlations and create hypothetical variables. Factor analysis is considered an expeditious way of creating a minimal number of factors from observed conditions to relate them to another option to see if the hypothetical relation works in a different scenario. This method can be used to test for a specific hypothesis, which is the intent of this study. For

this study, factor analysis will be used to test a specific expectation of the behavior of water in a distribution system. This is sometimes referred to as a confirmatory factor analysis (Kim & Mueller, 1978).

There are a few important terms to remember when using factor analysis. The square of the factor loading for a specific variable is called the communality. The uniqueness component is one minus the communality. Where the equivalent to the correlations between factors and variables in which a single factor that is common between them is known as factor loading. Factor loading is also observed where multiple factors that are common to each other and are orthogonal to each other (Kim & Mueller, 1978). The factorial complexity refers to the number of factors that possess significant loading on a given variable.

Factor analysis utilizes a series of matrices and determinants to develop the correspondence between the measured observed values and the common factors. The results of the common factors are given in a correlation matrix. The guiding principle to factor analysis is if the number of given factors in a problem is given as the variable K, this can be inferred that the ranking of any corresponding adjusted correlation matrix is also given as the variable K. This principle reinforces that through an inferential process the underlying common factors discovered through the examination of the adjusted correlation matrix (Kim & Mueller, 1978). This observation is how factor analysis was discovered to be a viable method in its early infancy.

To properly use factor analysis, one has to possess the perceived belief that the measured variables provided by the data are a linear combination of some underlying source variable. These underlying factors existence is assumed and are contained in the

observed variables. It is further believed that a correspondence exists and that the use of factor analysis exploits those underlying factors. To reduce uncertainties, like many scientific methods, certain postulates are relied upon. This refers to the basic assumptions that must be adhered to while using factor analysis if errors are to be minimized. Some assumptions may be more appropriate than others depending on the problem. Ultimately all assumptions are subject to doubt. This is why it is key to maintain consistency throughout the entire process. Developing a parsimonious model provides more confidence in the results provided by the factor analysis. Adopting the principle of parsimony generally leads to a single conclusion to where a single conclusion is otherwise impossible. Another theorem that is popular in the use of factor analysis is rank theorem. Rank theorem indicates the correspondence between the number of common factors and the number of dimensions of the adjusted correlations matrix. This can be a useful tool as rank theorem will allow a hypothesis to be rejects of not based on the results.

The objective is exploratory factor analysis is to reduce a set of given variables from a large number to two or three underlying factors. The purpose is confirmatory factor analysis is to utilize two or three underlying factors to create a set of larger variables to confirm the truth of a hypothesis (Kim & Mueller, 1978).

There have been many uses for factor analysis since its discovery. Economists has used factor analysis to generate further study of a series of variables that are uncorrelated when it is discovered that highly correlated variables develop results that may be misleading or inaccurate. In the field of political science, factor analysis has been used to determine attributes from various nations to determine the correlation between wealth and size to classify nations and what factors are applicable in doing such. Psychologists and

educators have used factor analysis to determine stimuli perception and facilitate a method of categorizing human behavior as it relates to language.

Factor analysis has its restrictions. Those include the assumption that one group of common factors exist for a particular data set. When more than one set of common factors are found the exact configuration of those factors cannot be determined without adding additional assumptions to the scenario. The rank theorem only applies when causal operations, the rules for combining factors to create variables, meet a specific set of conditions. If the observed correlations are contaminated with too many outliers sampling and measurement errors can occur. The real-world conditions with or without sampling and measurement errors may not necessarily fit into the factor model as predicted.

Principal Component Analysis (PCA) is popular multivariate technical mainly used to reduce the dimensionality of *p* multi-attributes to two or three dimensions (Bloetscher, Review of Nitrification and Distribution System Water Quality, 2017). PCA is a special case of factor analysis (where k, the number of factors, equals p, the number of variables). While FA assumes a number of factors, PCA is used to reduce the number of variables to factor sets, while maximizing the unchanged variability in order to obtain independent (non-correlated) factors (Bloetscher, Review of Nitrification and Distribution System Water Quality, 2017). PCA uses a multivariate statistical parameter called an eigenvalue, which is a measure of the amount of variation explained by each principal component. PCA summarizes the variation in a correlated multi-attribute to a set of uncorrelated components, each of which is a particular linear combination of the original variables (Pleitez, 2012). PCA is the simplest of the true eigenvector-based multivariate analyses. The benefit of using PCA include:

- Visualizing the correlations between variables to hopefully be able to limit the number of variables to be measured afterwards;
- Obtaining non-correlated factors which are linear combinations of the initial variables so as to use these factors in modeling methods such as linear regression, logistic regression or discriminant analysis.
- Visualizing observations in a 2- or 3-dimensional space in order to identify uniform or atypical groups of observations.

Two methods are commonly used for determining the number of factors to be used for interpreting the results: A Scree Plot is a simple line segment plot that shows the fraction of total variance in the data as explained or represented by each component (IOS, 2012). The number of factors to be kept corresponds to the first turning point found on the curve. However, these representations are only reliable if the sum of the variability percentages associated with the axes of the representation space are sufficiently high. If this percentage is high (for example 80%), the representation can be considered as reliable. If the percentage is reliable, it is recommended to produce representations on several axis pairs in order to validate the interpretation made on the first two factor axes.

The principle objective of principal component analysis is to take a set of collected data points that may be correlated and reduce them to a smaller set of uncorrelated variables whose information will be relatable to the original information (Stedmon & Bro, 2008). This smaller set of uncorrelated variables are easier to decipher and comprehend. Using this smaller set of variables is easier to analyze. An important observation in the use of principal component analysis is that the first two principal components typically make up

for the majority of the total variation. If this is the case approximating the remainder of the distribution can be estimated by plotting the variables.

The principal component analysis technique is similar to that of factor analysis. There are some of the differences between principal component analysis and factor analysis. Principal component analysis does not use and underlying statistical model of the measured variables. Instead principal component analysis looks at rationalizing the total variation in those measured variables based on the maximum variance properties of the principal components. Factor analysis utilizes an underlying statistical model that separates the total variance properties of the principal components (Dunteman, 1989). Principal component analysis finds itself to be similar to a few other statistical methods. Some of those methods include the discriminant analysis and canonical correlation analysis. Both of these methods involve the use of linear combinations of correlated variables in which the weights of that linear combination are obtained by maximizing some sort of statistical property (Dunteman, 1989).

In many instances principal component analysis can be used before factor analysis in order to review the dimensionality of a common factor space (Dunteman, 1989). Principal component analysis can be also used before factor analysis to find a subset of variables from an existing larger set of variables. The objective is to use a set of variables that have high correlation using principal components versus using the principal components as the original variables. Multicollinearity problems can be addresses by using principal component analysis with regression analysis. Since principal component analysis represents the variation of the original variables; it can also be used for displaying multivariate data graphically (Dunteman, 1989).

Through searching for a limited number of uncorrelated linear combinations found using the original variables, principal component analysis is able to capture most of the relevant data from those original variables. To complete this task principal component analysis using mathematical weights that are calculated to provide the maximum variation. This can also be accomplished by squaring the sum of the correlations of the principal components with the original values (Dunteman, 1989). The first few principal components account for a lot of the variation that is found to be present in the original variables. These variables have the highest probability to possess the highest squared multiple correlation with the original value. The physical weights of the variables and their magnitude are used to interpret how well the account for any variability in the data. The magnitude of these variables provide insight into the relative contribution that principal component variable will provide.

When visualizing the first principal components through geometrical principles the first principal components seen is the closest to fit the observed values. This minimizes the squared distance sum of the observation from the line in the variable space that represents the observed component (Dunteman, 1989). As the components progress, the second, third, and so on are the residuals of the first and those first three components form the hyperplane (Dunteman, 1989). The hyperplane is the three-dimensional plane of closest fit (Dunteman, 1989). The number of principal components cannot exceed the number of given variables. The farther one travels from center in the geometric plane, the less relevant the principal component variable becomes.

The variance of a linear composite equation is shown in Equation 7,

Equation 7 Variance of Linear Composite

$$\sum_{i=1}^{p} a_i x_i \; and \; is \; \sum_{i=1}^{p} \sum_{j=1}^{p} a_i a_j \sigma_{ij}$$

Where σ_{ij} is the covariance between the *i*th and *j*th variables (Dunteman, 1989). This representation of a linear composite is pretty straightforward generalization of the variance of two variables, $y = a_1 x_1 + a_2 x_2$ which can also be viewed as $a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2a_1 a_2 \sigma_{12}$. This can be expressed in a simpler form through matrix algebra as $a'Ca$, where a appears as the vector of variable weights and the C shows as the covariance matrix (Dunteman, 1989). In principal component analysis the weight of vector a is found to maximize the $a'Ca$ relationship given the given constraint in Equation 8,

Equation 8 Constraint to Weight Factor

$$\sum_{i=1}^{p} a_i^2 = a'a = 1$$

If this variable is not constrained, the variance can become quite large when the model selects weights. A covariance matrix or a correlation matrix can complete the make-up of a linear composite. In the same fashion, a principal component analysis can also be based on these same matrices (Dunteman, 1989). The key to preventing large weights is to verify that all of the measured variables are the same dimension or measurement. The reason is the software will not be able to differentiate the size differences and the solution will be inaccurate. Verification of the dimensions used in all measurements and the ensuring that the dimensions work with the other variables is paramount to the success of any principal component analysis. Additionally, dependencies must exist in the original variables of there will be just as many principal components as variables and the analysis will not work. When the linear dependencies exist then the dimensionality of those

principal components will be reduced proportionately. It is the goal of principal component analysis to determine the linear transformations of the same observed variables (Dunteman, 1989). Another point to the variance of a principal component analysis is that sum of all of the variances of the principal components must be equal to the sum of the variances of the original variables (Dunteman, 1989). If this statement is found to be true in this case, then the majority of the information in the measured variables will be captured in the first for principal components. Remember though, the principal components are uncorrelated and must also make an independent contribution to the variance in regards to the measured variables.

Some things to watch in principal component analysis come from some notable people who helped fine tune principal component analysis as a viable solution for reducing measured components into useable data. Henry F Kaiser in 1960 removing the principal components with the largest latent roots less than one from the correlation matrix (Kaiser, 1960). He further opined that principal components with a value less than one contain less information that the single standard value in which the value is one (Kaiser, 1960). At issues is the concern that principal components could be arbitrarily discarded due to their value by may contain important information not otherwise seen as the early stages. It has been suggested by Ian Jolliffe that Kaisers rule is to strict and removes too much valuable information from the analysis (Jolliffe, 1972). Jolliffe's idea was to lower the cutoff to .7 instead of 1. His thought process was that too much information was being discarded and the fact that the latent root was over 1 would result in a higher sampling error. By lowering the cutoff of the latent roots to .7 these errors were shown to be reduced or eliminated.

An important point to remember is that all of these rules are arbitrary in nature. If they are used, they should be used with caution. An example of this is if one uses Jolliffe's claim that the cut off should be .7, then twice as many principal components would be retain then is Kaisar's rule of the principal component less than 1. The general take away from any of these rules is to review the data and determine if more of less variables are to remain. It is also an observation that smaller principal components are more difficult to work with than larger principal components.

Another approach to principal component analysis is using the geometric properties of the variables and plot them in one, two, or three-dimensional space. By plotting these quadrants, one is able to visualize where the vector resides in space in relation to other vectors. This visualization may be providing the opportunity to see connections to the other variables that may have otherwise gone unobserved. Using scatter plots, a researcher can observe a correlation that was otherwise unnoticed. Using this method is a valuable tool in the researcher's tool box for finding an effective and useful solution for their work. Another point of observation is that the first two principal components always form a forty-five-degree angle. The direction cosine of the principal components is the cosine of the coordinate axis forms with the specific variable coordinate axis. There are many direction cosines for a specific principal component as there are observed variables or the number of elements in each latent vector (Dunteman, 1989). The objective in using this method is to identify the largest principal component and ensure them for a forty-five-degree angle. Anything outside of that area is not a relevant solution. This is a good method when there are two principal components.

If there are three or more principal components, it is key to examine the latent vectors and identify the largest principal component (Dunteman, 1989). There is generally one that is largest than the rest and differ from the others. If using a correlation matrix, the correlations rarely equal to one another. If the correlation matrices are similar to each other than the largest principal component will also be similar in its magnitude. The principle component is directed in such a manner that it faces the variable with the largest variance. This maximizes the variation of the largest principal component. There is a rare instance in reality that one or more of the latent roots are equal to exactly zero or that two or more latent roots equal each other. There are instances where the roots are close to zero or each other and in these cases the correlation matrix of principal components are considered unstable (Dunteman, 1989). It is important to plot latent vectors whether it is in the standard Cartesian plane or in three dimensions. One of the reasons is one can visual any outliers that may exist in a set of principle components. Clusters can also be searched for and found in visualized data as well.

Varimax rotation is a method used in principal component analysis similar to that done in factor analysis. Varimax rotation is one of the more popular orthogonal rotations procedures (Dunteman, 1989). This procedure is an orthogonal rotation of the principal component axes find themselves to be perpendicular to the new axes. In a varimax rotations scenario the original coordinate axes are rotated to maximize what is called the varimax criterion (Dunteman, 1989). This solution is a unique solution much like principal component solution. The method rotates the original coordinate axes to a new coordinate axes which include small or large loading of the variables on it (Dunteman, 1989). This equates to an orthogonal transformation of the original principal components. The

reasoning behind rotating the principal components is to make them easier to identify and research. This rotation will cause some variable to be reduced or eliminated and thus easier to find a solution to a difficult problem.

Principal component analysis and factor analysis share a common quality of reducing the dimensionality of a given or observed set of variables to make them simpler to analyze. There are some important differences that must be recognized so the correct system is used. The largest items to consider is that principal component analysis decomposes the total variance of the variables. Factor analysis works to find a decomposition of a reduced correlation matrix. It forms a diagonal matrix that contain unique variables that are closely associated with the variables given (Dunteman, 1989). An important note in the use of principal component analysis is that this procedure does not look at the underlying model and only decomposes the correlation matrix. This said the principal component analysis does not know the difference between a common variance and a unique variance. Factor analysis does not have this issue as it does possess an underlying model. Factor analysis relies on a series of assumptions that are established early in the process. The most important assumption is that the *i*th variable in a set of variables, x_i, can be shown to be nothing more than a linear combination of hypothetical unobserved common factors (Dunteman, 1989). The *i*th variable is a unique factor to that specific variable. The important observation is that factor analysis provides the x_i set of variables, shown as a linear function of unobserved common factors. The principal component analysis expresses the principal component as a linear function of x_i. The model used in a principal component analysis assumes that the standardized variable x_i can be express as shown in Equation 9.

Equation 9 Factor Analysis Model Equation

$$\sum_{j=1}^{k} l_{ij} f_j + u_1$$

Where i_j is where the loading of the *i*th variables occurs, f_j is the factor score for the *j*th factor, and u_i is a unique factor score for the *i*th variable (Dunteman, 1989). To make factor analysis work properly a number of assumptions have to be made. Namely that the mean, the correlations between common factors, and the correlation between common factors and unique factors are all zero. The other assumption is that the variance must equal one.

On the other hand, there are typically two ways that a principal component analysis is used to estimate the loading of the factors in a factor analysis model. The first is called the iterative principal component analysis. One starts with a principal component analysis of a correlation matrix. The values are examined with assumptions and considerations considered, to locate the factors accounting for the correlations within the observed values. In this method Kaiser's criterion of only retaining components that contain a value greater than one is generally used in this application. The other method utilizes a standard goodness of fit test that could look at how well a given set of variables fit in the correlation matrix. In principal component analysis the number of factors that are common to each other a calculated and the initial communality estimate for each term is determined by using the summing the square of the loading for the retained components (Dunteman, 1989). To process this the computation of the sum of squares is computed from the initial communality. Those estimates are then inserted into the principal diagonal for each variable substituting the original. A new reduced correlation matrix is formed and the

second principal component analysis is conducted (Dunteman, 1989). This process is repeated, reinserting the new sums in the correlation matrix until all of the forms converge.

Another method employed by some computer software packages is the maximum likelihood factor analysis model. In the maximum likelihood factor analysis, the variables are assumed to be part of a multivariate normal distributions in which the principal component factor analysis does not require any distributional assumptions. The computed matrices generated possess their estimated parameters that were generated by the original observed variables. In the maximum likelihood factor analysis, the user is required to provide the number of factors to be extracted. The results of using this method have found similar solutions to the standard principal component analysis.

Fortunately, this process is conducted by computers and is not a manual operation. An extensive knowledge of mathematics is needed to conduct a principal component analysis as mentioned earlier. To conduct these processes manually would take hours if not day to complete.

A final note on principal component analysis. The most important point in using this or factor analysis, is that the data used be valid and highly correlated. If the measured variables are not correlated, there will be as many principal components as there are variables. Principal component analysis can be used to reduce the dimensionality of an observed data set. This reduction of variables makes it simpler to analyze and form a relatable hypothesis and solution to the question presented.

Discriminant function analysis (DA) is also closely related to PCA in that they both look for linear combinations of variables to explain the data. Using these techniques, the factors affecting each component can be determined, allowing the investigator to evaluate

common and differ results by the factors most affecting the results. A biplot display of both factors and factor loadings is an effective way of studying the relationships between variables, and the interrelationship between observations and the variables (Fernandez, 2010). The correlations among the multivariate attributes used in the DA analysis are revealed by the angles between any two DA loading vectors. For each variable, a DA load vector is created by connecting the *X*-*Y* origin and the multiplied value of F1 and F2 loadings in the biplot (Pleitez, 2012). The angles between any two variable vectors will be:

1. Narrower (<45°) if the correlations between these two attributes are positive and larger
2. Wider (around 90°) if the correlation is not significant
3. Closer to 180° (>135°) if the correlations between these two attributes are negative and stronger

The membership of each factor in question can be grouped using Mahalanobis distances. In general, the Mahalanobis distance is a measure of distance between two points in the space defined by two or more correlated variables (Pleitez, 2012). Using Mahalanobis distances for each variable, the location of the point that represents the means for variables in the multivariate space defined by the variables in the model can be determined. These points are called group centroids (XL Stat, 2017). For each case, the Mahalanobis distances from each of the group centroids can be computed. Since the probability that a case belongs to a particular group is proportional to the Mahalanobis distance from the group's centroid, each case is classified as belonging to the group to which it is closest, that is, where the Mahalanobis distance is smallest (Pleitez, 2012).

Linear Regression

Ultimately the goal is to determine if the condition has a consequence – i.e. the potential for failure. If so, one needs to know what that consequence is – in this case

operation or inactive. The values were assigned for operational (1) or inactive (0) of aquifer storage units in the United States, as the dichotomous dependent variable. The impact of these factors can be developed via a linear regression model. The model would be developed as follows (Bloetscher, Infrastructure Management, 2019):

Equation 10 Condition Index

$$CI = w_1C_1+w_2C_2+w_3C_3+w_4C_4+\ldots w_iC_i$$

where:

- CI = Condition index
- w = weighting factor
- C is condition factor

If one knows the consequence, the weights can be found:

$$f(x) = c_1x_1 + c_2x_2 + c_3x_3 + .. + .c_nx_n$$

where the values of c_n are real numbers and

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

are the factors which are a compilation of the original variable to maximize variance. It assumes these constraints and linear variables in the matrices are non-negative. If there are negative values, they must be made positive as follows

$$x_i^+ = \begin{cases} x_i \text{ if } x_i \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

$$x_i^- = \begin{cases} -x_i \text{ if } x_i > 0 \\ 0 \text{ otherwise} \end{cases}$$

the linear regression model provides a mechanism to model the data to determine if differences between the active and inactive projects exists.

Development of Model Data

To develop a factor analysis model there are four basic steps. The first step is to gather the data that is to be used and prepare the data for the relevant covariance matrix (Chowdhury, Campagne, & McLellan, 2010). This data can be collected in a logical manner that will allow the results to be reasonably acceptable for use. The data must be arranged in a systematic way to form a data matrix.

The next step is to use the computer software to extract the initial common factors from the collected data. The objective of this step is to find the number of factors needed that can adequately explain the observed correlations. There are different ways the software can do this. The first is the maximum likelihood method, the next is the least-squares method, additionally alpha factoring can be used, or image factoring, and lastly principal component analysis. The XLStat software used for this project uses the principal component analysis to complete this step.

The third step in this process is to rotate the factors to a terminal solution to make interpretation easier. There are certain restrictions that are generally used in this step of the process. They include the verification that there are common k factors. The underlying factors must be orthogonal to each other. The first factor discovered typically accounts for much of the variance in the problem. The second factor makes up most of the residual variance that is left unexplained by the first factor. This process continues until all of the factors have been analyzed.

The last step is to construct factor scales to be used for further analysis. The reason most users of factor analysis enjoys this method is to develop a factor structure among a set of observed or measured conditions that is reduced. Obtaining factor scales to use on other studies is the way to complete this objective. These scales can be used to analyze other data that is relevant to the first dataset. The computer software used create weights that are combined with the observed values to develop the underlying factor and factor scales. Using factor scales finds that the underlying factors are highly correlated with each other but may not fit the hypothetical factor. (Kim & Mueller, 1978).

Choosing the factors that fit in your model is key to success. In previous experiments to predict trihalomethanes using factor analysis, items like pH, Temperature, bromide, chlorine dose were all used to attempt to provide accurate factors (Rodrigues, Esteves da Silva, & Antunes, 2007). Ensuring that all of the data correlates is key to the success of any factor analysis (Rodrigues, Esteves da Silva, & Antunes, 2007). One thing observed in the Rodrigues study is that due to environmental changes that cannot be factored in, the factor analysis completed by his team in Portugal cannot be used to predict future trihalomethane problems (Rodrigues, Esteves da Silva, & Antunes, 2007).

The most important part of using factor analysis is to complete the fit test. That is, the observed correlations and the variables produced by the factor analysis model need fit. To complete this task, the x-squared method is employed. If the data fits, when it is squared there should not be a statistically significant error in the results. If there is, the factors must be rejected.

Results and Discussion

Data was gathered from four water distribution systems from 4 water treatment plants. For the 4th system, the data was incomplete. Initial analysis was performed. The system operates with a higher pH that the other three systems, does not quench, then add ammonia for chloramination, maintains a combined chlorine residual between 0.2 and 05 mg/L. No biofilms have been detected in this system despite four types of pipe. As a result, it was not useful in the analysis and was deleted. The others were analysis using PCA analysis, the discriminant analysis and finally via linear regression. The variables analyzed:

Water Plant
AC PIPE
CAST PIPE
CNC PIPE
DIP
PVC
GI
DIAMETER
ECOLI E. coli by Colilert
FRESCG-FLD Free Residual Chlorine-Field
NH3H_FREE_CALC-LAB Free Ammonia _Lab
Calc.
PH-FLD pH-Field
TCOLIB Total Coliform
TEMP-FLD Temp-Field
TRESCG-FLD Total Chlorine Residual-Field
D-O2-FLD Dissolved Oxygen-Field
NO2 Nitrite, as N
NO3 Nitrite, as N
NO3NO2 Nitrate+Nitrite as N
TOC Total Organic Carbon
HPC35 HPC (R2A Agar at 35c)

NH3 Ammonia
NH3H-FLD Ammonia Hach Test-Field §

Table 1 outlines the summary statistics for the variables analyzed. There were 16,030 observations at the three plants over a period of 3 years. Pipe diameters were from 1.5 to 42 inches. pH varied from 8.2 to 9.1.

Table 1 Summary Statistics for Variables Analyzed

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Water_Plant	16030	0	16030	2.000	9.000	5.488	2.968
AC	16030	0	16030	0.000	1.000	0.025	0.156
CAS	16030	0	16030	0.000	1.000	0.007	0.085
CNC	16030	0	16030	0.000	1.000	0.007	0.082
DIP	16030	0	16030	0.000	1.000	0.290	0.454
PVC	16030	0	16030	0.000	1.000	0.671	0.470
GI	16030	0	16030	0.000	0.000	0.000	0.000
DIAMETER	16030	0	16030	1.500	42.000	10.600	6.693
ECOLI E. coli by Colilert	16030	0	16030	0.000	1.000	0.001	0.025
FRESCG-FLD Free Residual Chlorine-Field	16030	0	16030	0.000	5.300	0.178	0.678
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	16030	0	16030	0.000	0.660	0.027	0.044
PH-FLD pH-Field	16030	0	16030	0.000	9.180	8.268	0.473
TCOLIB Total Coliform	16030	0	16030	0.000	1.000	0.014	0.116
TEMP-FLD Temp-Field	16030	0	16030	0.000	60.000	26.371	3.318
TRESCG-FLD Total Chlorine Residual-Field	16030	0	16030	0.000	6.400	3.172	0.969
D-O2-FLD Dissolved Oxygen-Field	16030	0	16030	0.000	10.500	2.725	3.758
NO2 Nitrite, as N	16030	0	16030	0.000	0.825	0.023	0.069
NO3 Nitrite, as N	16030	0	16030	0.000	1.110	0.034	0.077
NO3NO2 Nitrate+Nitrite as N	16030	0	16030	0.000	1.420	0.031	0.100
TOC Total Organic Carbon	16030	0	16030	0.000	9.500	0.073	0.599
HPC35 HPC (R2A Agar at 35c)	16030	0	16030	0.000	5700.000	55.398	381.600
NH3 Ammonia	16030	0	16030	0.000	1.180	0.208	0.281
NH3H-FLD Ammonia Hach Test-Field §	16030	0	16030	0.000	4.300	0.085	0.271

Table 2 outlines the correlation between variables. The galvanized pipe variable was dropped by the analysis as there were no observations in the remaining 3 systems. Significant correlations (P>0.05) are highlighted in black. Major correlations (P>0.05 and correlation > 0.5) include:

- Total ammonia and Free ammonia
- Free ammonia and dissolved oxygen
- Dissolved oxygen and total ammonia

otherwise there were no strong correlations. The Bartlett's analysis also shows there is no correlation significantly different from 0 between variables Table 3. The Scree plot in Figure 1 shows that to get to 70 percent of variance within the samples 12 factors are required. The preference is that the factors be limited to 3 or 4. The larger figure indicates that variability is strewn throughout the 16,030 samples.

While the goal was to reduce the 23 variables to 3 or 4, 12 is still an improvement. Using the eigenvectors (table 5), the factor loadings were created in XLStat (Table 5).

Table 2 Correlation among Variables

Variables	Water_Plant	AC	CAS	CNC	DIP	PVC	DIAMETER	ECOLI E. coli by Colilert	FRESCG-FLD Free Residual Chlorine-Field	NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	PH-FLD pH-Field	TCOLIB Total Coliform	TEMP-FLD Temp-Field	TRESCG-FLD Total Chlorine Residual-Field	D-O2-FLD Dissolved Oxygen-Field	NO2 Nitrite, as N	NO3 Nitrite, as N	NO3NO2 Nitrate+Nitrite as N	TOC Total Organic Carbon	HPC35 HPC (R2A Agar at 35c)	NH3 Ammonia	NH3H-FLD Ammonia Hach Test-Field §
Water_Plant	**1**	**-0.186**	**-0.094**	**-0.069**	**0.043**	**0.049**	-0.009	**0.019**	0.015	-0.012	**-0.051**	**0.044**	**-0.025**	**-0.097**	**-0.023**	**0.098**	-0.007	**0.017**	**0.031**	**0.056**	0.000	-0.001
AC	**-0.186**	**1**	-0.014	-0.013	**-0.102**	**-0.228**	**-0.090**	0.012	0.014	**0.119**	**0.124**	-0.008	**-0.025**	**0.134**	**0.016**	**-0.030**	**0.027**	0.007	0.002	-0.014	**0.037**	-0.010
CAS	**-0.094**	-0.014	**1**	-0.007	**-0.055**	**-0.122**	**0.116**	-0.002	-0.006	**0.056**	**0.069**	-0.010	**0.031**	-0.015	0.006	**0.027**	**0.041**	**0.031**	0.005	-0.009	0.002	-0.007
CNC	**-0.069**	-0.013	-0.007	**1**	**-0.053**	**-0.118**	**0.388**	-0.002	0.008	-0.007	-0.011	-0.010	-0.015	**0.057**	**0.025**	**-0.020**	-0.015	**-0.018**	-0.007	-0.010	0.015	-0.012
DIP	**0.043**	**-0.102**	**-0.055**	**-0.053**	**1**	**-0.913**	**0.229**	0.006	**-0.021**	**-0.021**	**-0.031**	-0.006	0.002	**-0.053**	0.004	0.014	**0.052**	**0.027**	-0.002	0.007	-0.013	0.009
PVC	**0.049**	**-0.228**	**-0.122**	**-0.118**	**-0.913**	**1**	**-0.280**	-0.009	0.015	**-0.028**	**-0.022**	0.012	0.003	-0.001	-0.014	-0.005	**-0.064**	**-0.031**	0.001	0.001	-0.003	-0.002
GI																						
DIAMETER	-0.009	**-0.090**	**0.116**	**0.388**	**0.229**	**-0.280**	**1**	-0.002	0.010	**0.028**	**0.042**	**-0.021**	-0.003	**0.045**	**0.016**	-0.013	**0.039**	0.005	0.013	-0.011	0.007	**-0.023**
ECOLI E. coli by Colilert	**0.019**	0.012	-0.002	-0.002	0.006	-0.009	-0.002	**1**	-0.007	0.002	-0.004	**0.062**	0.007	**-0.025**	0.010	0.001	-0.001	0.006	-0.003	-0.002	0.007	-0.008
FRESCG-FLD Free Residual Chlorine-Field	0.015	0.014	-0.006	0.008	**-0.021**	0.015	0.010	-0.007	**1**	**-0.145**	-0.011	**-0.022**	**0.048**	**0.017**	**-0.167**	**-0.080**	**-0.094**	**-0.067**	**-0.032**	**-0.031**	**-0.173**	**-0.071**
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	-0.012	**0.119**	**0.056**	-0.007	**-0.021**	**-0.028**	**0.028**	0.002	**-0.145**	**1**	**0.202**	**-0.016**	**0.093**	**0.205**	**0.711**	**0.166**	**0.297**	**0.188**	**0.193**	**-0.047**	**0.857**	**-0.189**
PH-FLD pH-Field	**-0.051**	**0.124**	**0.069**	-0.011	**-0.031**	**-0.022**	**0.042**	-0.004	-0.011	**0.202**	**1**	**-0.037**	**0.323**	**0.305**	**0.045**	**-0.074**	**-0.045**	**-0.029**	**0.028**	**-0.067**	**0.093**	**0.017**
TCOLIB Total Coliform	**0.044**	-0.008	-0.010	-0.010	-0.006	0.012	**-0.021**	**0.062**	**-0.022**	**-0.016**	**-0.037**	**1**	-0.006	**-0.106**	-0.003	**0.020**	-0.001	-0.006	**0.016**	**0.056**	-0.012	**-0.016**
TEMP-FLD Temp-Field	**-0.025**	**-0.025**	**0.031**	-0.015	0.002	0.003	-0.003	0.007	**0.048**	**0.093**	**0.323**	-0.006	**1**	**0.020**	**-0.017**	0.004	-0.005	**-0.033**	**0.048**	0.014	**-0.024**	0.006
TRESCG-FLD Total Chlorine Residual-Field	**-0.097**	**0.134**	-0.015	**0.057**	**-0.053**	-0.001	**0.045**	**-0.025**	**0.017**	**0.205**	**0.305**	**-0.106**	**0.020**	**1**	**0.092**	**-0.360**	**-0.204**	**-0.220**	-0.002	**-0.255**	**0.165**	**0.042**
D-O2-FLD Dissolved Oxygen-Field	**-0.023**	**0.016**	0.006	**0.025**	0.004	-0.014	**0.016**	0.010	**-0.167**	**0.711**	**0.045**	-0.003	**-0.017**	**0.092**	**1**	**0.379**	**0.557**	**0.408**	**0.128**	-0.008	**0.879**	**-0.227**
NO2 Nitrite, as N	**0.098**	**-0.030**	**0.027**	**-0.020**	0.014	-0.005	-0.013	0.001	**-0.080**	**0.166**	**-0.074**	**0.020**	0.004	**-0.360**	**0.379**	**1**	**0.554**	**0.526**	**0.145**	**0.143**	**0.264**	**-0.105**
NO3 Nitrite, as N	-0.007	**0.027**	**0.041**	-0.015	**0.052**	**-0.064**	**0.039**	-0.001	**-0.094**	**0.297**	**-0.045**	-0.001	-0.005	**-0.204**	**0.557**	**0.554**	**1**	**0.721**	**0.068**	**0.086**	**0.416**	**-0.136**
NO3NO2 Nitrate+Nitrite as N	**0.017**	0.007	**0.031**	**-0.018**	**0.027**	**-0.031**	0.005	0.006	**-0.067**	**0.188**	**-0.029**	-0.006	**-0.033**	**-0.220**	**0.408**	**0.526**	**0.721**	**1**	**-0.038**	**0.035**	**0.283**	**-0.098**
TOC Total Organic Carbon	**0.031**	0.002	0.005	-0.007	-0.002	0.001	0.013	-0.003	**-0.032**	**0.193**	**0.028**	**0.016**	**0.048**	-0.002	**0.128**	**0.145**	**0.068**	**-0.038**	**1**	**0.020**	**0.151**	**-0.038**
HPC35 HPC (R2A Agar at 35c)	**0.056**	-0.014	-0.009	-0.010	0.007	0.001	-0.011	-0.002	**-0.031**	**-0.047**	**-0.067**	**0.056**	0.014	**-0.255**	-0.008	**0.143**	**0.086**	**0.035**	**0.020**	**1**	**-0.048**	-0.009
NH3 Ammonia	0.000	**0.037**	0.002	0.015	-0.013	-0.003	0.007	0.007	**-0.173**	**0.857**	**0.093**	-0.012	**-0.024**	**0.165**	**0.879**	**0.264**	**0.416**	**0.283**	**0.151**	**-0.048**	**1**	**-0.232**
NH3H-FLD Ammonia Hach Test-Field §	-0.001	-0.010	-0.007	-0.012	0.009	-0.002	**-0.023**	-0.008	**-0.071**	**-0.189**	**0.017**	**-0.016**	0.006	**0.042**	**-0.227**	**-0.105**	**-0.136**	**-0.098**	**-0.038**	-0.009	**-0.232**	**1**

Values in bold are different from 0 with a significance level alpha=0.05

Table 3 Bartlett's Sphericity Test

Chi-square (Observed value)	-Inf
Chi-square (Critical value)	291.102
DF	253
alpha	0.05

Test interpretation:

H0: There is no correlation significantly different from 0 between the variables.

Ha: At least one of the correlations between the variables is significantly different from 0.

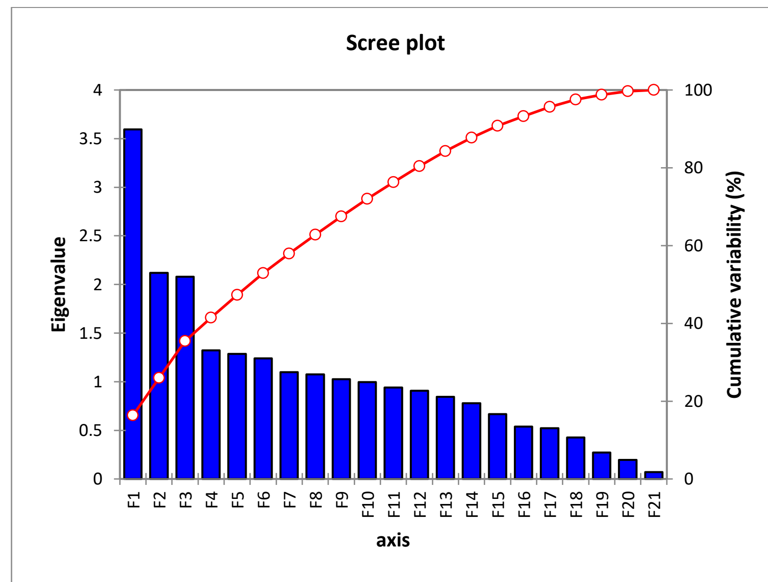


Figure 1 Scree Plot of PCA Factor Contribution to Variance Among Factors

Table 4 Eigenvectors from PCA Analysis

Eigenvectors:	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Water_Plant	0.005	-0.149	0.073	0.015	-0.312	0.461	-0.233	0.044	0.135	-0.072	0.008	0.034
AC	0.030	0.164	-0.089	-0.297	0.172	-0.463	0.352	0.135	-0.117	-0.294	0.096	0.054
CAS	0.027	0.018	-0.090	0.058	0.357	-0.022	0.119	-0.137	-0.086	0.818	-0.007	-0.027
CNC	0.003	0.036	-0.187	0.635	0.201	-0.012	0.143	-0.013	0.019	-0.327	0.023	0.026
DIP	0.022	-0.195	-0.589	-0.236	-0.227	0.093	-0.118	0.017	0.017	0.019	-0.035	-0.011
PVC	-0.036	0.124	0.647	0.205	0.063	0.069	-0.049	-0.034	0.035	-0.011	-0.001	-0.007
GI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DIAMETER	0.021	-0.025	-0.372	0.512	0.169	0.138	0.013	-0.032	0.030	-0.025	0.006	0.013
ECOLI E. coli by Colilert	0.005	-0.018	-0.002	-0.034	-0.048	0.065	0.366	0.115	0.679	0.073	0.482	-0.375
FRESCG-FLD Free Residual Chlorine-Field	-0.113	-0.013	0.003	0.024	0.163	0.014	-0.179	0.723	-0.048	0.011	0.131	0.147
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	0.395	0.318	-0.034	-0.015	-0.127	0.094	0.075	0.000	-0.051	0.045	-0.057	-0.065
PH-FLD pH-Field	0.036	0.347	-0.088	-0.231	0.393	0.263	-0.068	-0.054	0.128	-0.126	-0.090	0.020
TCOLIB Total Coliform	0.000	-0.089	0.043	-0.029	-0.103	0.163	0.554	0.062	0.272	0.043	-0.398	0.621
TEMP-FLD Temp-Field	0.006	0.136	-0.030	-0.247	0.451	0.482	-0.061	0.051	0.064	-0.120	-0.083	-0.039
TRESCG-FLD Total Chlorine Residual-Field	-0.027	0.512	-0.110	0.031	-0.073	-0.091	-0.148	-0.026	0.095	-0.059	0.013	0.061
D-O2-FLD Dissolved Oxygen-Field	0.470	0.128	0.008	0.058	-0.127	-0.017	-0.008	-0.004	0.025	0.008	-0.063	-0.041
NO2 Nitrite, as N	0.312	-0.334	0.093	-0.041	0.172	0.056	-0.027	-0.019	-0.048	-0.067	0.156	0.093
NO3 Nitrite, as N	0.397	-0.240	0.025	-0.030	0.192	-0.118	-0.109	-0.011	0.081	-0.062	0.044	0.089
NO3NO2 Nitrate+Nitrite as N	0.335	-0.281	0.053	-0.039	0.241	-0.170	-0.198	-0.014	0.191	-0.070	0.022	0.127
TOC Total Organic Carbon	0.106	0.059	0.006	-0.023	-0.086	0.328	0.257	-0.013	-0.482	0.008	0.629	0.273
HPC35 HPC (R2A Agar at 35c)	0.022	-0.244	0.059	-0.059	0.081	0.191	0.372	-0.051	-0.310	-0.206	-0.285	-0.532
NH3 Ammonia	0.449	0.239	-0.001	0.055	-0.208	0.024	0.022	0.001	-0.006	0.034	-0.075	-0.065
NH3H-FLD Ammonia Hach Test-Field §	-0.155	-0.020	-0.015	-0.106	0.070	-0.024	-0.075	-0.640	0.115	-0.180	0.221	0.200

Table 5 Factor Loading

Factor loadings:	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Water_Plant	0.009	-0.216	0.106	0.017	-0.354	0.513	-0.244	0.046	0.137	-0.072	0.008	0.032
AC	0.057	0.238	-0.128	-0.342	0.195	-0.515	0.369	0.140	-0.119	-0.293	0.093	0.052
CAS	0.052	0.026	-0.130	0.066	0.405	-0.025	0.124	-0.142	-0.087	0.816	-0.007	-0.025
CNC	0.005	0.053	-0.269	0.730	0.227	-0.013	0.150	-0.013	0.020	-0.326	0.022	0.025
DIP	0.041	-0.284	-0.849	-0.271	-0.257	0.104	-0.124	0.018	0.017	0.019	-0.034	-0.011
PVC	-0.069	0.181	0.933	0.236	0.071	0.077	-0.051	-0.036	0.035	-0.011	-0.001	-0.007
GI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DIAMETER	0.040	-0.036	-0.537	0.589	0.192	0.154	0.014	-0.033	0.030	-0.025	0.005	0.012
ECOLI E. coli by Colilert	0.010	-0.026	-0.002	-0.039	-0.055	0.072	0.383	0.119	0.688	0.072	0.467	-0.357
FRESCG-FLD Free Residual Chlorine-Field	-0.214	-0.019	0.005	0.028	0.185	0.016	-0.187	0.749	-0.048	0.011	0.127	0.140
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	0.749	0.463	-0.050	-0.018	-0.144	0.105	0.078	0.000	-0.051	0.045	-0.055	-0.062
PH-FLD pH-Field	0.068	0.505	-0.127	-0.265	0.446	0.293	-0.071	-0.056	0.129	-0.126	-0.087	0.019
TCOLIB Total Coliform	-0.001	-0.129	0.062	-0.033	-0.117	0.182	0.580	0.065	0.275	0.043	-0.386	0.591
TEMP-FLD Temp-Field	0.012	0.198	-0.044	-0.284	0.511	0.537	-0.064	0.053	0.065	-0.119	-0.081	-0.038
TRESCG-FLD Total Chlorine Residual-Field	-0.050	0.746	-0.159	0.035	-0.083	-0.101	-0.155	-0.027	0.097	-0.059	0.013	0.058
D-O2-FLD Dissolved Oxygen-Field	0.891	0.186	0.011	0.067	-0.144	-0.019	-0.009	-0.004	0.025	0.008	-0.061	-0.039
NO2 Nitrite, as N	0.591	-0.486	0.134	-0.047	0.195	0.063	-0.028	-0.019	-0.048	-0.067	0.152	0.088
NO3 Nitrite, as N	0.753	-0.350	0.036	-0.035	0.218	-0.131	-0.114	-0.011	0.082	-0.062	0.043	0.084
NO3NO2 Nitrate+Nitrite as N	0.636	-0.409	0.077	-0.045	0.273	-0.189	-0.208	-0.014	0.193	-0.070	0.022	0.121
TOC Total Organic Carbon	0.201	0.086	0.008	-0.027	-0.097	0.365	0.269	-0.014	-0.489	0.008	0.610	0.260
HPC35 HPC (R2A Agar at 35c)	0.042	-0.355	0.085	-0.068	0.092	0.213	0.390	-0.053	-0.314	-0.205	-0.276	-0.507
NH3 Ammonia	0.851	0.348	-0.002	0.064	-0.236	0.027	0.023	0.001	-0.006	0.034	-0.073	-0.062
NH3H-FLD Ammonia Hach Test-Field §	-0.294	-0.029	-0.021	-0.121	0.080	-0.026	-0.078	-0.663	0.117	-0.180	0.215	0.191

Table 5 indicates that the major factors defining each factor are as follows:

- Factor 1 – Nitrogen species, including ammonia, and dissolved oxygen.
- Factor 2 – pH and chlorine residual
- Factor 3 – ductile iron pipe (negative impact)
- Factor 4 – CNC pipe (positive)
- Factor 5 – Temperature, CAS pipe and pH
- Factor 6 – water plant, AC pipe and water temperature
- Factor 7 – coliform presence
- Factor 8 – Free Chlorine residual

- Factor 9 – E Coli and TOC (Negative)
- Factor 10 – CAS pipe
- Factor 11 – TOC
- Factor 12 – HPC (Negative)

Table 6 outlines the correlations among variables in the factors which confirms the factor loadings described above. Correlations for F1 and F2 are shown in Figure 2. Those variables within a total of 45 degrees should be indicated as correlated. The longer the line to the data point, the stronger the correlations. The graph shows nitrogen species grouped within 45 degrees with longlines in two places, but all on the positive side of the graphic. This confirms F1 and F2. Similar plots were created for the first 4 factors and all confirm the correlation matrix (see Figure 3-7). Based on the results from Table 7, Table 8 shows the percentage of variability found in each factor as derived from each variable, which also confirms the bullets above.

Table 6 Factor Loading Correlations

Correlations between variables and factors:												
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Water_Plant	0.009	-0.216	0.106	0.017	-0.354	0.513	-0.244	0.046	0.137	-0.072	0.008	0.032
AC	0.057	0.238	-0.128	-0.342	0.195	-0.515	0.369	0.140	-0.119	-0.293	0.093	0.052
CAS	0.052	0.026	-0.130	0.066	0.405	-0.025	0.124	-0.142	-0.087	0.816	-0.007	-0.025
CNC	0.005	0.053	-0.269	0.730	0.227	-0.013	0.150	-0.013	0.020	-0.326	0.022	0.025
DIP	0.041	-0.284	-0.849	-0.271	-0.257	0.104	-0.124	0.018	0.017	0.019	-0.034	-0.011
PVC	-0.069	0.181	0.933	0.236	0.071	0.077	-0.051	-0.036	0.035	-0.011	-0.001	-0.007
GI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DIAMETER	0.040	-0.036	-0.537	0.589	0.192	0.154	0.014	-0.033	0.030	-0.025	0.005	0.012
ECOLI E. coli by Colilert	0.010	-0.026	-0.002	-0.039	-0.055	0.072	0.383	0.119	0.688	0.072	0.467	-0.357
FRESCG-FLD Free Residual Chlorine-Field	-0.214	-0.019	0.005	0.028	0.185	0.016	-0.187	0.749	-0.048	0.011	0.127	0.140
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	0.749	0.463	-0.050	-0.018	-0.144	0.105	0.078	0.000	-0.051	0.045	-0.055	-0.062
PH-FLD pH-Field	0.068	0.505	-0.127	-0.265	0.446	0.293	-0.071	-0.056	0.129	-0.126	-0.087	0.019
TCOLIB Total Coliform	-0.001	-0.129	0.062	-0.033	-0.117	0.182	0.580	0.065	0.275	0.043	-0.386	0.591
TEMP-FLD Temp-Field	0.012	0.198	-0.044	-0.284	0.511	0.537	-0.064	0.053	0.065	-0.119	-0.081	-0.038
TRESCG-FLD Total Chlorine Residual-Field	-0.050	0.746	-0.159	0.035	-0.083	-0.101	-0.155	-0.027	0.097	-0.059	0.013	0.058
D-O2-FLD Dissolved Oxygen-Field	0.891	0.186	0.011	0.067	-0.144	-0.019	-0.009	-0.004	0.025	0.008	-0.061	-0.039
NO2 Nitrite, as N	0.591	-0.486	0.134	-0.047	0.195	0.063	-0.028	-0.019	-0.048	-0.067	0.152	0.088
NO3 Nitrite, as N	0.753	-0.350	0.036	-0.035	0.218	-0.131	-0.114	-0.011	0.082	-0.062	0.043	0.084
NO3NO2 Nitrate+Nitrite as N	0.636	-0.409	0.077	-0.045	0.273	-0.189	-0.208	-0.014	0.193	-0.070	0.022	0.121
TOC Total Organic Carbon	0.201	0.086	0.008	-0.027	-0.097	0.365	0.269	-0.014	-0.489	0.008	0.610	0.260
HPC35 HPC (R2A Agar at 35c)	0.042	-0.355	0.085	-0.068	0.092	0.213	0.390	-0.053	-0.314	-0.205	-0.276	-0.507
NH3 Ammonia	0.851	0.348	-0.002	0.064	-0.236	0.027	0.023	0.001	-0.006	0.034	-0.073	-0.062
NH3H-FLD Ammonia Hach Test-Field §	-0.294	-0.029	-0.021	-0.121	0.080	-0.026	-0.078	-0.663	0.117	-0.180	0.215	0.191

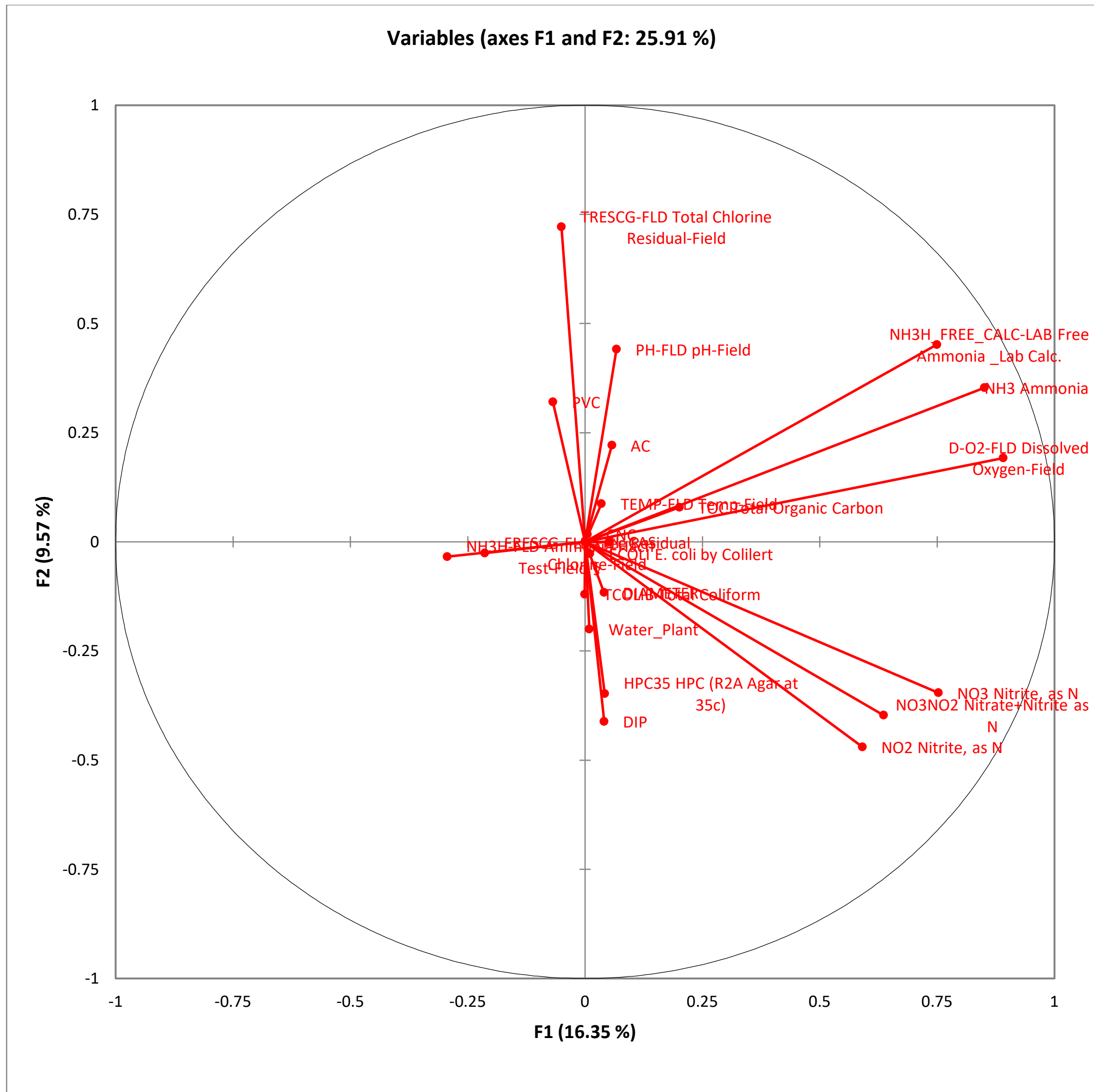


Figure 2 Factor diagram showing grouped correlations among variables constituting F1 and F2

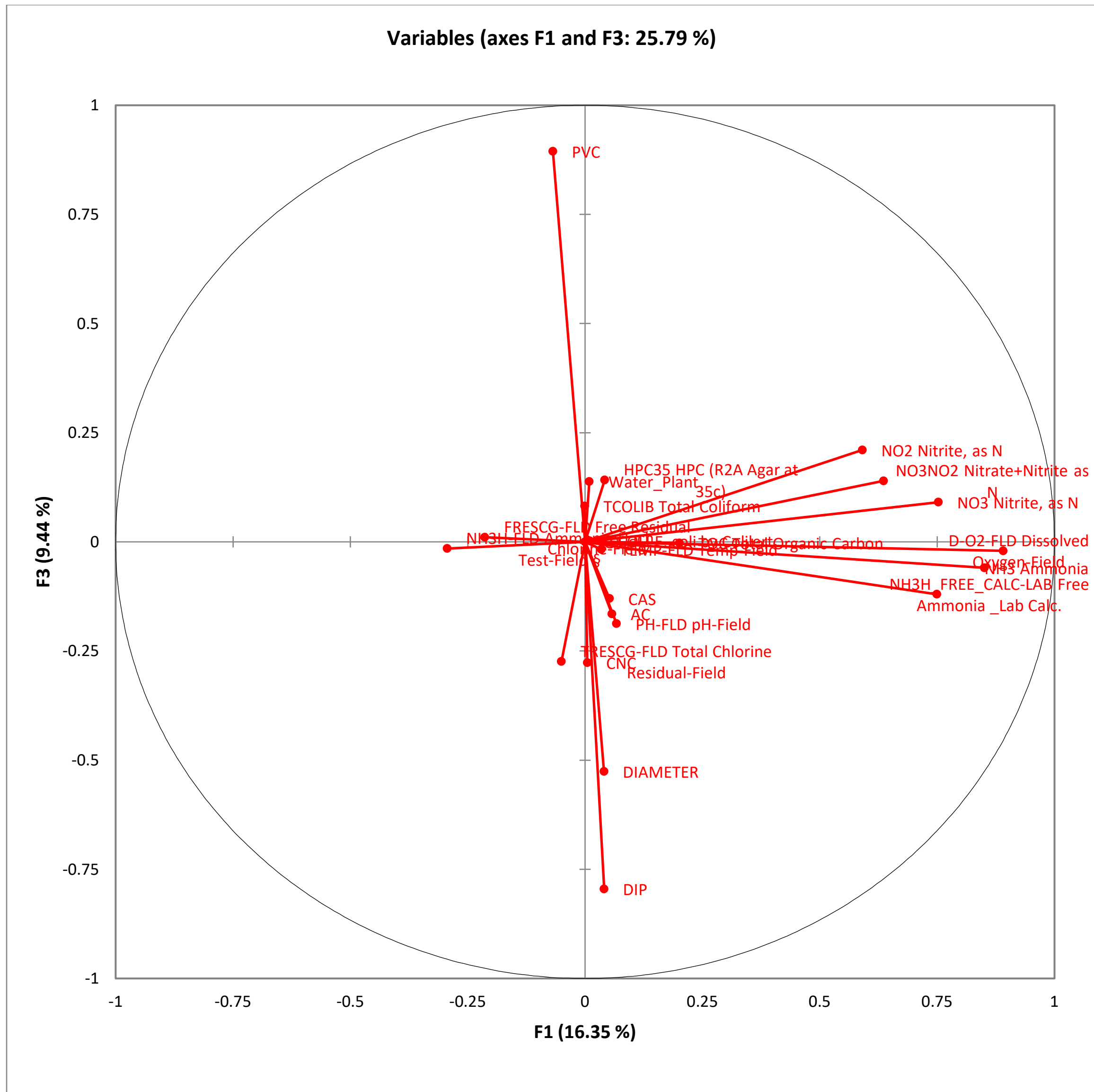


Figure 3 Factor diagram showing grouped correlations among variables constituting F1 and F3

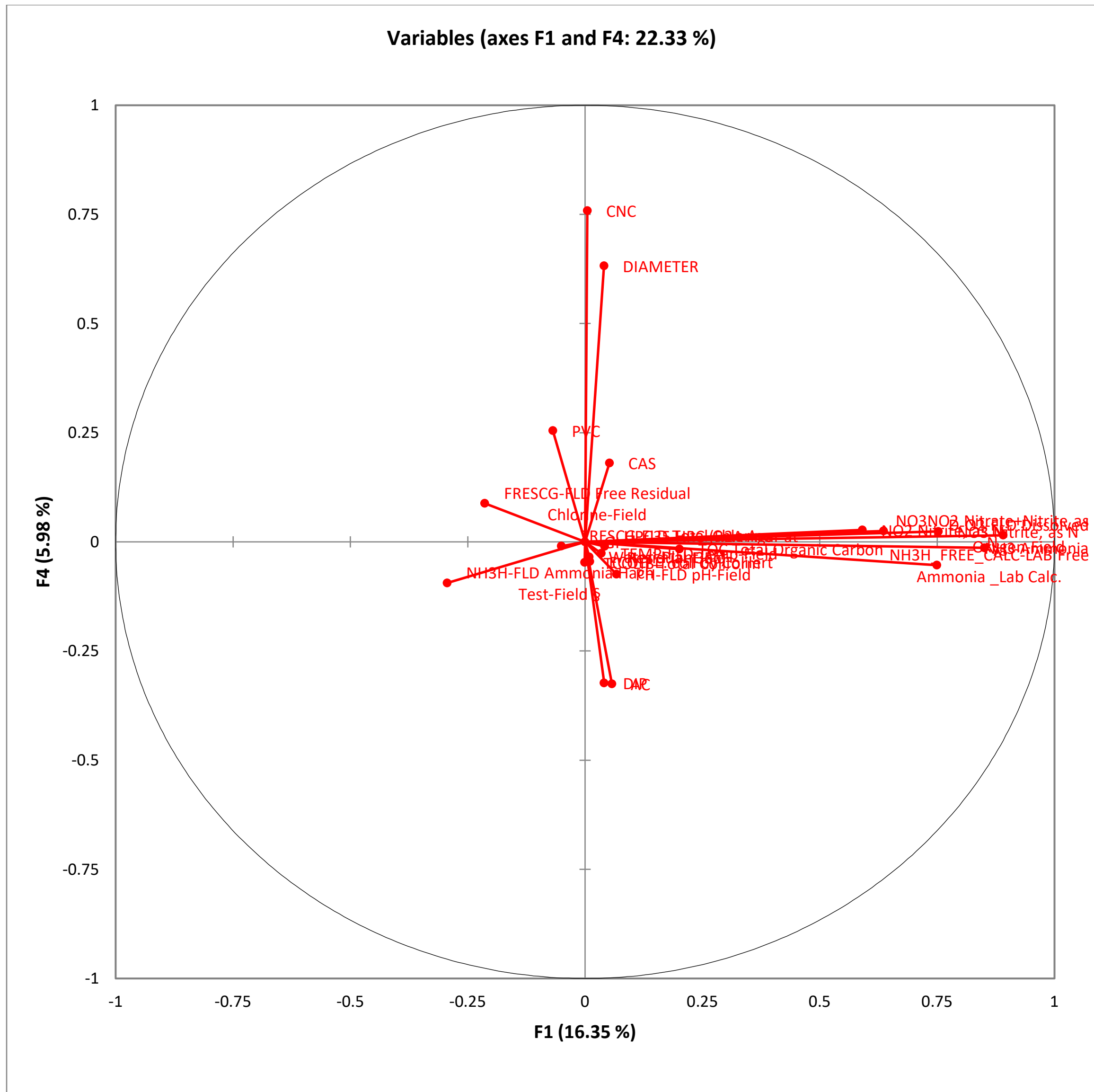


Figure 4 Factor diagram showing grouped correlations among variables constituting F1 and F4

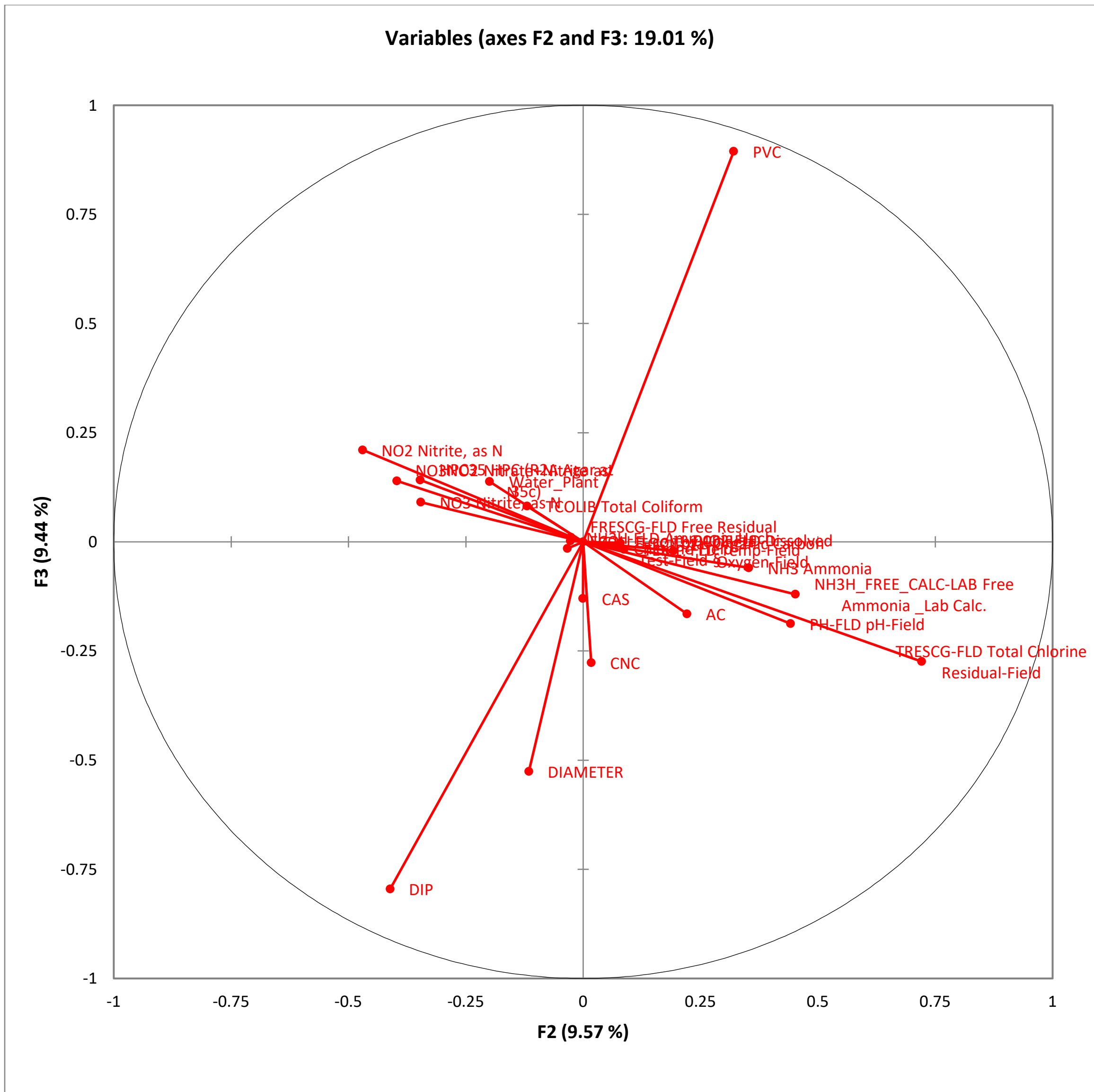


Figure 5 Factor diagram showing grouped correlations among variables constituting F3 and F2.

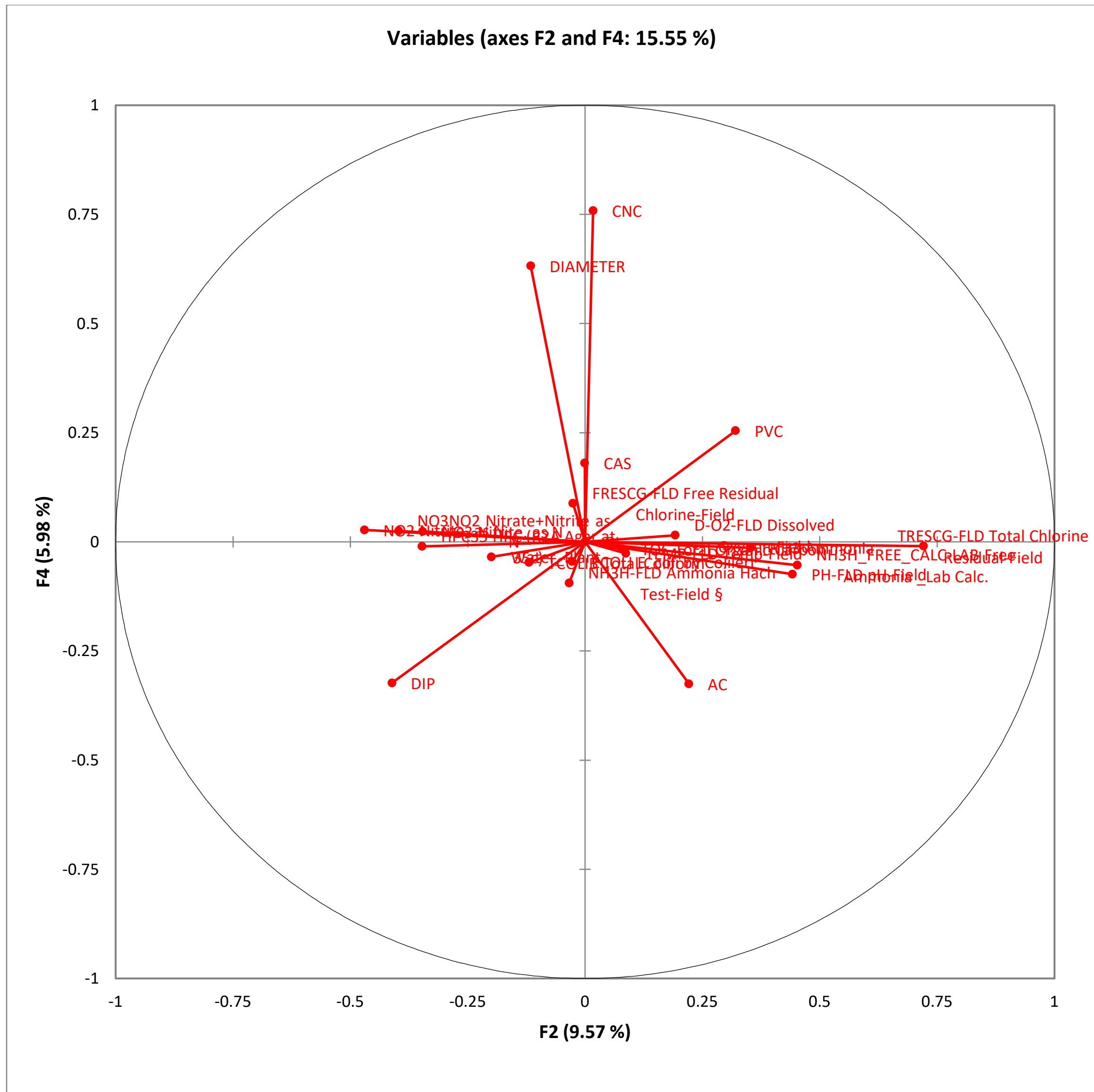


Figure 6 Factor diagram showing grouped correlations among variables constituting F4 and F2.

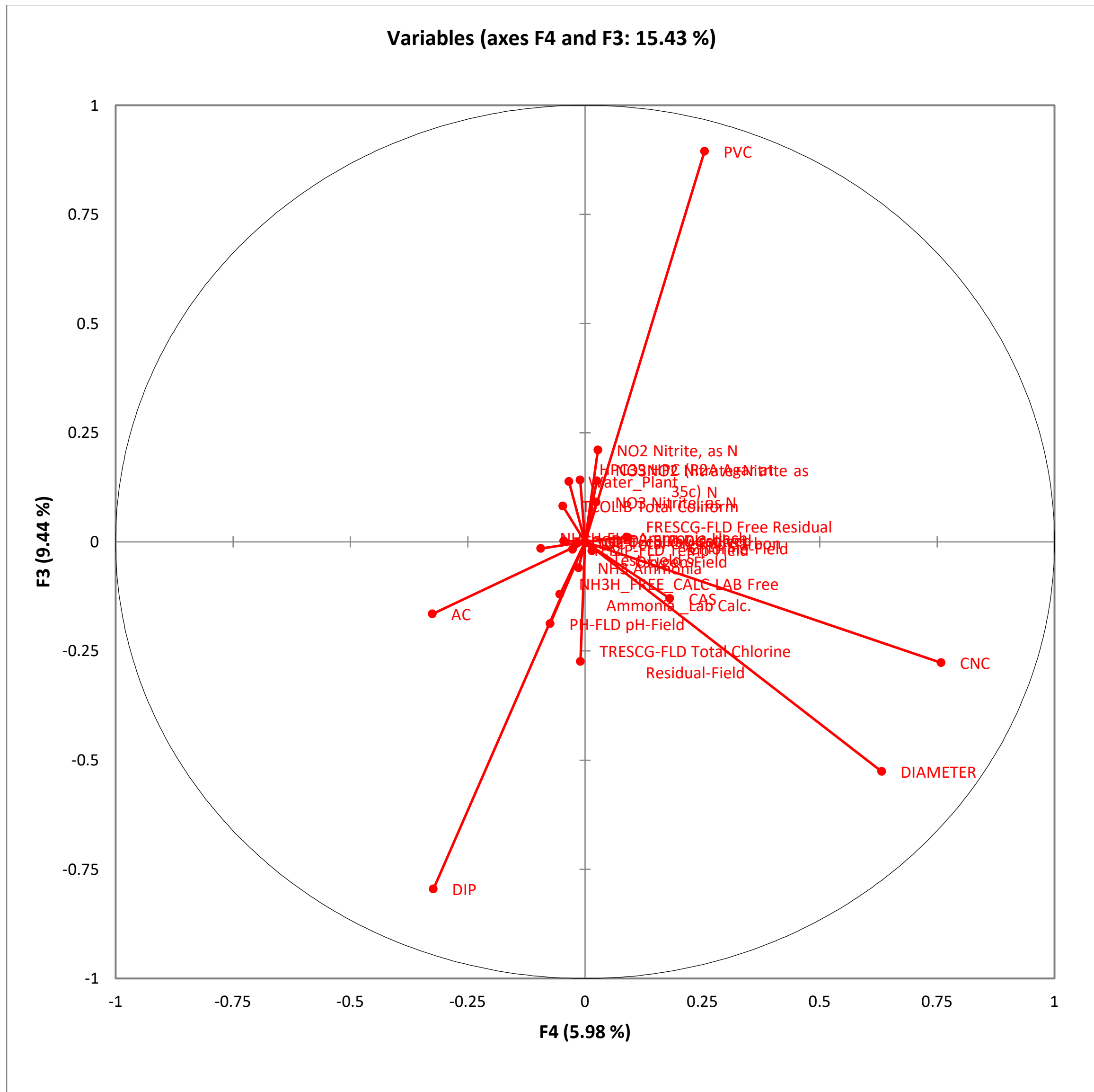


Figure 7 Factor diagram showing grouped correlations among variables constituting F3 and F4.

Table 7 Factor loading percentages for Factors developed from PCA analysis.

Contribution of the variables (%):	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Water_Plant	0.002	2.208	0.536	0.023	9.735	21.209	5.439	0.193	1.833	0.524	0.007	0.116
AC	0.091	2.682	0.787	8.843	2.942	21.393	12.410	1.815	1.380	8.637	0.917	0.296
CAS	0.075	0.033	0.812	0.334	12.758	0.049	1.411	1.878	0.740	66.851	0.006	0.071
CNC	0.001	0.131	3.491	40.371	4.022	0.014	2.057	0.016	0.037	10.661	0.051	0.067
DIP	0.047	3.802	34.673	5.577	5.148	0.869	1.402	0.030	0.028	0.035	0.122	0.012
PVC	0.131	1.548	41.868	4.208	0.394	0.481	0.240	0.118	0.123	0.012	0.000	0.005
GI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DIAMETER	0.045	0.061	13.869	26.265	2.854	1.903	0.018	0.104	0.091	0.063	0.003	0.017
ECOLI E. coli by Colilert	0.003	0.031	0.000	0.113	0.232	0.418	13.379	1.321	46.153	0.526	23.196	14.071
FRESCG-FLD Free Residual Chlorine-Field	1.269	0.017	0.001	0.058	2.662	0.020	3.189	52.255	0.227	0.011	1.726	2.157
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	15.595	10.126	0.119	0.023	1.616	0.885	0.558	0.000	0.256	0.201	0.321	0.427
PH-FLD pH-Field	0.127	12.014	0.777	5.314	15.451	6.906	0.456	0.287	1.634	1.599	0.811	0.039
TCOLIB Total Coliform	0.000	0.788	0.186	0.082	1.064	2.667	30.656	0.388	7.393	0.185	15.876	38.592
TEMP-FLD Temp-Field	0.004	1.842	0.092	6.119	20.319	23.244	0.378	0.257	0.410	1.431	0.693	0.156
TRESCG-FLD Total Chlorine Residual-Field	0.071	26.259	1.209	0.094	0.540	0.826	2.178	0.069	0.909	0.351	0.017	0.372
D-O2-FLD Dissolved Oxygen-Field	22.082	1.635	0.006	0.340	1.604	0.028	0.007	0.002	0.061	0.007	0.400	0.165
NO2 Nitrite, as N	9.726	11.124	0.863	0.169	2.956	0.315	0.071	0.035	0.229	0.447	2.449	0.857
NO3 Nitrite, as N	15.769	5.767	0.062	0.090	3.682	1.387	1.181	0.011	0.654	0.387	0.193	0.788
NO3NO2 Nitrate+Nitrite as N	11.254	7.880	0.282	0.155	5.802	2.884	3.925	0.019	3.645	0.491	0.050	1.604
TOC Total Organic Carbon	1.125	0.346	0.003	0.054	0.737	10.733	6.587	0.018	23.261	0.006	39.581	7.434
HPC35 HPC (R2A Agar at 35c)	0.049	5.945	0.344	0.347	0.659	3.658	13.853	0.260	9.603	4.229	8.112	28.319
NH3 Ammonia	20.132	5.724	0.000	0.307	4.330	0.057	0.046	0.000	0.004	0.115	0.565	0.427
NH3H-FLD Ammonia Hach Test-Field §	2.403	0.039	0.021	1.113	0.493	0.055	0.560	40.926	1.332	3.233	4.904	4.008

The next step in the process is to develop a varimax rotation of the factors to attempt to rotate the axes between factors to improve the fit of the variables. Table 8 and Table 9 outline the impact of each factor as a result of the varimax rotation and the new composite factor loadings (now referred to as D1-D12). Table 10 is the percentage of each new factor as a percentage of variance in the results. As was identified in Table 6 and Table 11, shows the contribution of the variable to each varimax factor. Figure 8, Figure 9, Figure 10, Figure 11, Figure 12, and Figure 13, show the varimax diagrams for the factors D1-D4. Figure 14, Figure 15, Figure 16, Figure 17, Figure 18, and Figure 19 shows the varimax biplots for each of the same variables.

Table 8 Factor Loadings after rotation

Results after the Varimax rotation:												
Rotation matrix:												
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
D1	0.806	0.580	0.046	0.013	-0.021	0.017	-0.016	0.016	0.002	0.034	0.096	-0.006
D2	0.465	-0.609	-0.369	-0.020	-0.219	-0.213	0.009	0.037	0.014	0.008	-0.036	0.425
D3	-0.106	0.224	-0.854	-0.360	0.134	0.108	0.000	-0.003	-0.010	-0.112	0.032	-0.192
D4	-0.017	0.035	-0.336	0.880	0.210	-0.058	0.095	-0.019	-0.042	0.198	-0.029	-0.109
D5	0.243	-0.359	0.119	-0.030	0.733	0.217	-0.033	0.017	0.070	-0.309	0.289	-0.166
D6	-0.011	-0.111	-0.055	0.103	-0.345	0.682	-0.178	0.201	0.273	0.248	0.425	-0.018
D7	-0.085	0.084	-0.026	-0.015	0.223	-0.174	-0.855	0.124	-0.212	0.229	0.075	0.234
D8	-0.085	0.051	0.024	-0.155	0.194	-0.129	0.442	0.541	-0.339	0.391	0.344	0.191
D9	-0.019	0.098	0.001	-0.075	0.310	0.073	0.049	0.297	0.659	0.228	-0.487	0.263
D10	-0.054	0.103	-0.013	0.211	-0.163	0.039	-0.087	0.649	-0.091	-0.675	-0.110	0.092
D11	-0.213	0.252	-0.037	0.095	0.049	-0.181	0.096	-0.283	0.331	-0.273	0.507	0.561
D12	-0.018	0.082	-0.004	0.023	0.145	0.583	0.106	-0.242	-0.454	-0.046	-0.305	0.510

Table 9 Factor loadings after varimax rotation showing improvement in the explanation of variables as a result of the rotation.

Factor loadings after Varimax rotation:												
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
Water_Plant	-0.041	0.092	0.039	-0.106	0.703	0.179	0.095	-0.054	0.006	-0.265	0.134	0.205
AC	0.021	0.037	0.084	-0.084	-0.803	0.082	0.077	-0.052	0.026	-0.142	0.066	0.257
CAS	-0.007	0.038	0.020	0.009	-0.009	0.013	0.001	-0.007	-0.004	0.949	0.025	0.068
CNC	0.007	-0.015	-0.046	0.881	-0.082	0.002	-0.004	-0.007	0.009	-0.088	-0.008	-0.028
DIP	-0.004	0.008	0.976	-0.019	0.122	-0.019	-0.020	0.009	-0.002	-0.045	-0.014	-0.053
PVC	-0.003	-0.024	-0.966	-0.109	0.165	-0.012	-0.006	0.011	-0.007	-0.065	-0.011	-0.042
GI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DIAMETER	0.013	0.007	0.282	0.755	0.150	-0.013	0.032	-0.003	-0.007	0.173	0.014	0.071
ECOLI E. coli by Colilert	-0.010	0.018	0.005	0.004	-0.014	0.068	-0.005	0.028	0.958	-0.005	0.041	-0.025
FRESCG-FLD Free Residual Chlorine-Field	-0.250	0.005	-0.025	0.019	0.003	-0.089	0.808	0.013	0.012	-0.038	0.017	0.139
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	0.880	0.043	0.007	-0.011	-0.049	0.012	-0.021	0.102	-0.008	0.036	0.125	0.151
PH-FLD pH-Field	0.107	-0.043	-0.015	0.007	-0.077	0.040	0.018	0.089	-0.042	0.124	0.030	0.822
TCOLIB Total Coliform	0.035	-0.052	-0.012	-0.010	0.040	0.842	-0.008	-0.064	0.141	0.025	-0.124	0.028
TEMP-FLD Temp-Field	0.040	-0.017	-0.001	-0.006	0.009	-0.023	0.009	0.978	0.025	-0.005	-0.023	0.042
TRESCG-FLD Total Chlorine Residual-Field	0.299	-0.464	-0.029	0.079	-0.114	-0.255	0.006	-0.068	0.008	-0.058	-0.145	0.491
D-O2-FLD Dissolved Oxygen-Field	0.865	0.331	-0.001	0.025	-0.002	-0.012	-0.034	-0.017	0.004	-0.017	0.015	-0.022
NO2 Nitrite, as N	0.152	0.773	-0.009	-0.020	0.071	0.057	-0.004	0.007	-0.006	0.013	0.214	-0.090
NO3 Nitrite, as N	0.352	0.808	0.048	0.017	-0.041	-0.025	-0.004	-0.018	0.002	0.027	-0.032	-0.032
NO3NO2 Nitrate+Nitrite as N	0.201	0.845	0.013	-0.004	-0.008	-0.054	0.007	-0.026	0.025	0.016	-0.167	0.025
TOC Total Organic Carbon	0.165	-0.029	-0.004	0.001	0.017	-0.068	0.006	-0.035	0.053	0.028	0.898	0.012
HPC35 HPC (R2A Agar at 35c)	-0.139	0.192	-0.004	0.032	-0.084	0.492	-0.087	0.148	-0.250	-0.056	0.311	-0.164
NH3 Ammonia	0.940	0.154	-0.008	0.004	0.008	-0.006	-0.031	-0.017	-0.001	-0.022	0.052	0.025
NH3H-FLD Ammonia Hach Test-Field §	-0.380	0.013	-0.002	-0.005	0.034	-0.137	-0.627	-0.008	0.039	-0.079	0.029	0.319

Table 10 percent of variability for each factor D1-D12.

Percentage of variance after Varimax rotation:												
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
Variability (%)	13.420	10.778	9.006	6.305	5.671	5.004	4.874	4.611	4.581	4.855	4.822	5.568
Cumulative %	13.420	24.198	33.204	39.509	45.180	50.184	55.058	59.669	64.250	69.106	73.928	79.496

Table 11 Contribution of the variable to each varimax factor

Contribution of the variables (%) after Varimax rotation:	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
Water_Plant	0.056	0.355	0.075	0.815	39.581	2.912	0.842	0.282	0.004	6.556	1.681	3.434
AC	0.015	0.056	0.353	0.507	51.631	0.604	0.553	0.268	0.067	1.878	0.407	5.399
CAS	0.002	0.061	0.021	0.006	0.006	0.016	0.000	0.005	0.001	84.281	0.057	0.373
CNC	0.002	0.009	0.109	55.977	0.538	0.000	0.001	0.004	0.007	0.721	0.005	0.065
DIP	0.001	0.003	48.097	0.027	1.184	0.032	0.038	0.008	0.000	0.190	0.019	0.228
PVC	0.000	0.025	47.105	0.860	2.175	0.012	0.003	0.012	0.005	0.400	0.012	0.141
GI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DIAMETER	0.006	0.002	4.008	41.133	1.797	0.016	0.097	0.001	0.004	2.789	0.019	0.415
ECOLI E. coli by Colilert	0.003	0.013	0.001	0.001	0.016	0.420	0.003	0.079	90.993	0.002	0.156	0.052
FRESCG-FLD Free Residual Chlorine-Field	2.110	0.001	0.031	0.025	0.001	0.727	60.842	0.016	0.013	0.135	0.029	1.578
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	26.207	0.076	0.003	0.009	0.194	0.013	0.041	1.025	0.006	0.119	1.472	1.858
PH-FLD pH-Field	0.390	0.077	0.012	0.004	0.478	0.149	0.031	0.780	0.175	1.429	0.085	55.187
TCOLIB Total Coliform	0.042	0.113	0.008	0.007	0.130	64.406	0.006	0.406	1.973	0.057	1.439	0.062
TEMP-FLD Temp-Field	0.053	0.012	0.000	0.003	0.007	0.048	0.008	94.218	0.061	0.002	0.051	0.144
TRESCG-FLD Total Chlorine Residual-Field	3.031	9.094	0.044	0.451	1.038	5.918	0.003	0.456	0.006	0.311	1.993	19.685
D-O2-FLD Dissolved Oxygen-Field	25.358	4.632	0.000	0.046	0.000	0.013	0.108	0.027	0.001	0.028	0.021	0.041
NO2 Nitrite, as N	0.783	25.189	0.004	0.029	0.406	0.293	0.001	0.005	0.004	0.015	4.313	0.664
NO3 Nitrite, as N	4.187	27.539	0.116	0.022	0.132	0.057	0.002	0.031	0.000	0.066	0.098	0.085
NO3NO2 Nitrate+Nitrite as N	1.367	30.133	0.008	0.001	0.005	0.261	0.005	0.068	0.064	0.023	2.632	0.050
TOC Total Organic Carbon	0.922	0.035	0.001	0.000	0.023	0.419	0.004	0.123	0.281	0.076	76.059	0.011
HPC35 HPC (R2A Agar at 35c)	0.657	1.560	0.001	0.075	0.561	21.970	0.706	2.151	6.179	0.296	9.123	2.198
NH3 Ammonia	29.916	1.006	0.004	0.001	0.005	0.004	0.092	0.028	0.000	0.045	0.251	0.050
NH3H-FLD Ammonia Hach Test-Field §	4.893	0.007	0.000	0.002	0.093	1.711	36.615	0.006	0.152	0.579	0.078	8.281

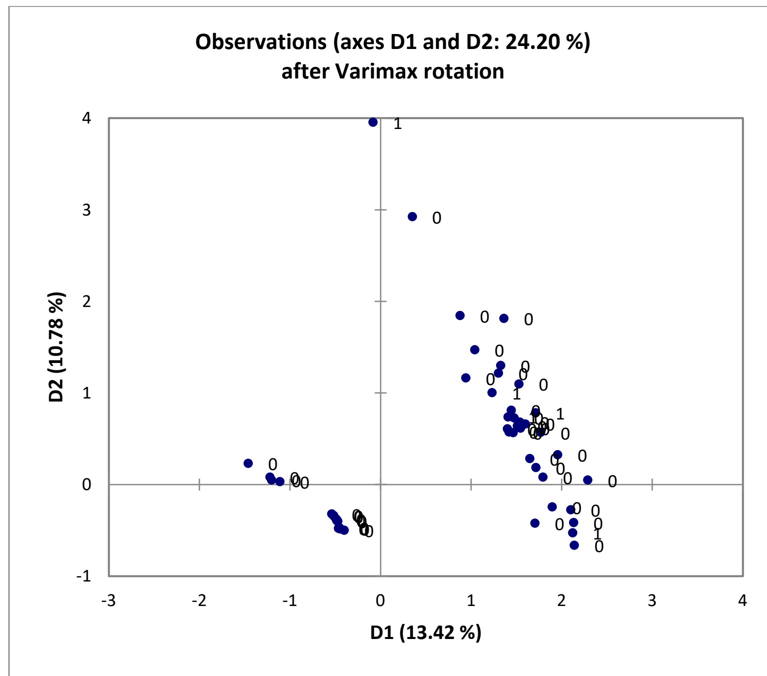


Figure 8 Factor diagram showing grouped correlations among variables constituting D1 and D2.

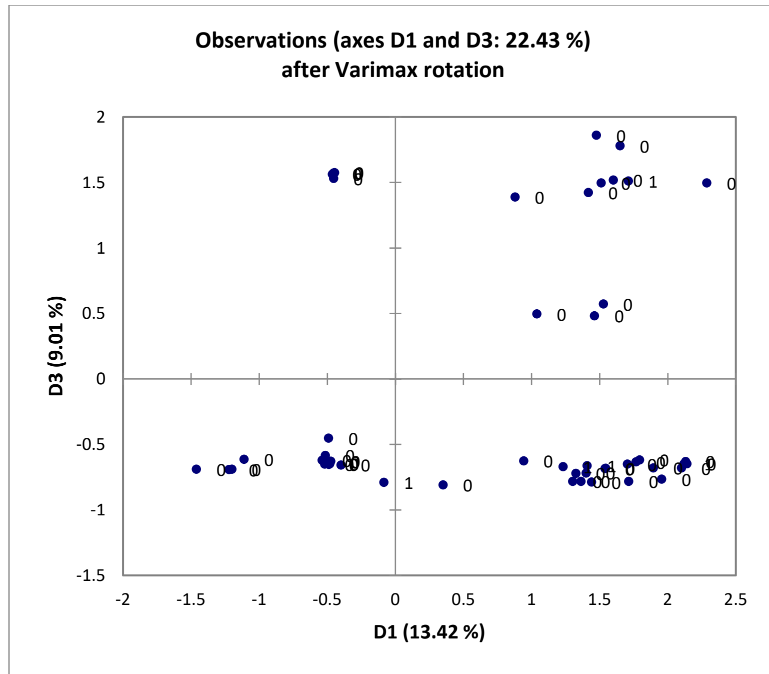


Figure 9 Factor diagram showing grouped correlations among variables constituting D1 and D3

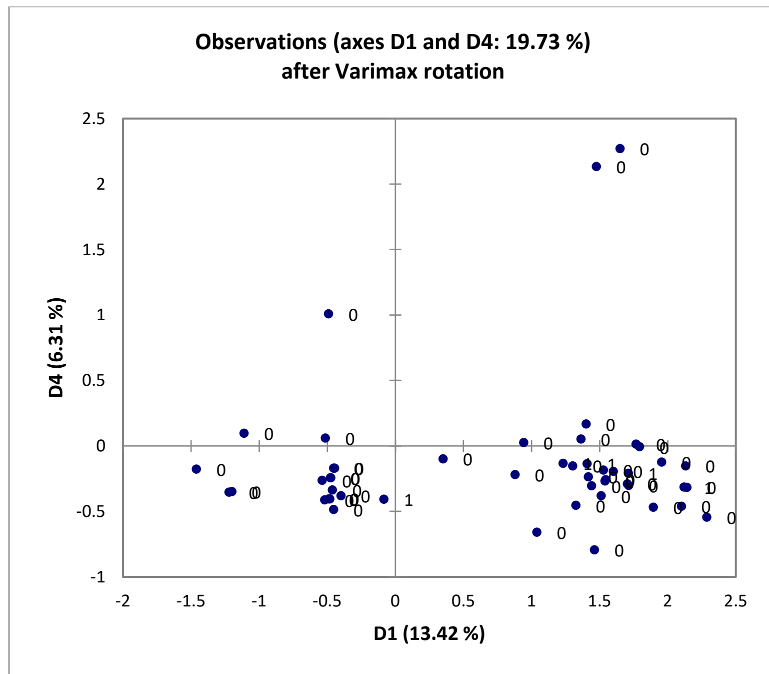


Figure 10 Factor diagram showing grouped correlations among variables constituting D1 and D4

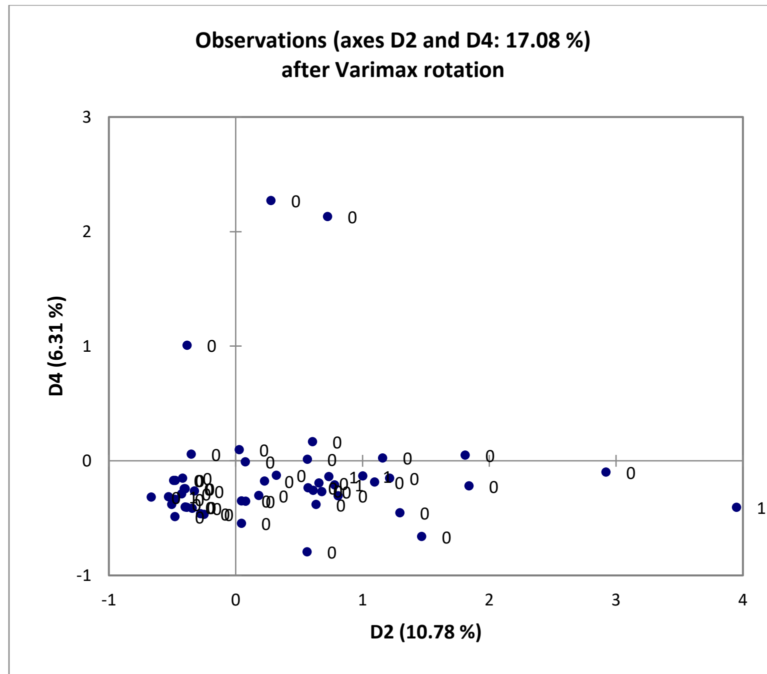


Figure 11 Factor diagram showing grouped correlations among variables constituting D4 and D2.

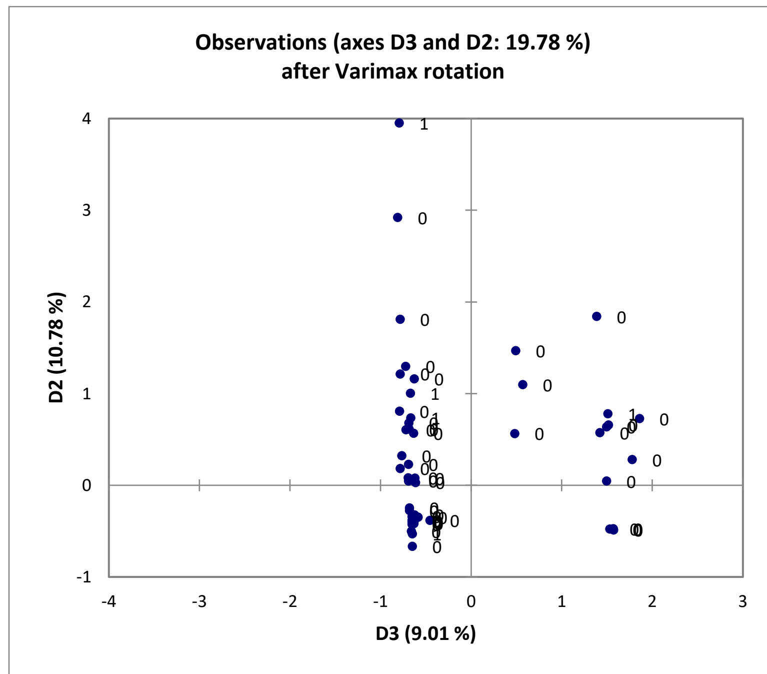


Figure 12 Factor diagram showing grouped correlations among variables constituting D3 and D2.

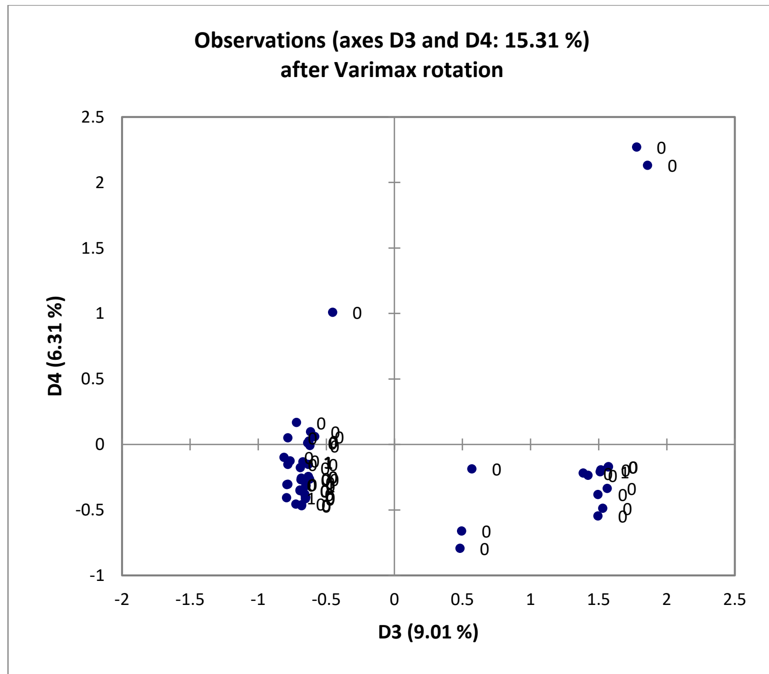


Figure 13 Factor diagram showing grouped correlations among variables constituting D3 and D4

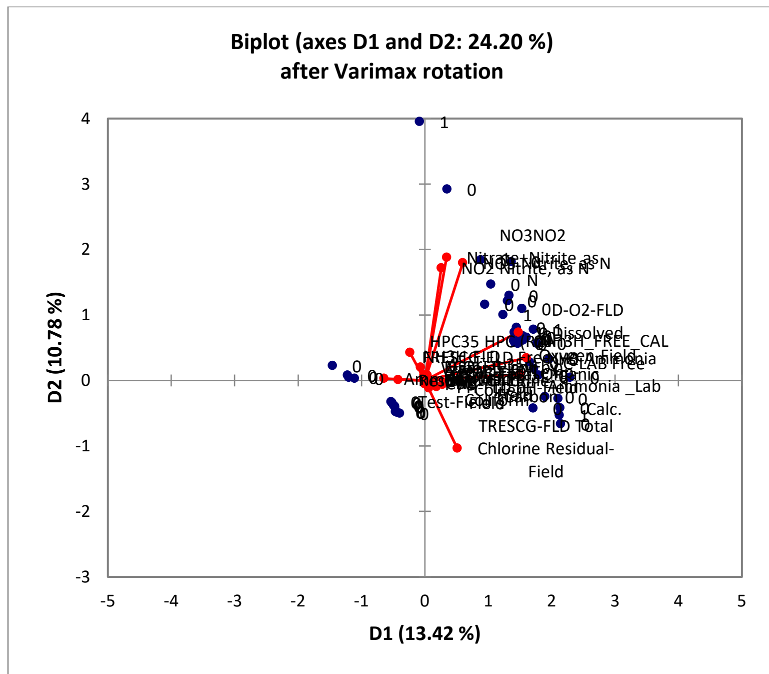


Figure 14 Varimax biplot showing grouped correlations among variables constituting D1 and D2

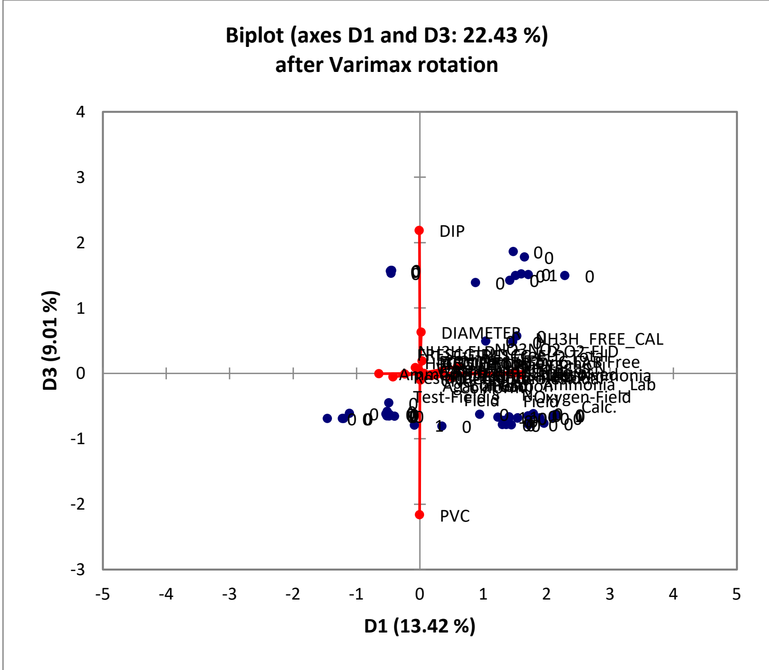


Figure 15 Varimax biplot showing grouped correlations among variables constituting D1 and D3

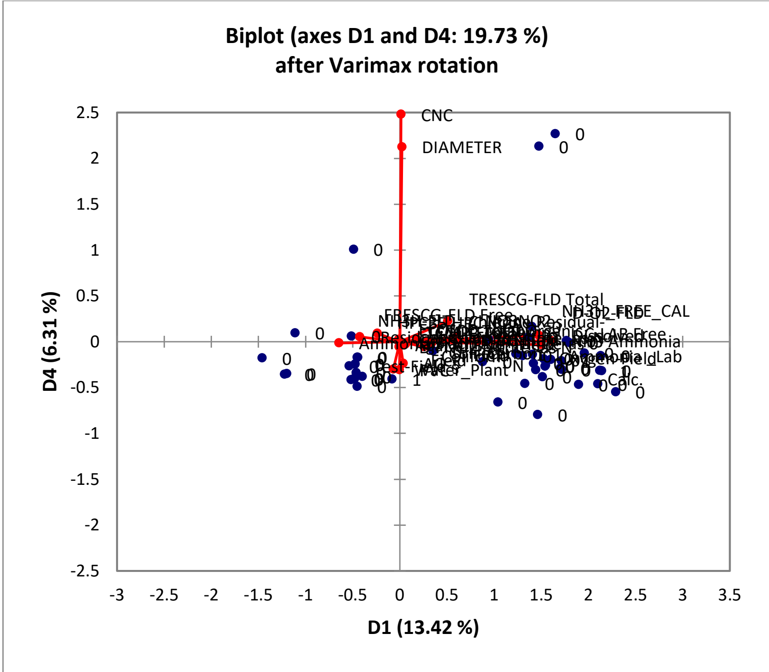


Figure 16 Varimax biplot showing grouped correlations among variables constituting D1 and D4

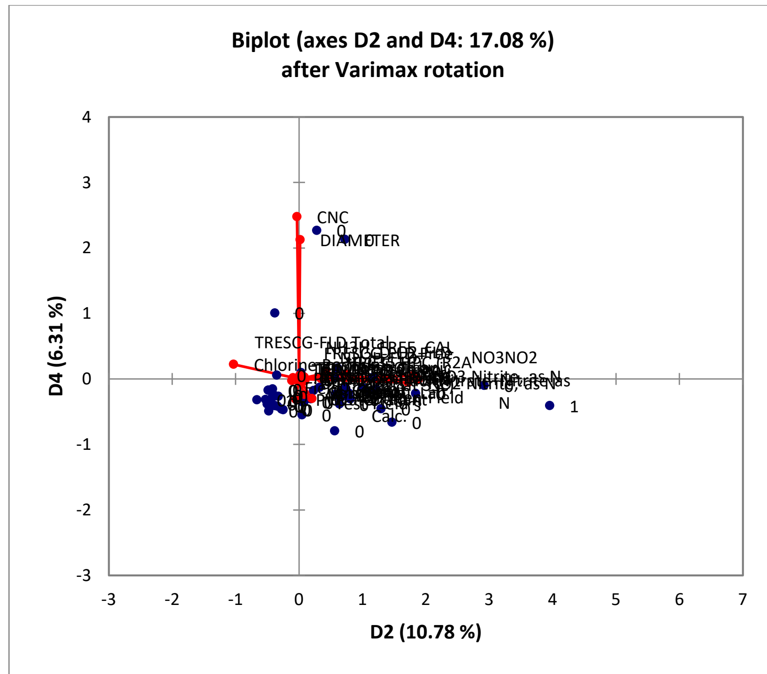


Figure 17 Varimax biplot showing grouped correlations among variables constituting D4 and D2

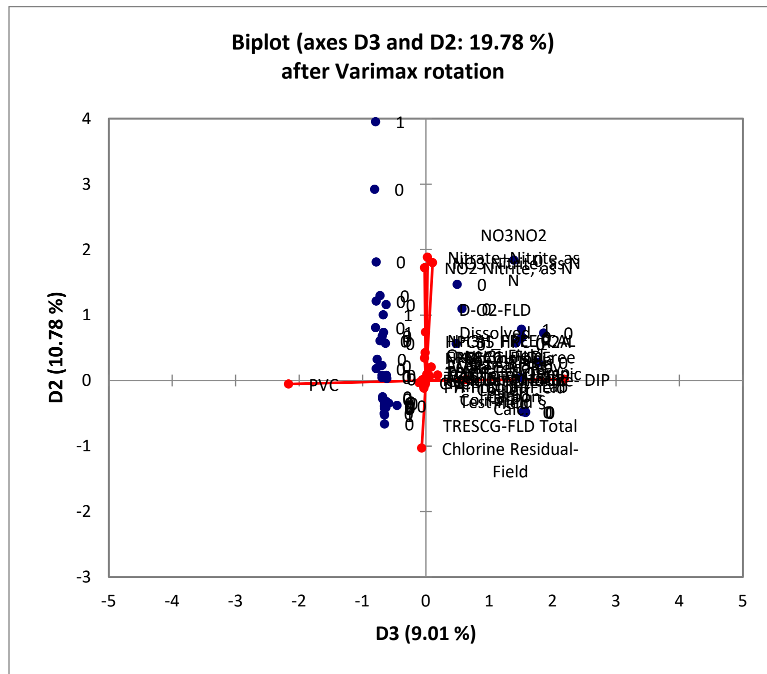


Figure 18 Varimax biplot showing grouped correlations among variables constituting D13 and D2

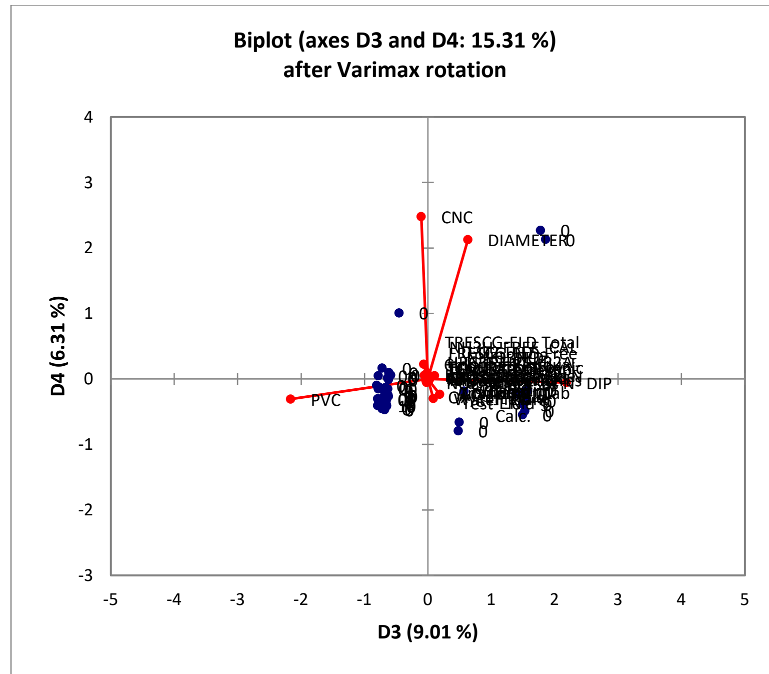


Figure 19 Varimax biplot showing grouped correlations among variables constituting D3 and D4

Discriminant Analysis

The results the PCA did not indicate clear correlations. The discriminant analysis was then performed to determine if there was a difference between those sits with nitrification and those without. The data indicates that they can be separated as noted in the different centroids:

Table 12 Table of Centroids

Centroids:

	F1
0	0.089
1	-1.619

Figure 20 and Figure 21 shows the separate centroids and the two populations of data. Table 13 confirms that the factors in DA will predict presence or absence of the nitrification over 98 percent of the time.

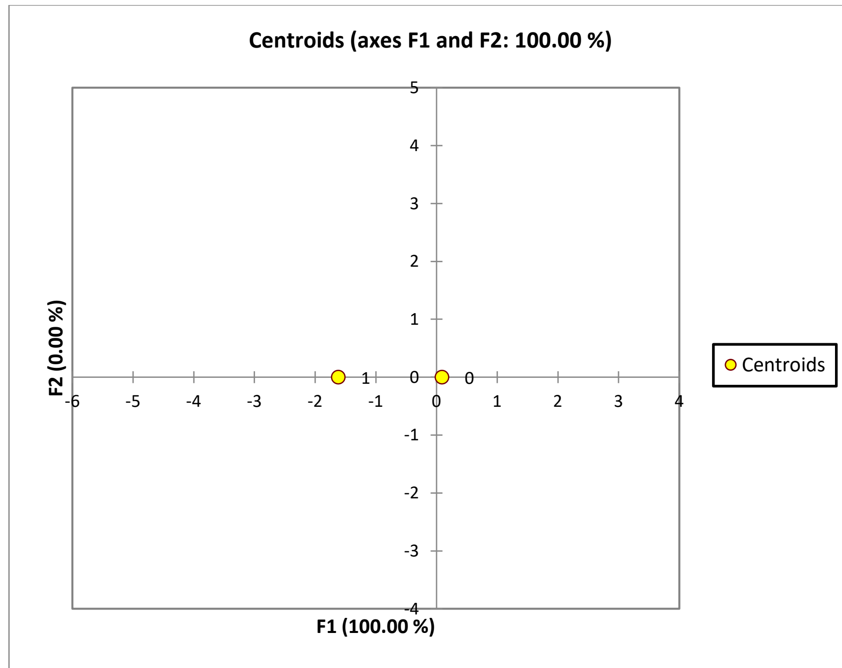


Figure 20 DA analysis showing centroids of the two populations

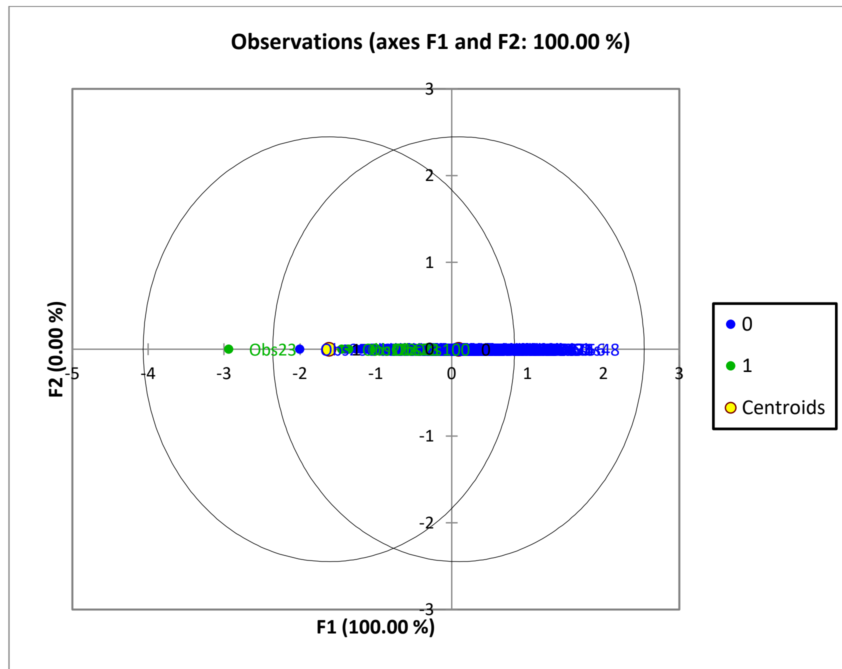


Figure 21 DA showing the centroids and the populations–

Table 13 Confusion matrix for the estimation sample:

from \ to	0	1	Total	% correct
0	14893	298	15191	98.04%
1	673	166	839	19.79%
Total	15566	464	16030	93.94%

Linear Regression

Linear regression was pursued to develop a predictive equation as a means to predict nitrification in other systems. As a part of the linear regression analysis, there were a series of variable removed from the analysis because they did not provide any variability in the analysis:

Table 14 Removed Variables

Variable	Reason
PVC	Multicollinearity
GI	Constant
PC	0.876

Table 15 outlines the goodness of fit tests for the linear regression analysis. These results lead to the model parameters shown in Table 16. Figure 22 shows the relative weight of each factor. Issue that are important to the equation are:

- Water plant
- Free ammonia
- Nitrite
- Residual chlorine (negative)
- Ammonia (negative) and
- Dissolved oxygen (negative)

Table 15 Goodness of Fit Statistics

Observations	16030.000
Sum of weights	16030.000
DF	16008.000
R^2	0.127
Adjusted R^2	0.125
MSE	0.043
RMSE	0.208
MAPE	82.775
DW	1.388
Cp	22.000
AIC	-50274.894
SBC	-50105.885

Table 16 Model Parameters

Model parameters:

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	0.091	0.031	2.956	0.003	0.031	0.151
Water_Plant	0.013	0.001	21.713	< 0.0001	0.011	0.014
AC	0.009	0.011	0.824	0.410	-0.013	0.031
CAS	-0.019	0.020	-0.933	0.351	-0.058	0.020
CNC	0.105	0.022	4.726	< 0.0001	0.061	0.148
DIP	-0.006	0.004	-1.677	0.094	-0.014	0.001
PVC	0.000	0.000				
GI	0.000	0.000				
DIAMETER	-0.002	0.000	-8.418	< 0.0001	-0.003	-0.002
ECOLI E. coli by Colilert	0.162	0.066	2.452	0.014	0.032	0.292
FRESCG-FLD Free Residual Chlorine-Field	-0.001	0.002	-0.344	0.731	-0.006	0.004
NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.	0.790	0.078	10.083	< 0.0001	0.637	0.944
PH-FLD pH-Field	0.006	0.004	1.352	0.176	-0.003	0.014
TCOLIB Total Coliform	0.036	0.014	2.533	0.011	0.008	0.064
TEMP-FLD Temp-Field	0.001	0.001	2.144	0.032	0.000	0.003
TRESCG-FLD Total Chlorine Residual-Field	-0.053	0.002	-25.302	< 0.0001	-0.057	-0.049
D-O2-FLD Dissolved Oxygen-Field	-0.004	0.001	-4.163	< 0.0001	-0.006	-0.002
NO2 Nitrite, as N	0.380	0.032	11.992	< 0.0001	0.318	0.443
NO3 Nitrite, as N	-0.156	0.036	-4.397	< 0.0001	-0.226	-0.087
NO3NO2 Nitrate+Nitrite as N	0.015	0.025	0.586	0.558	-0.034	0.064
TOC Total Organic Carbon	0.003	0.003	0.972	0.331	-0.003	0.008
HPC35 HPC (R2A Agar at 35c)	0.000	0.000	2.550	0.011	0.000	0.000
NH3 Ammonia	-0.058	0.017	-3.305	0.001	-0.092	-0.023

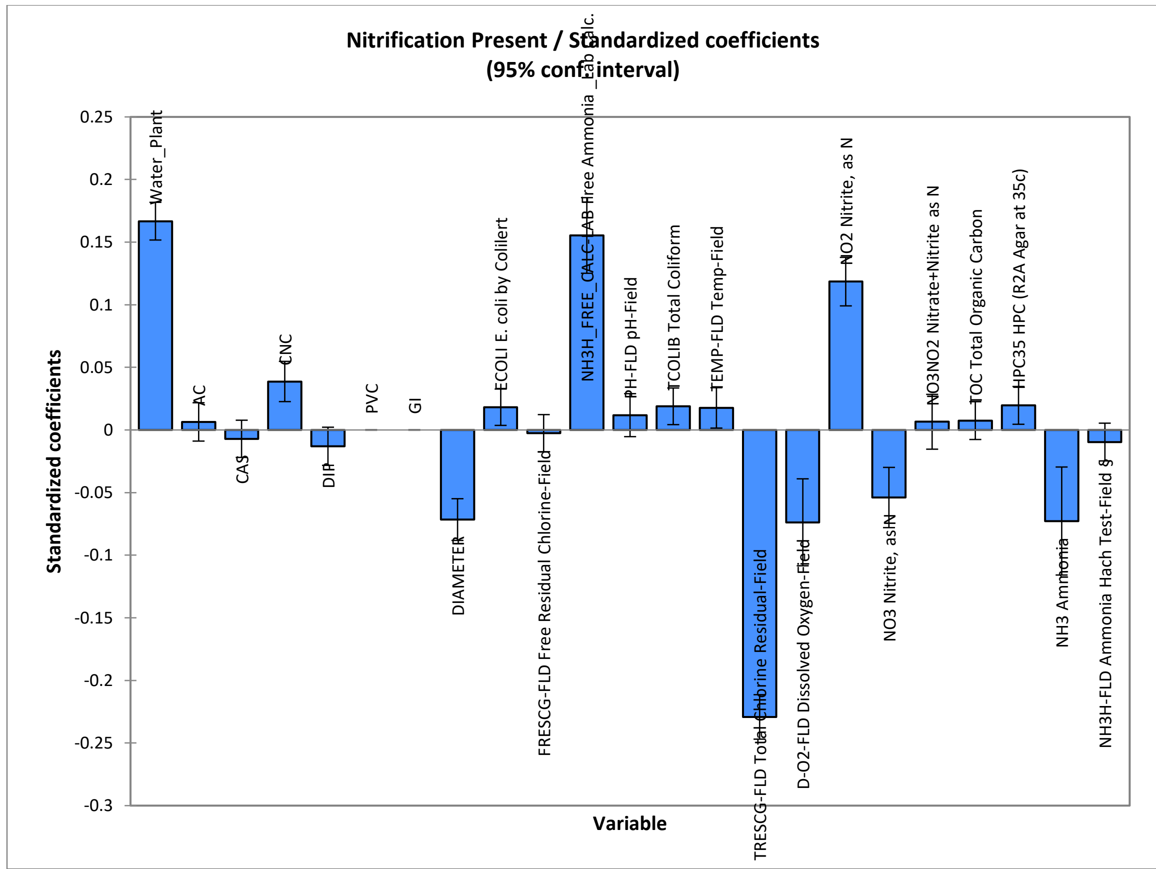


Figure 22 Nitrification Present

The full equation is:

Nitrification Present = 9.08729382469469E-02

+1.25137908536162E-02*Water Plant

+9.20282026348584E-03*AC

-1.85842255436452E-02*CAS

+0.104900752874613*CNC

-6.39471312694251E-03*DIP

-2.38111710877115E-03*DIAMETER

+0.161987151519582

*ECOLI E. coli by Colilert

-8.59199131895524E-04*FRESCG-FLD Free Residual Chlorine-Field

+0.790186455930821*NH3H_FREE_CALC-LAB Free Ammonia _Lab Calc.

+5.57149205995384E-03*PH-FLD pH-Field

+3.63300241629198E-02*TCOLIB Total Coliform

+1.45290410065473E-03*TEMP-FLD Temp-Field

-5.26540380854557E-02*TRESCG-FLD Total Chlorine Residual-Field

-4.37088468780576E-03*D-O2-FLD Dissolved Oxygen-Field

+0.38038773739833*NO2 Nitrite, as N

-0.156443301790404*NO3 Nitrite, as N+1.46233527948516E-02*NO3NO2

Nitrate+Nitrite as N

+2.7871144923777E-03*TOC Total Organic Carbon

+1.14589703640961E-05*HPC35 HPC (R2A Agar at 35c)

-5.76660832919546E-02*NH3 Ammonia

-0.007928667229052*NH3H-FLD Ammonia Hach Test-Field §

Figure 23, Figure 24, and Figure 25 show the results of applying the equation to the data. Figure 23 shows the two populations can be separated. Figure 24 shows the residuals are separated graphically. The prediction of nitrification does not appear to be as decisive as through the Discriminant analysis (see Figure 25).

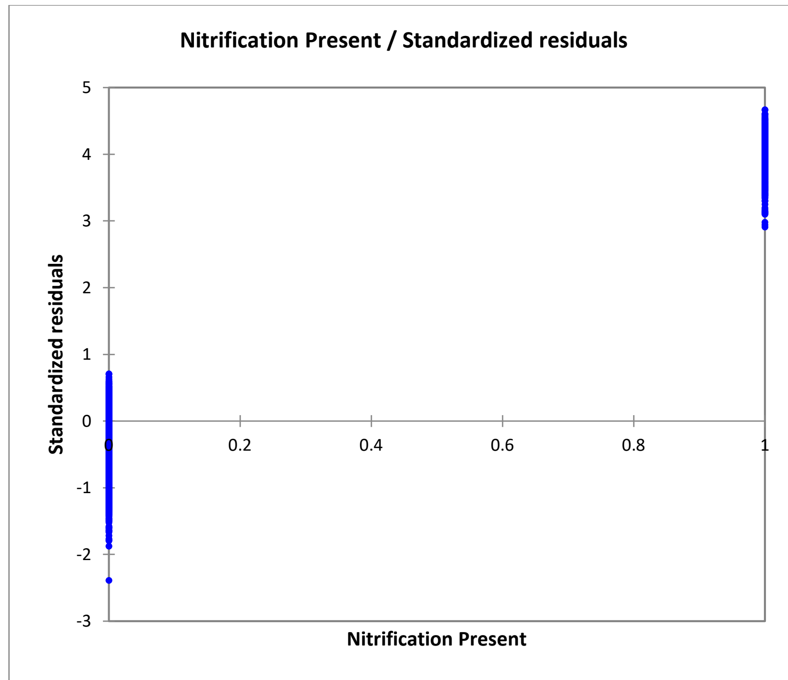


Figure 23 Differentiation of nitrification and non-nitrification results.

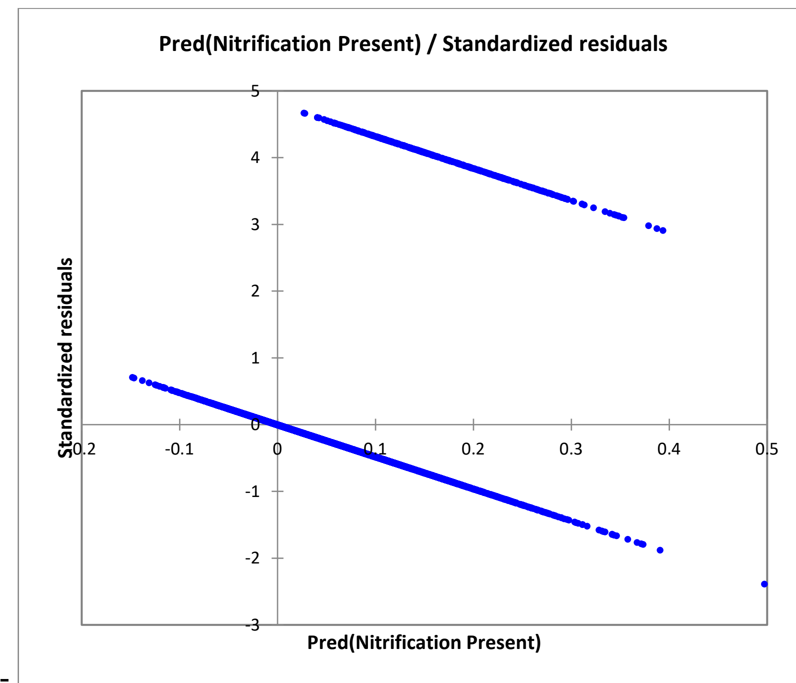


Figure 24 Standardized residuals show separation of t nitrification and non-nitrification samples

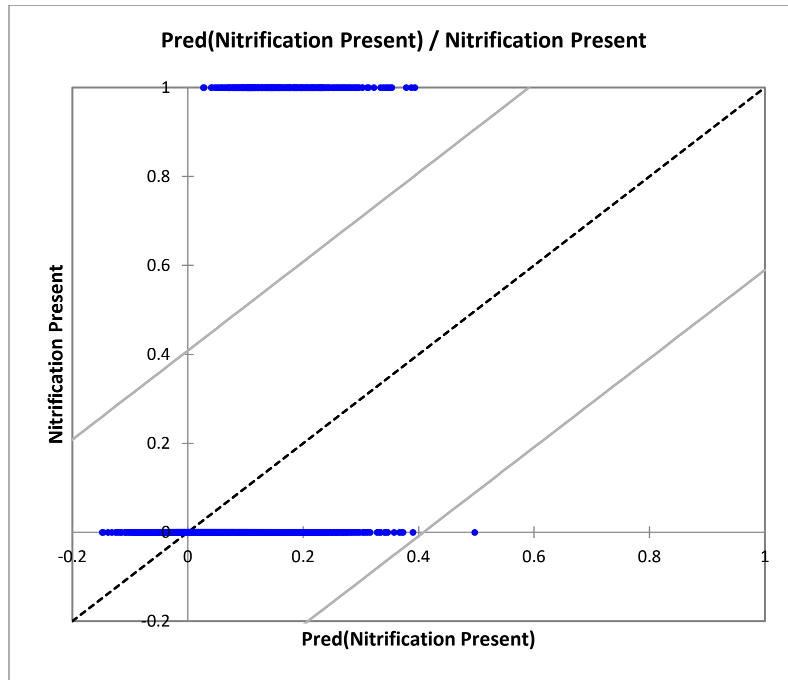


Figure 25 Prediction of Nitrification

Data Modification

Based on the fact that only 3 datasets were available for analysis, the investigators separated the data-set into two section: 2013/2014, and 2015-2017. The first data set was evaluated with principal component analysis a, discriminant analysis and linear regression to create the summary data, and then applied it in a predictive sense to the 2015 to 2017 data. In addition certain things from the larger analysis were changed. Since the initial principal component analysis indicated that the nitrogen species were all correlated in Factor 1, nitrate and ammonia were retained, but all others were eliminated as redundant. In addition, further analysis of the data showed that many results listed as "0" for ammonia and nitrate were actually points where the analysis was not collected. As a result, these points were eliminated from the data (cutting the total of 16,031 data points to under 8,500. The dissolved oxygen (DO) was measured after the sample is taken. DO can be introduced

into the analysis during sampling so is spurious in this data set. Therefore DO was also eliminated. Temperature is more relevant to DO anyway. The values for E. Coli and coliforms were primarily 0 throughout, and the full analysis demonstrated little value (since they appear infrequently). They were also eliminated. Finally, HPC has a detection limit of 20 CFU/100 mL. All values under 20 were revised to ½ the detection limit as is standard practice. The plants have different treatment as well, so were separated to a yes/no response.

Having made these changes, the PCA, DA and linear regression were run again using the XL Stat software. Table 17 shows the summary statistics. There are 2912 data points in the 2013/2014 dataset. However, the correlations remained poor (see Table 18). The PCA results were more favorable. The Scree plot (See Figure 26 indicated that the factors could be reduced to 8 as opposed to 12. Table 17 shows the factor loadings. The primary factors were:

- Factor 1 – nitrogen species/chlorine residual
- Factor 2 – PVC and DI pipe
- Factor 3 - pH
- Factor 4 – Water Plant 2
- Factor 5 –Water Plant 9
- Factor 6 – CAS
- Factor 7 – Temperature
- Factor 8 – CAS again

The factor loading table (Table 20) confirms these variables as the primary drivers. Correlations between variable (Table 21) and contributions of variables to the factor (Table 22) both confirmed these results as well.

Table 17 Summary Statistics 2013-2014 Data Only

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Nitrification Present	2912	0	2912	0.0	1.0	0.0	0.2
Collection_Date	2912	0	2912	2013.0	2014.0	2013.3	0.5
Water Plant 2	2912	0	2912	0.0	1.0	0.2	0.4
Water Plant 3	2912	0	2912	0.0	1.0	0.3	0.4
Water Plant 8	2912	0	2912	0.0	1.0	0.3	0.4
Water Plant 9	2912	0	2912	0.0	1.0	0.2	0.4
AC	2912	0	2912	0.0	1.0	0.0	0.2
CAS	2912	0	2912	0.0	1.0	0.0	0.1
CNC	2912	0	2912	0.0	1.0	0.0	0.1
DIP	2912	0	2912	0.0	1.0	0.3	0.4
PVC	2912	0	2912	0.0	1.0	0.7	0.5
DIAMETER	2912	0	2912	2.0	42.0	10.3	6.5
PH-FLD pH-Field	2912	0	2912	7.2	8.9	8.3	0.2
TEMP-FLD Temp-Field	2912	0	2912	19.0	32.5	25.7	1.9
TRESCG-FLD Total Chlorine Residual	2912	0	2912	0.0	5.0	3.1	0.8
NO3NO2 Nitrate+Nitrite as N	2912	0	2912	0.0	1.4	0.1	0.2
NH3 Ammonia	2912	0	2912	0.1	4.3	0.6	0.2
HPC	2912	0	2912	10.0	5700.0	39.6	274.5

Table 18 Correlation Matrix for 2013-2014 Data

Variables	llection_Da	/ater Plant	/ater Plant	/ater Plant	/ater Plant	AC	CAS	CNC	DIP	PVC	DIAMETER	-FLD pH-Fi	-FLD Temp	al Chlorine	Nitrate+Ni	H3 Ammon	HPC
Collection	**1.000**	-0.025	0.015	0.036	-0.030	-0.023	0.027	0.019	0.024	-0.022	0.022	-0.005	0.096	-0.003	0.177	-0.083	0.058
Water Pla	-0.025	**1.000**	-0.338	-0.335	-0.279	0.319	0.138	-0.042	0.092	-0.218	0.024	0.146	0.019	-0.023	0.061	-0.034	-0.018
Water Pla	0.015	-0.338	**1.000**	-0.392	-0.327	-0.108	-0.047	0.123	-0.143	0.163	-0.044	-0.267	0.019	0.173	-0.179	0.074	-0.039
Water Pla	0.036	-0.335	-0.392	**1.000**	-0.324	-0.107	-0.046	-0.048	0.003	0.051	0.065	0.367	-0.024	-0.267	0.248	-0.104	0.065
Water Pla	-0.030	-0.279	-0.327	-0.324	**1.000**	-0.089	-0.038	-0.040	0.060	-0.013	-0.048	-0.258	-0.015	0.126	-0.138	0.068	-0.010
AC	-0.023	0.319	-0.108	-0.107	-0.089	**1.000**	-0.013	-0.013	-0.105	-0.254	-0.086	0.227	-0.044	0.142	-0.016	0.040	-0.009
CAS	0.027	0.138	-0.047	-0.046	-0.038	-0.013	**1.000**	-0.006	-0.045	-0.110	0.111	0.071	0.068	-0.055	0.096	-0.027	-0.006
CNC	0.019	-0.042	0.123	-0.048	-0.040	-0.013	-0.006	**1.000**	-0.048	-0.115	0.378	-0.042	0.001	0.061	-0.036	0.003	-0.007
DIP	0.024	0.092	-0.143	0.003	0.060	-0.105	-0.045	-0.048	**1.000**	-0.908	0.258	-0.078	0.015	-0.074	0.080	-0.024	0.026
PVC	-0.022	-0.218	0.163	0.051	-0.013	-0.254	-0.110	-0.115	-0.908	**1.000**	-0.298	-0.011	-0.010	0.019	-0.081	0.013	-0.019
DIAMETER	0.022	0.024	-0.044	0.065	-0.048	-0.086	0.111	0.378	0.258	-0.298	**1.000**	0.049	0.004	-0.009	0.042	-0.034	0.010
PH-FLD pH	-0.005	0.146	-0.267	0.367	-0.258	0.227	0.071	-0.042	-0.078	-0.011	0.049	**1.000**	-0.007	0.370	-0.201	0.183	-0.087
TEMP-FLD	0.096	0.019	0.019	-0.024	-0.015	-0.044	0.068	0.001	0.015	-0.010	0.004	-0.007	**1.000**	0.017	0.040	0.023	-0.034
TRESCG-FI	-0.003	-0.023	0.173	-0.267	0.126	0.142	-0.055	0.061	-0.074	0.019	-0.009	0.370	0.017	**1.000**	-0.583	0.419	-0.155
NO3NO2 N	0.177	0.061	-0.179	0.248	-0.138	-0.016	0.096	-0.036	0.080	-0.081	0.042	-0.201	0.040	-0.583	**1.000**	-0.468	0.135
NH3 Amm	-0.083	-0.034	0.074	-0.104	0.068	0.040	-0.027	0.003	-0.024	0.013	-0.034	0.183	0.023	0.419	-0.468	**1.000**	-0.098
HPC	0.058	-0.018	-0.039	0.065	-0.010	-0.009	-0.006	-0.007	0.026	-0.019	0.010	-0.087	-0.034	-0.155	0.135	-0.098	**1.000**

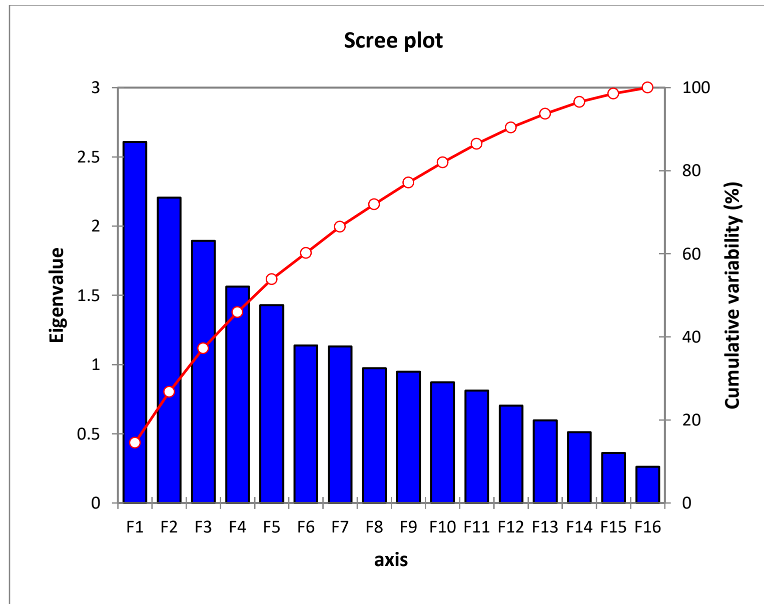


Figure 26 Scree Plot Revision - 2013-2014 Data Only

Table 19 Components of Factors from the Scree Plot

	F1	F2	F3	F4	F5	F6	F7	F8
Nitrification Present	0.205	0.207	0.262	-0.188	0.060	-0.050	0.154	-0.177
Collection_Date	0.176	0.036	-0.108	0.060	-0.131	-0.108	-0.514	0.346
Water Plant 2	0.050	-0.294	0.175	0.549	-0.154	0.086	0.031	-0.113
Water Plant 3	-0.226	0.124	-0.344	-0.108	-0.466	-0.383	0.163	0.124
Water Plant 8	0.272	0.200	0.442	-0.368	0.047	-0.036	0.015	0.058
Water Plant 9	-0.098	-0.052	-0.281	-0.045	0.614	0.369	-0.225	-0.083
AC	-0.076	-0.208	0.270	0.382	-0.086	-0.118	-0.136	-0.315
CAS	0.078	-0.084	0.035	0.141	-0.181	0.486	0.124	0.656
CNC	-0.016	-0.123	-0.104	-0.289	-0.424	0.319	-0.252	-0.439
DIP	0.224	-0.502	-0.095	-0.224	0.192	-0.295	0.168	0.123
PVC	-0.197	0.591	0.005	0.102	-0.055	0.198	-0.089	-0.032
DIAMETER	0.114	-0.304	-0.033	-0.366	-0.279	0.319	-0.142	-0.050
PH-FLD pH-Field	-0.139	-0.058	0.600	-0.091	-0.061	0.047	-0.105	0.154
TEMP-FLD Temp-Field	0.052	0.008	-0.093	0.042	-0.067	0.294	0.588	-0.090
TRESCG-FLD Total Chlorine Res	-0.487	-0.158	0.106	-0.095	0.000	0.000	-0.106	0.112
NO3NO2 Nitrate+Nitrite as N	0.494	0.118	-0.061	0.134	-0.075	0.015	-0.016	-0.013
NH3 Ammonia	-0.394	-0.100	0.124	-0.166	0.082	-0.037	0.109	0.127
HPC	0.129	0.028	-0.056	0.039	0.028	-0.154	-0.316	0.093

Table 20 Factor Loading

Factor loadings:								
	F1	F2	F3	F4	F5	F6	F7	F8
Nitrification Present	0.331	0.307	0.360	-0.235	0.072	-0.053	0.164	-0.174
Collection_Date	0.284	0.053	-0.148	0.075	-0.157	-0.115	-0.547	0.341
Water Plant 2	0.080	-0.436	0.241	0.687	-0.184	0.092	0.033	-0.111
Water Plant 3	-0.365	0.184	-0.473	-0.135	-0.557	-0.408	0.173	0.123
Water Plant 8	0.440	0.297	0.608	-0.460	0.056	-0.039	0.016	0.057
Water Plant 9	-0.158	-0.077	-0.386	-0.056	0.734	0.393	-0.239	-0.082
AC	-0.123	-0.309	0.372	0.477	-0.103	-0.125	-0.144	-0.311
CAS	0.125	-0.125	0.048	0.176	-0.216	0.519	0.132	0.647
CNC	-0.026	-0.182	-0.143	-0.362	-0.507	0.340	-0.268	-0.433
DIP	0.362	-0.746	-0.131	-0.280	0.229	-0.315	0.178	0.121
PVC	-0.318	0.877	0.007	0.128	-0.066	0.211	-0.095	-0.032
DIAMETER	0.185	-0.451	-0.045	-0.458	-0.334	0.341	-0.151	-0.049
PH-FLD pH-Field	-0.224	-0.086	0.826	-0.113	-0.072	0.050	-0.112	0.151
TEMP-FLD Temp-Field	0.084	0.012	-0.129	0.053	-0.080	0.313	0.625	-0.089
TRESCG-FLD Total Chlorine Residu	-0.787	-0.235	0.146	-0.119	0.000	0.000	-0.112	0.110
NO3NO2 Nitrate+Nitrite as N	0.798	0.175	-0.085	0.168	-0.090	0.016	-0.017	-0.013
NH3 Ammonia	-0.637	-0.148	0.171	-0.208	0.098	-0.040	0.116	0.125
HPC	0.209	0.041	-0.078	0.049	0.034	-0.165	-0.336	0.091

Table 21 Correlations between Variables within Factors

Correlations between variables and factors:								
	F1	F2	F3	F4	F5	F6	F7	F8
Nitrification Present	0.331	0.307	0.360	-0.235	0.072	-0.053	0.164	-0.174
Collection_Date	0.284	0.053	-0.148	0.075	-0.157	-0.115	-0.547	0.341
Water Plant 2	0.080	-0.436	0.241	0.687	-0.184	0.092	0.033	-0.111
Water Plant 3	-0.365	0.184	-0.473	-0.135	-0.557	-0.408	0.173	0.123
Water Plant 8	0.440	0.297	0.608	-0.460	0.056	-0.039	0.016	0.057
Water Plant 9	-0.158	-0.077	-0.386	-0.056	0.734	0.393	-0.239	-0.082
AC	-0.123	-0.309	0.372	0.477	-0.103	-0.125	-0.144	-0.311
CAS	0.125	-0.125	0.048	0.176	-0.216	0.519	0.132	0.647
CNC	-0.026	-0.182	-0.143	-0.362	-0.507	0.340	-0.268	-0.433
DIP	0.362	-0.746	-0.131	-0.280	0.229	-0.315	0.178	0.121
PVC	-0.318	0.877	0.007	0.128	-0.066	0.211	-0.095	-0.032
DIAMETER	0.185	-0.451	-0.045	-0.458	-0.334	0.341	-0.151	-0.049
PH-FLD pH-Field	-0.224	-0.086	0.826	-0.113	-0.072	0.050	-0.112	0.151
TEMP-FLD Temp-Field	0.084	0.012	-0.129	0.053	-0.080	0.313	0.625	-0.089
TRESCG-FLD Total Chlorine Resid	-0.787	-0.235	0.146	-0.119	0.000	0.000	-0.112	0.110
NO3NO2 Nitrate+Nitrite as N	0.798	0.175	-0.085	0.168	-0.090	0.016	-0.017	-0.013
NH3 Ammonia	-0.637	-0.148	0.171	-0.208	0.098	-0.040	0.116	0.125
HPC	0.209	0.041	-0.078	0.049	0.034	-0.165	-0.336	0.091

Table 22 Contribution Percent of Each Variable to Factor

Contribution of the variables (%):	F1	F2	F3	F4	F5	F6	F7	F8
Nitrification Present	4.192	4.268	6.847	3.547	0.364	0.247	2.371	3.126
Collection_Date	3.088	0.129	1.159	0.358	1.728	1.171	26.470	11.983
Water Plant 2	0.245	8.621	3.057	30.163	2.364	0.746	0.096	1.268
Water Plant 3	5.119	1.543	11.813	1.174	21.708	14.656	2.660	1.546
Water Plant 8	7.410	4.005	19.504	13.537	0.222	0.133	0.021	0.335
Water Plant 9	0.959	0.272	7.878	0.199	37.690	13.588	5.069	0.694
AC	0.578	4.321	7.290	14.561	0.747	1.384	1.842	9.941
CAS	0.602	0.704	0.121	1.982	3.259	23.664	1.535	43.040
CNC	0.026	1.506	1.084	8.378	17.987	10.187	6.338	19.291
DIP	5.025	25.209	0.904	5.035	3.689	8.702	2.818	1.511
PVC	3.876	34.898	0.002	1.049	0.304	3.925	0.800	0.105
DIAMETER	1.310	9.215	0.108	13.401	7.804	10.208	2.031	0.248
PH-FLD pH-Field	1.926	0.334	35.987	0.820	0.367	0.220	1.102	2.357
TEMP-FLD Temp-Field	0.273	0.006	0.874	0.179	0.445	8.620	34.538	0.818
TRESCG-FLD Total Chlorine Residu	23.743	2.498	1.130	0.900	0.000	0.000	1.116	1.253
NO3NO2 Nitrate+Nitrite as N	24.408	1.396	0.378	1.798	0.565	0.024	0.024	0.017
NH3 Ammonia	15.548	0.998	1.545	2.768	0.676	0.140	1.182	1.609
HPC	1.672	0.076	0.318	0.151	0.081	2.386	9.987	0.858

Table 23 Correlations between Variables and Factors after Varimax Rotation

Correlations between variables and factors after Varimax rotation:	D1	D2	D3	D4	D5	D6	D7	D8
Nitrification Present	-0.107	-0.030	0.163	-0.059	0.040	-0.004	0.005	-0.003
Collection_Date	-0.113	0.016	-0.015	-0.007	0.008	0.014	0.982	0.014
Water Plant 2	-0.024	0.088	-0.068	0.965	0.127	-0.018	-0.005	0.068
Water Plant 3	0.181	-0.106	-0.581	-0.461	0.599	0.053	0.027	0.001
Water Plant 8	-0.244	0.004	0.827	-0.276	0.152	-0.014	-0.019	-0.059
Water Plant 9	0.091	0.020	-0.189	-0.184	-0.947	-0.024	-0.003	-0.007
AC	0.056	0.055	0.033	0.171	0.024	-0.014	-0.023	-0.006
CAS	-0.045	0.011	0.001	0.062	0.006	0.014	0.014	0.991
CNC	0.031	-0.042	-0.113	-0.045	0.031	0.900	0.016	-0.043
DIP	-0.040	0.979	-0.002	0.024	-0.033	-0.029	0.011	-0.065
PVC	0.020	-0.954	0.008	-0.087	0.017	-0.117	-0.007	-0.082
DIAMETER	-0.018	0.330	0.182	0.034	0.025	0.713	0.008	0.130
PH-FLD pH-Field	0.410	-0.070	0.738	0.116	0.172	-0.006	0.005	0.089
TEMP-FLD Temp-Field	-0.036	0.005	-0.030	0.007	0.009	0.004	-0.041	0.036
TRESCG-FLD Total Chlorine Resid	0.831	-0.049	0.028	-0.067	-0.036	0.043	0.062	-0.042
NO3NO2 Nitrate+Nitrite as N	-0.821	0.058	0.106	0.006	0.060	-0.026	0.176	0.077
NH3 Ammonia	0.757	0.018	-0.017	-0.013	0.006	-0.049	-0.171	0.025
HPC	-0.070	0.014	0.004	-0.006	-0.004	-0.001	0.037	-0.004

Table 24 Contribution of Variables in percent after Varimax Rotation

Contribution of the variables (%) after Varimax rotation:	D1	D2	D3	D4	D5	D6	D7	D8
Nitrification Present	0.510	0.044	1.571	0.261	0.120	0.001	0.002	0.001
Collection_Date	0.565	0.013	0.013	0.003	0.005	0.015	93.278	0.020
Water Plant 2	0.027	0.389	0.277	70.575	1.207	0.023	0.002	0.451
Water Plant 3	1.460	0.555	19.970	16.102	26.927	0.205	0.068	0.000
Water Plant 8	2.657	0.001	40.359	5.752	1.720	0.014	0.036	0.334
Water Plant 9	0.372	0.020	2.115	2.568	67.156	0.044	0.001	0.004
AC	0.142	0.150	0.066	2.204	0.042	0.014	0.049	0.003
CAS	0.089	0.006	0.000	0.288	0.002	0.014	0.019	94.632
CNC	0.043	0.088	0.759	0.154	0.074	60.391	0.024	0.175
DIP	0.072	47.566	0.000	0.043	0.079	0.063	0.012	0.402
PVC	0.017	45.206	0.004	0.574	0.020	1.026	0.005	0.651
DIAMETER	0.015	5.401	1.953	0.085	0.045	37.817	0.007	1.638
PH-FLD pH-Field	7.471	0.244	32.133	1.027	2.223	0.003	0.003	0.761
TEMP-FLD Temp-Field	0.059	0.001	0.053	0.004	0.007	0.001	0.163	0.123
TRESCG-FLD Total Chlorine Resic	30.747	0.121	0.046	0.343	0.097	0.140	0.371	0.172
NO3NO2 Nitrate+Nitrite as N	30.009	0.169	0.664	0.002	0.271	0.050	2.988	0.569
NH3 Ammonia	25.530	0.015	0.017	0.013	0.003	0.176	2.835	0.062
HPC	0.216	0.010	0.001	0.002	0.001	0.000	0.135	0.001

As was the case with the full dataset, the discriminant analysis was able to discern separate centroids for the figures (See Figure 27 and Figure 28). However the Ma were very close to 0.4, which is not as useful as anticipated. Table 25 shows that the results predicted the correct answer nearly 100% of the time.

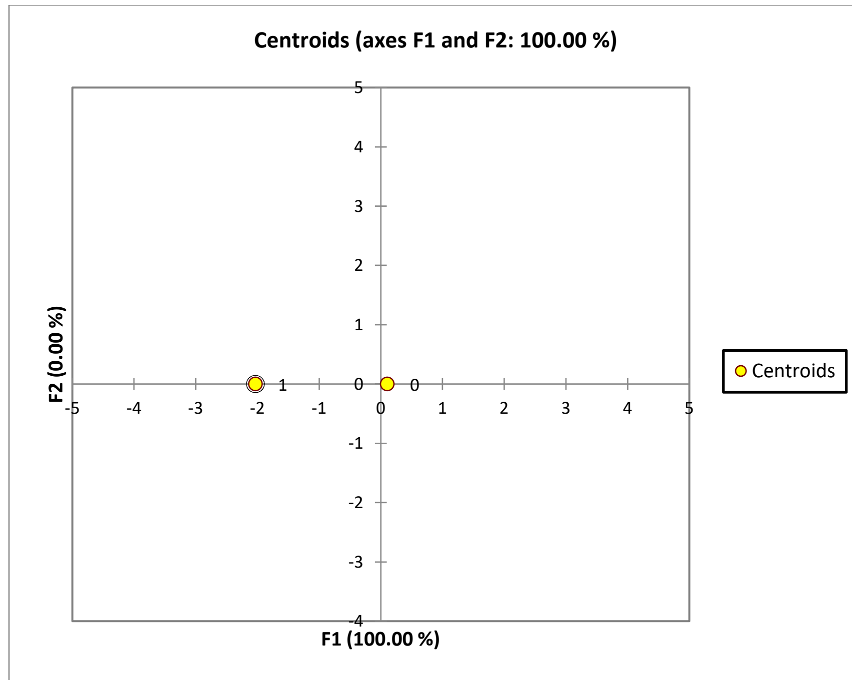


Figure 27 Centroids for the Plots for Nitrification

Table 25 Centroids for the Plots for Nitrification

0	0.107
1	2.033

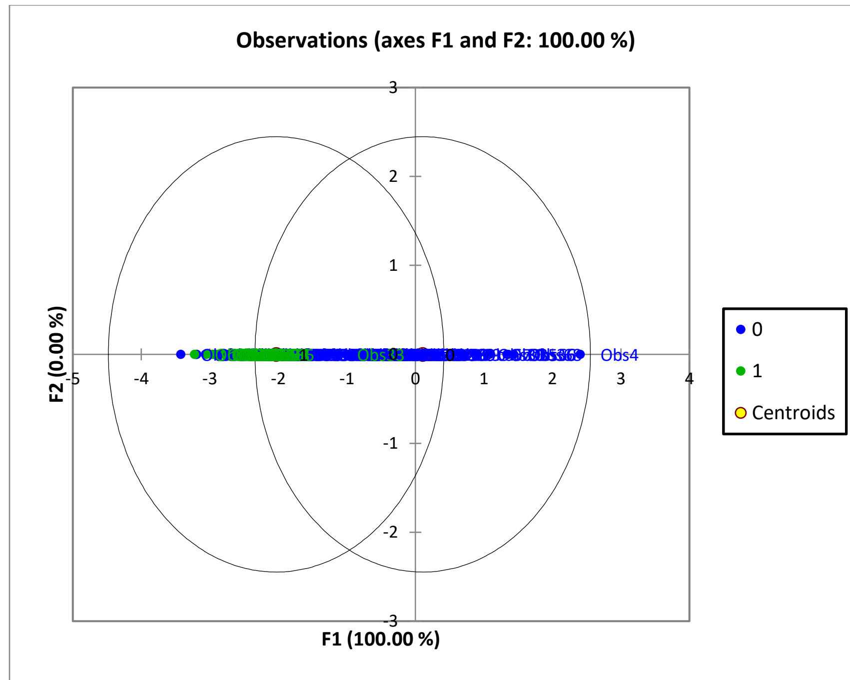


Figure 28 Universe for Observations for Nitrification

Table 26 Confusion Matrix for the Estimation Sample

Confusion matrix for the estimation sample:

from/to	0	1	Total	% Correct
0	2766	1	2767	99.96%
1	144	0	144	0.00%
Total	2910	1	2911	95.02%

Finally the linear regression model was created. The standard coefficients are shown in Table 27 and Figure 29. The actual equation developed by XLStat was:

Nitrification Present = -0.405-.0203*Water Plant 2-0.0028*Water Plant 3+0.144*Water Plant 8+.0119*AC-.0172*CAS+0.133*CNC-.0093*DIP-.00337*DIAMETER+.072*PH-FLD pH-Field+.0026*TEMP-FLD Temp-Field-

.062*TRESCG-FLD Total Chlorine Residual-Field-.063*NO3NO2 Nitrate+Nitrite as N-.00875*NH3 Ammonia-.000016*HPC

Figure 30 shows the prediction of the results. It should be noted that Water Plant 9 did not enter into the equation. An issue that arises is that the equation develops many answer less than 1. This was confirmed when the equation was applied against the 2015/2017 data (See Figure 31). The R value was 0.35.

Table 27 Standardized Coefficients

Table x+7 Standardized coefficients:

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Water Plant 2	-0.039	0.024	-1.670	0.095	-0.085	0.007
Water Plant 3	-0.006	0.022	-0.264	0.792	-0.049	0.037
Water Plant 8	0.296	0.027	10.842	< 0.0001	0.242	0.349
Water Plant 9	0.000	0.000				
AC	0.009	0.019	0.498	0.619	-0.027	0.046
CAS	-0.006	0.017	-0.324	0.746	-0.040	0.029
CNC	0.047	0.019	2.504	0.012	0.010	0.083
DIP	-0.019	0.018	-1.038	0.299	-0.055	0.017
PVC	0.000	0.000				
DIAMETER	-0.102	0.020	-5.207	< 0.0001	-0.140	-0.063
PH-FLD pH-Field	0.071	0.024	2.929	0.003	0.023	0.118
TEMP-FLD Temp-Field	0.022	0.017	1.313	0.189	-0.011	0.056
TRESCG-FLD Total Chlorine Residual-Field	-0.228	0.025	-9.097	< 0.0001	-0.277	-0.179
NO3NO2 Nitrate+Nitrite as N	-0.049	0.023	-2.097	0.036	-0.094	-0.003
NH3 Ammonia	-0.008	0.020	-0.428	0.669	-0.047	0.030
HPC	-0.020	0.017	-1.196	0.232	-0.054	0.013

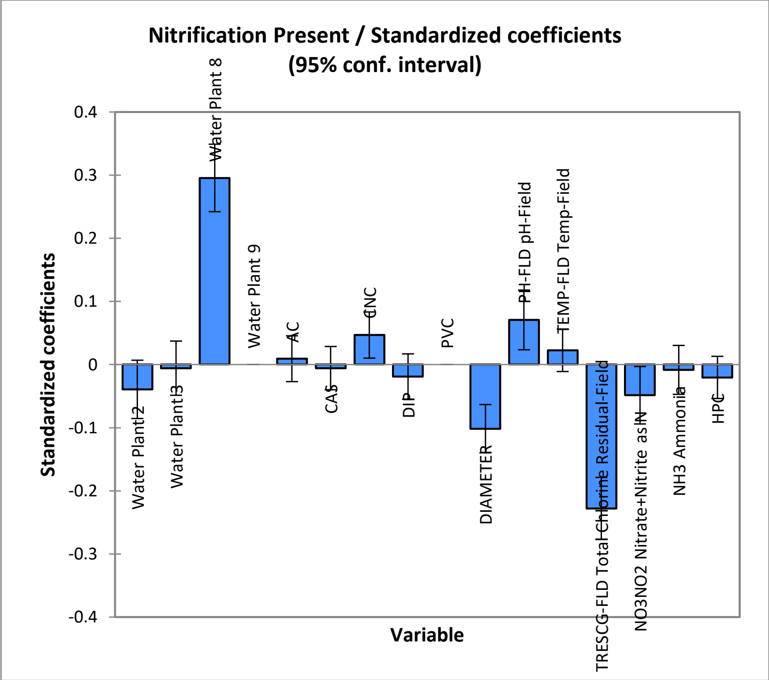


Figure 29 Coefficients for Linear Regression Mode

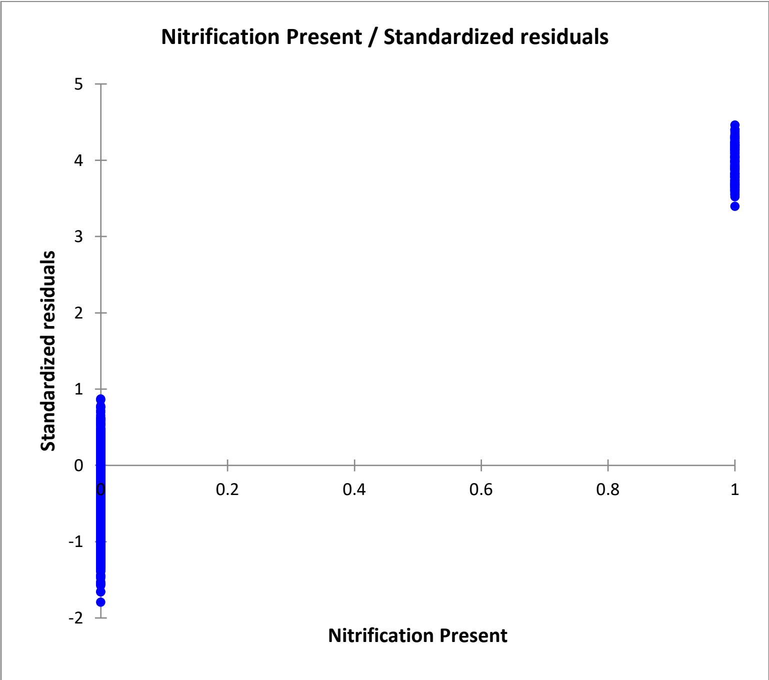


Figure 30 Prediction of Results Comparing Nitrification to Areas Without

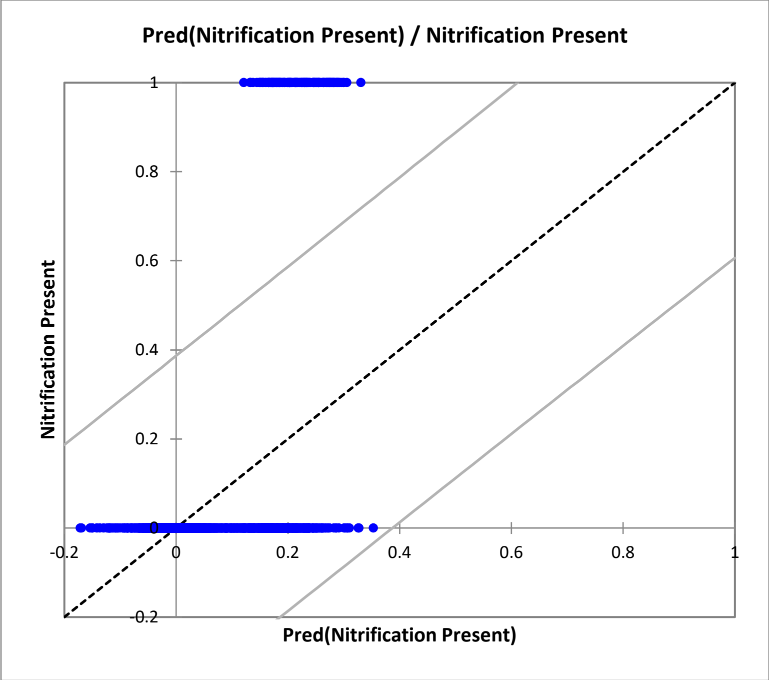


Figure 31 Prediction of the Model

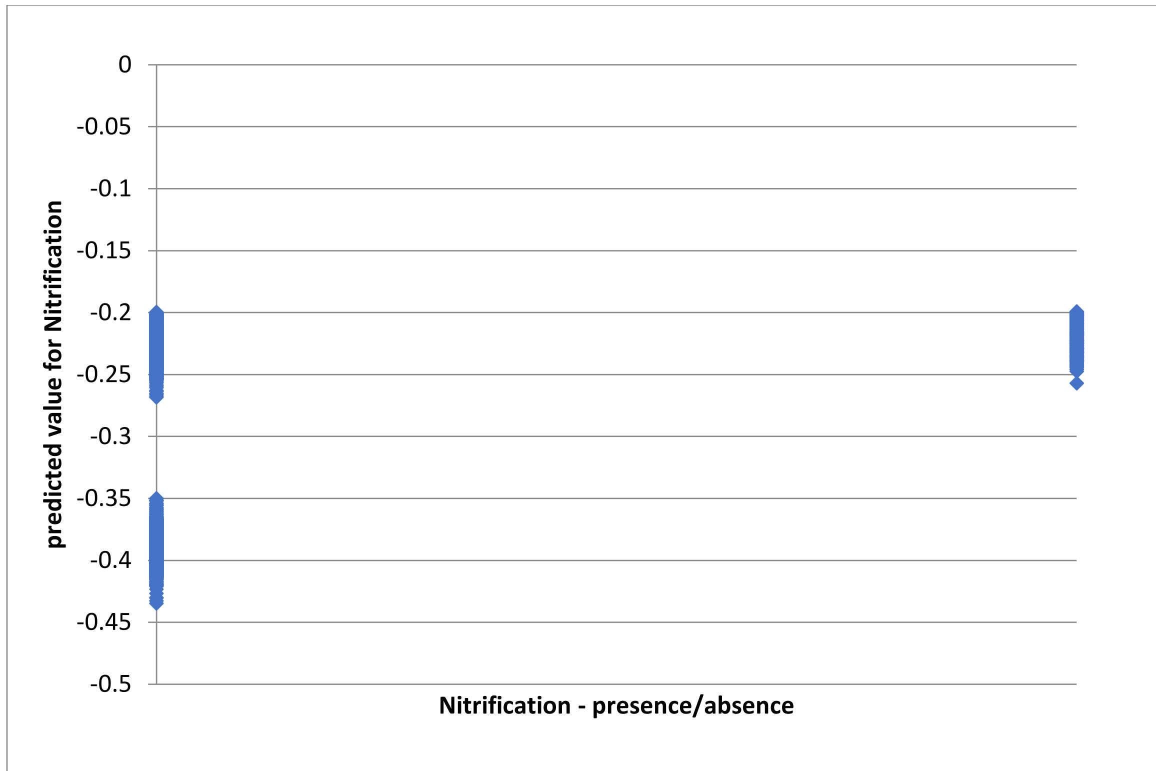


Figure 32 Results of Linear Regression Equation when Applied to 2015-2017 Data

Conclusions:

The result of this analysis indicated that Discriminant analysis provide the best predictive ability for nitrification in the distribution system. Issue that are important to predicting nitrification were:

- Water plant
- Free ammonia
- Nitrite
- Residual chlorine (negative)
- Ammonia (negative) and
- Dissolved oxygen (negative)

Ammonia use, especially when ammonia is quenched prior to rechloramination, should be avoided. While not a major factor in this analysis, pH does matter. A future project should focus on water plants with lower pH values. The plants were important however in the overall analysis. The utility tested has four water treatment plants that operate by three methods.

The first water treatment plant is known as water plant 8. The system is the largest of the four facilities. Water plant 8 is a thirty million gallon per day lime softening plant. They operate two Eimco softeners and one Accelator softener. At the time the initial lab samples were completed they were operating two-thirds of the plant using ozone and conventional non-biological filters and the other third employed a Tonka Anion Exchange fixed bed resin system for color removal. The thing that sets this facility apart from the other lime softening plant is the use of Ferric Chloride to aid in coagulation in the softeners. The facility observed increased clarity and less carryover into the filters using this method.

The next facility is also a lime softening plant identified as water plant 2. This facility is rated at 16.4 million gallons per day. This facility employee a Permuda softener to soften the water. They also have a Miex system to remove organic before the softener. This removes a lot of the contaminants prior to softening making the lime usage lower. This water plant does not use ferric chloride in their softener. They also have conventional non-biologic filters to assist in water treatment.

The last two facilities are sister plants and are identical. They are identified as water plant 3 and 9. Both of these facilities are nanofiltration plants. The treatment trains are multistage units and both facilities are rated at thirty million gallons per day. They use sand filters to remove any sand prior to the water entering the treatment trains. The water travels through the various stages of the nanofiltration membranes before it is passed through a degasification tower to remove hydrogen sulfide from the treated water.

All of the facilities employee the use of surficial aquifer wells that are approximately 150 deep. Most reside on the Biscayne aquifer. They all use chloramines to facilitate primary disinfection. The ages of the facilities vary and with upgrades that have occurred over time, judging their age is difficult. The oldest built plant is water plant 2 which was put in service in 1972. Next, water plant 8 was placed into service in 1982. Water plant 3 and 9 were originally built in the 1970's but in 2006 the utility completed replaced the plants to what exists today.

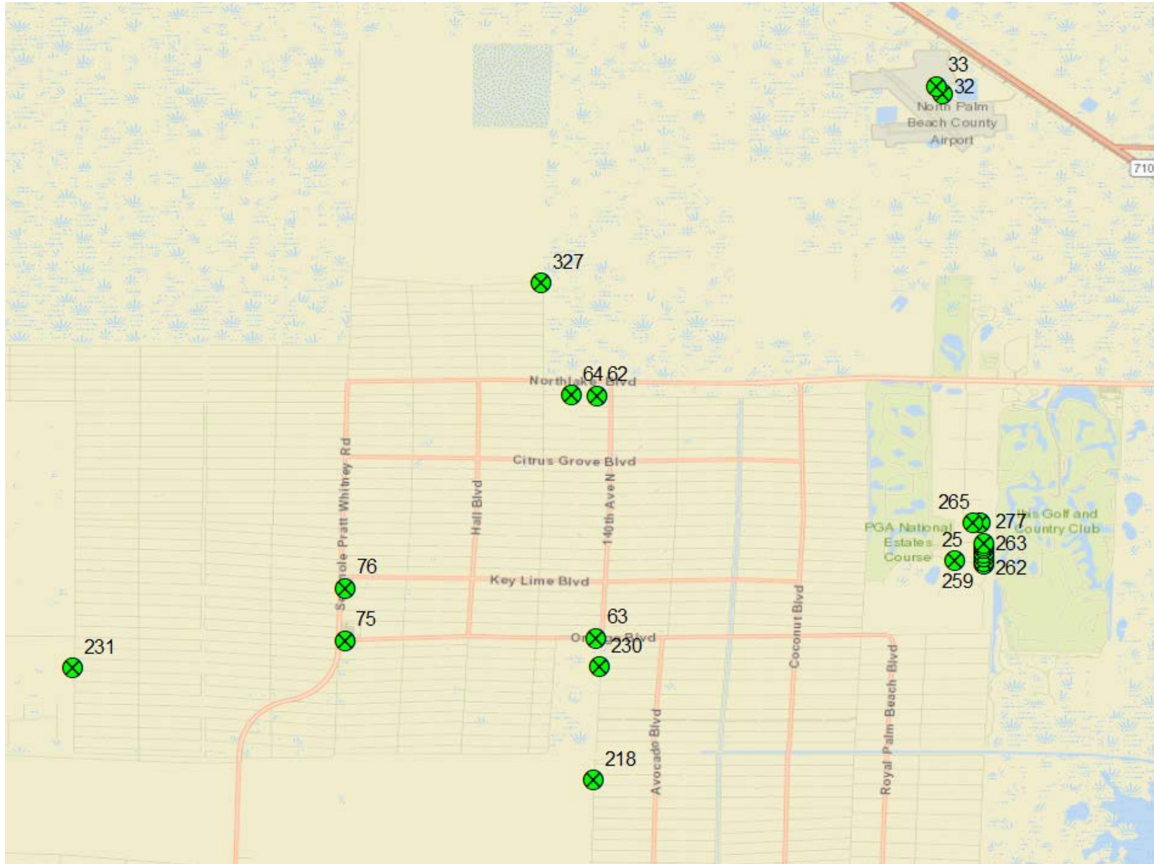


Figure 33 System 8 North

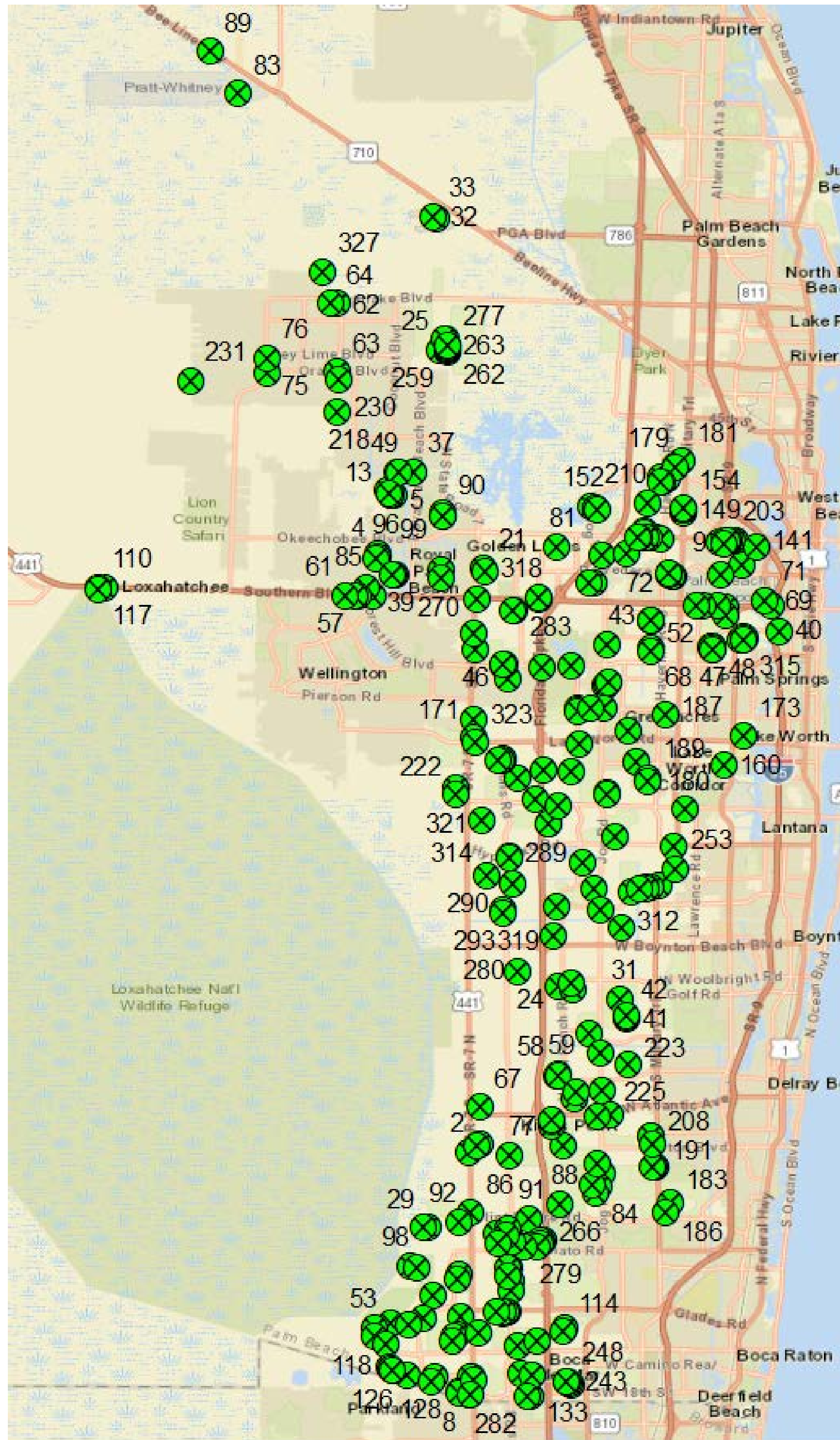


Figure 34 All Utility Sampling Locations

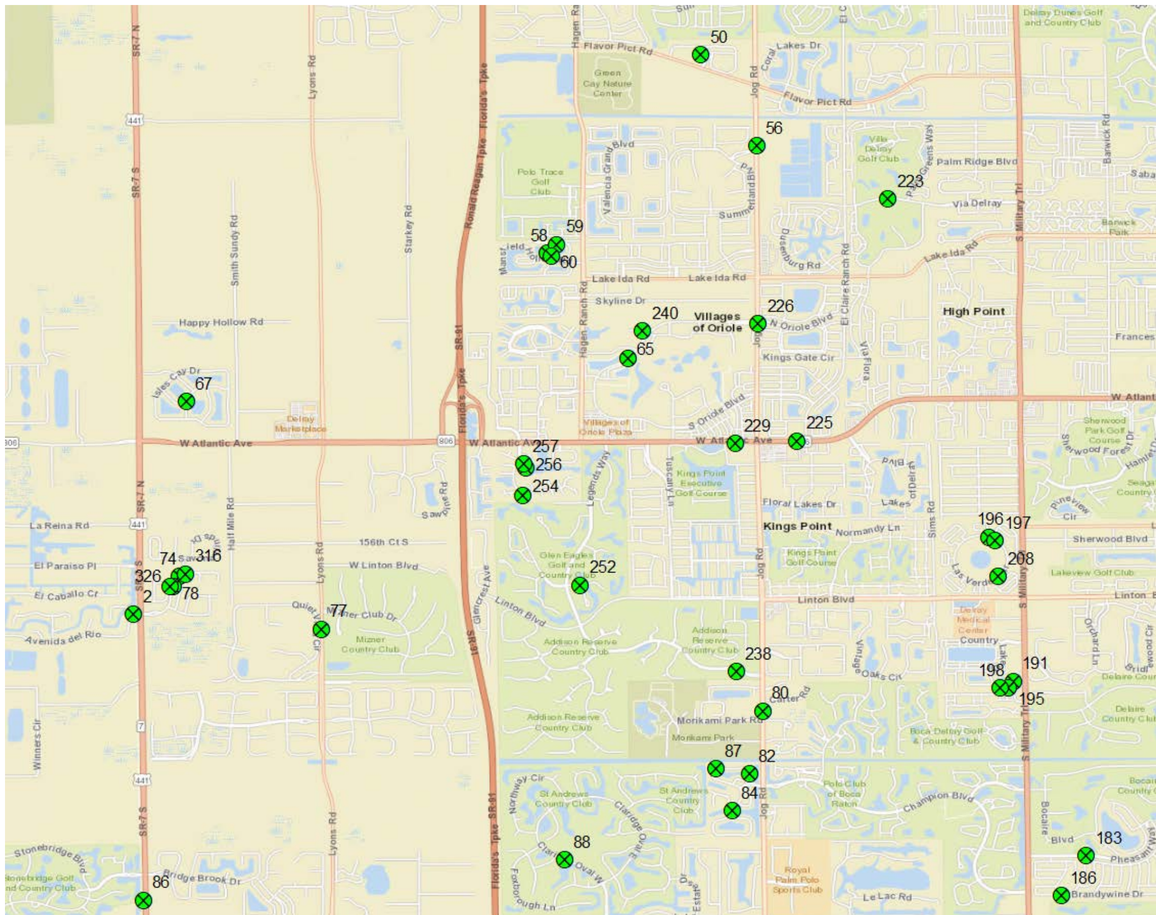


Figure 35 System 3 Sampling Locations

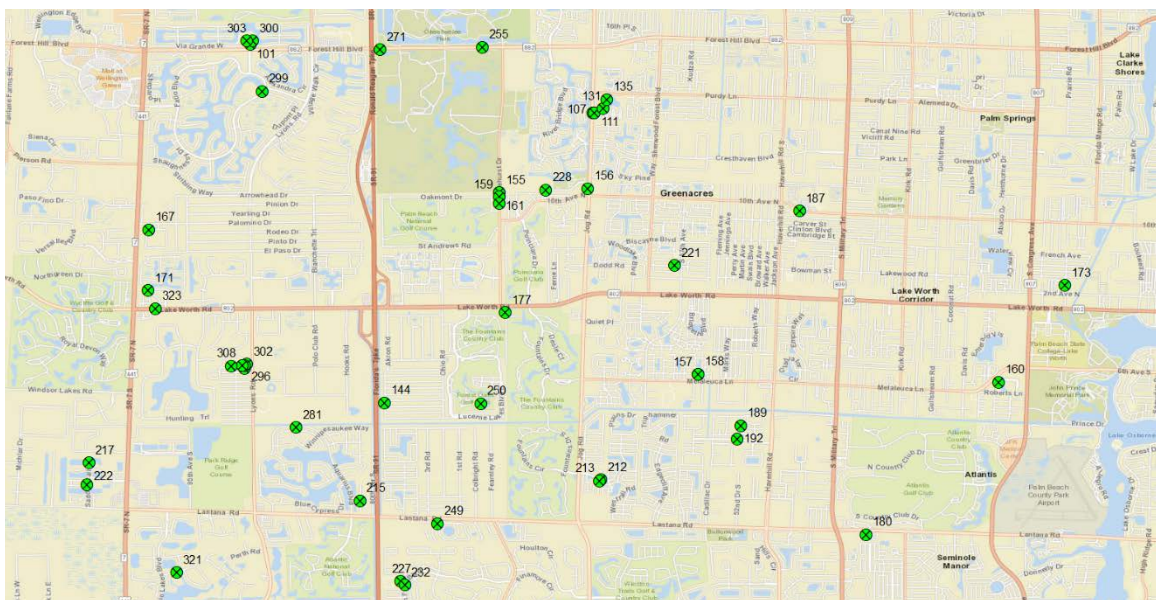


Figure 36 System 2 Sampling Locations

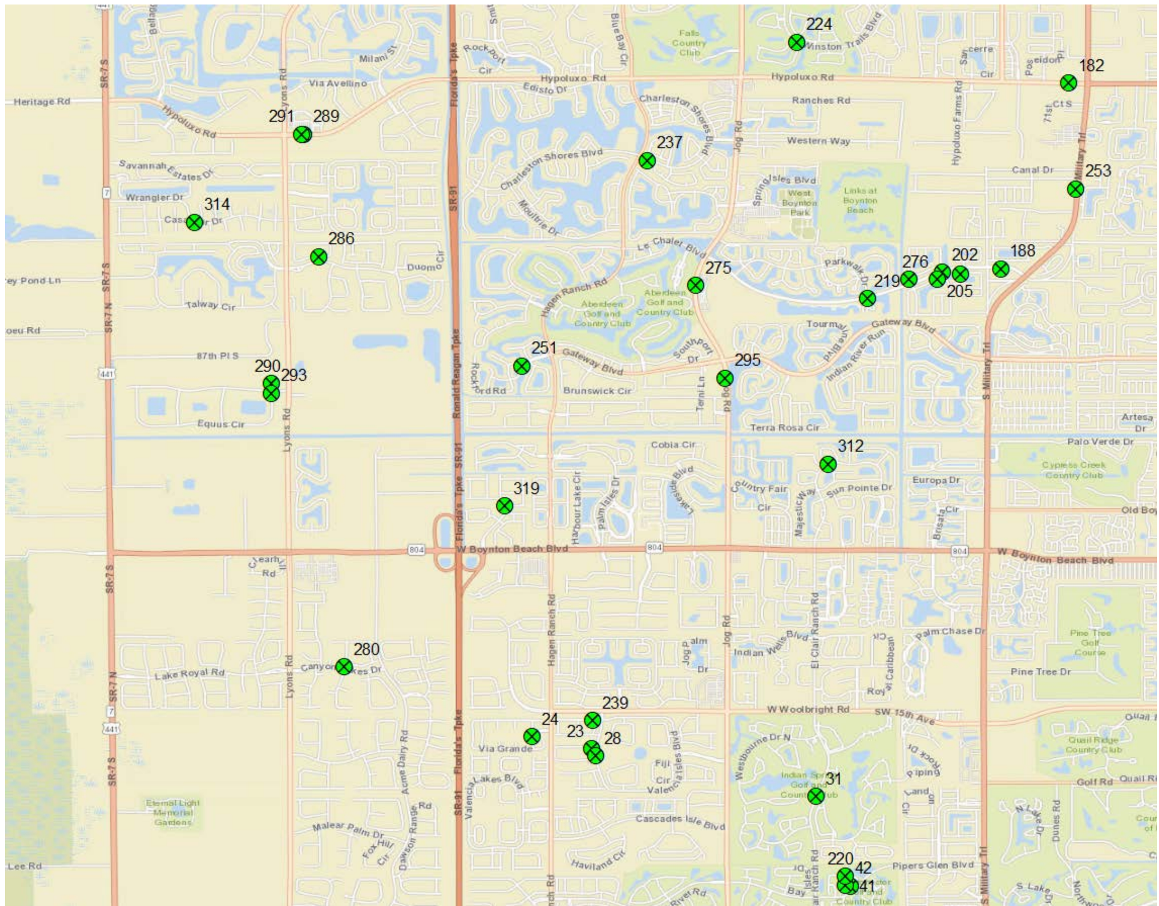


Figure 37 System 3 South Sampling Locations

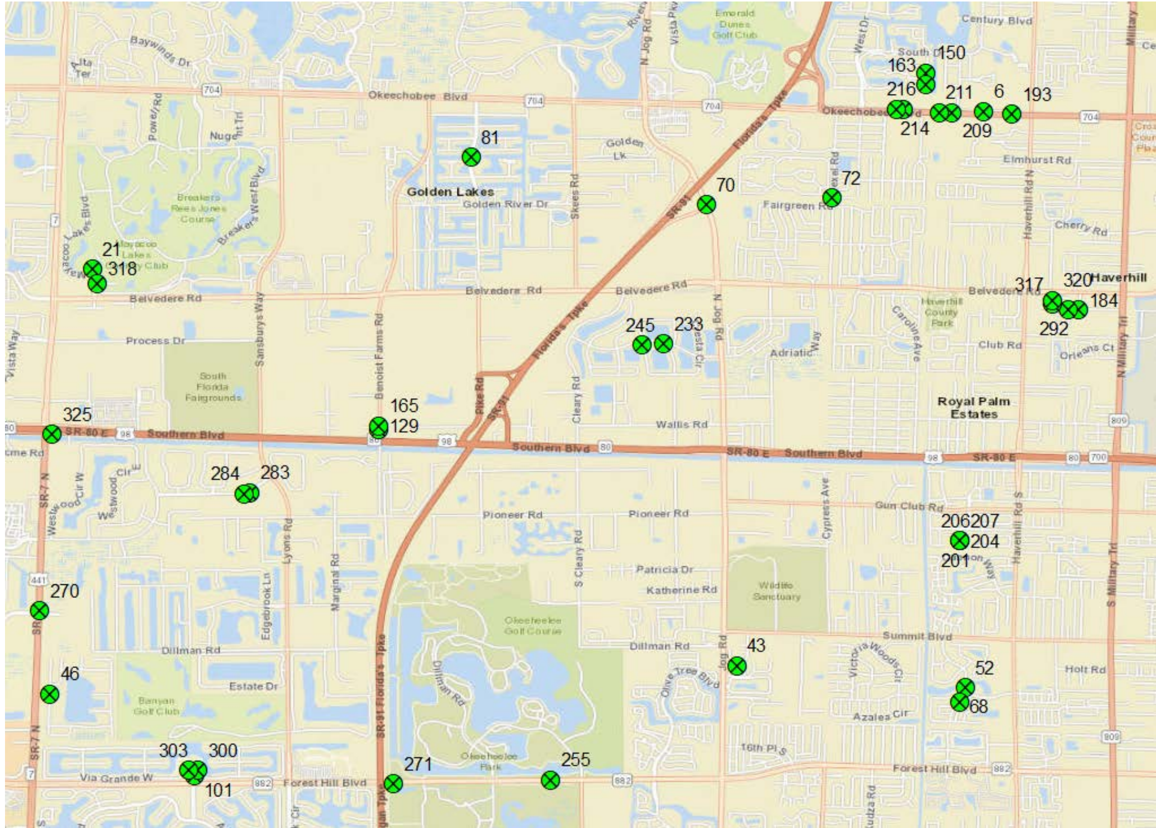


Figure 38 System 8 Sampling Locations

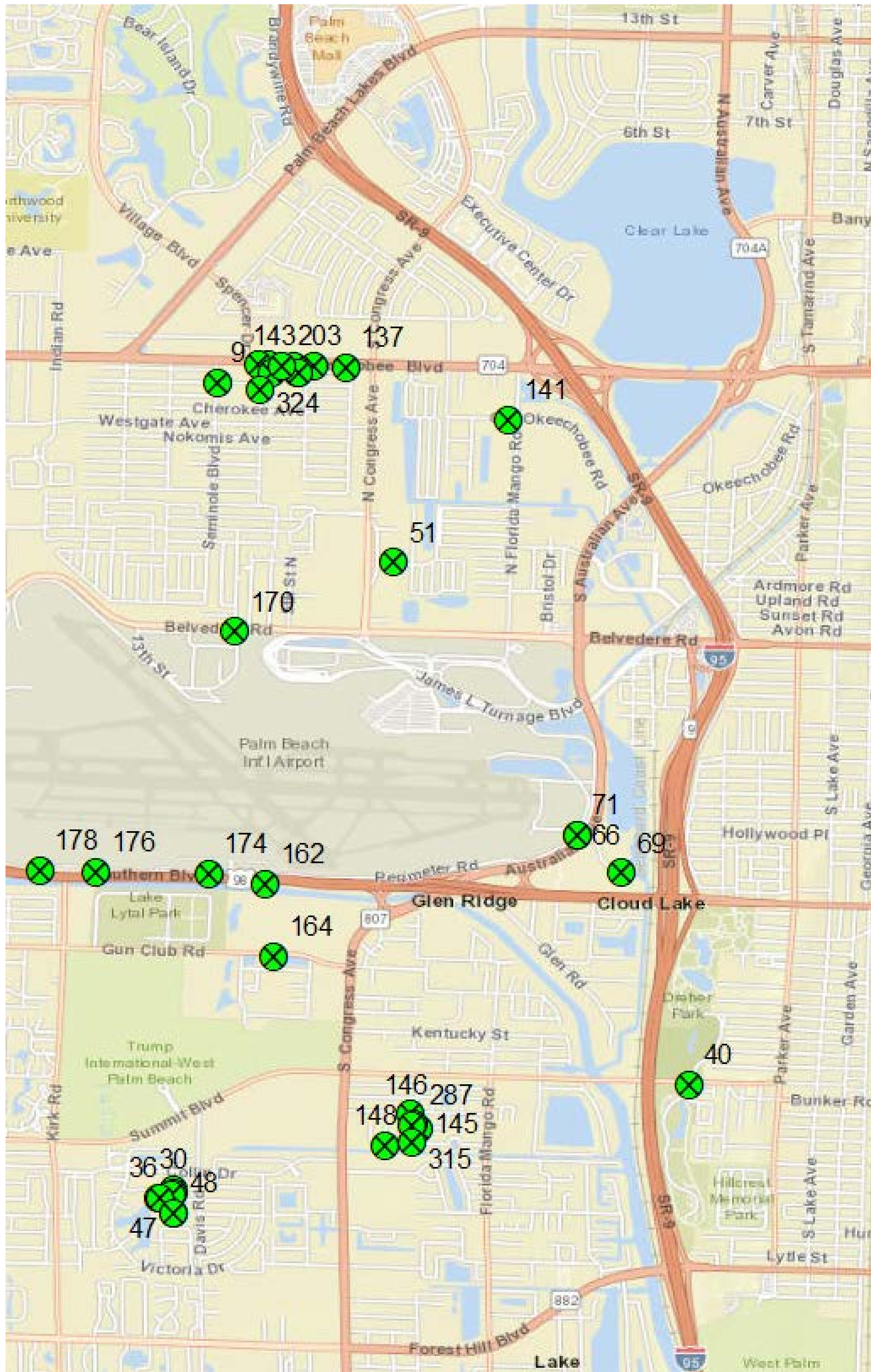


Figure 39 System 8 East Sampling Locations

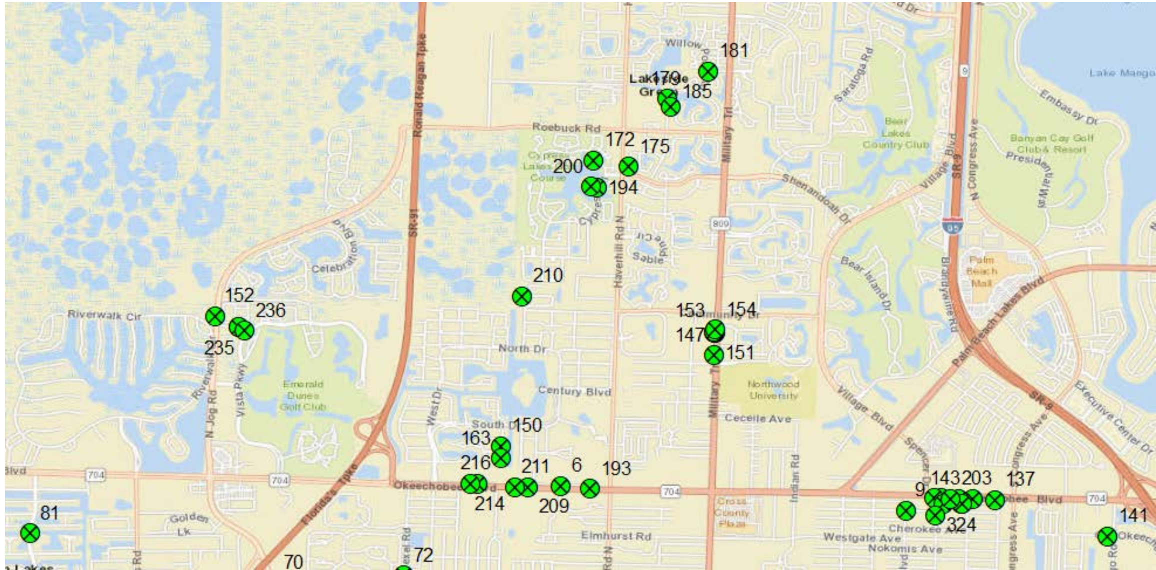


Figure 40 System 8 North Sampling Locations

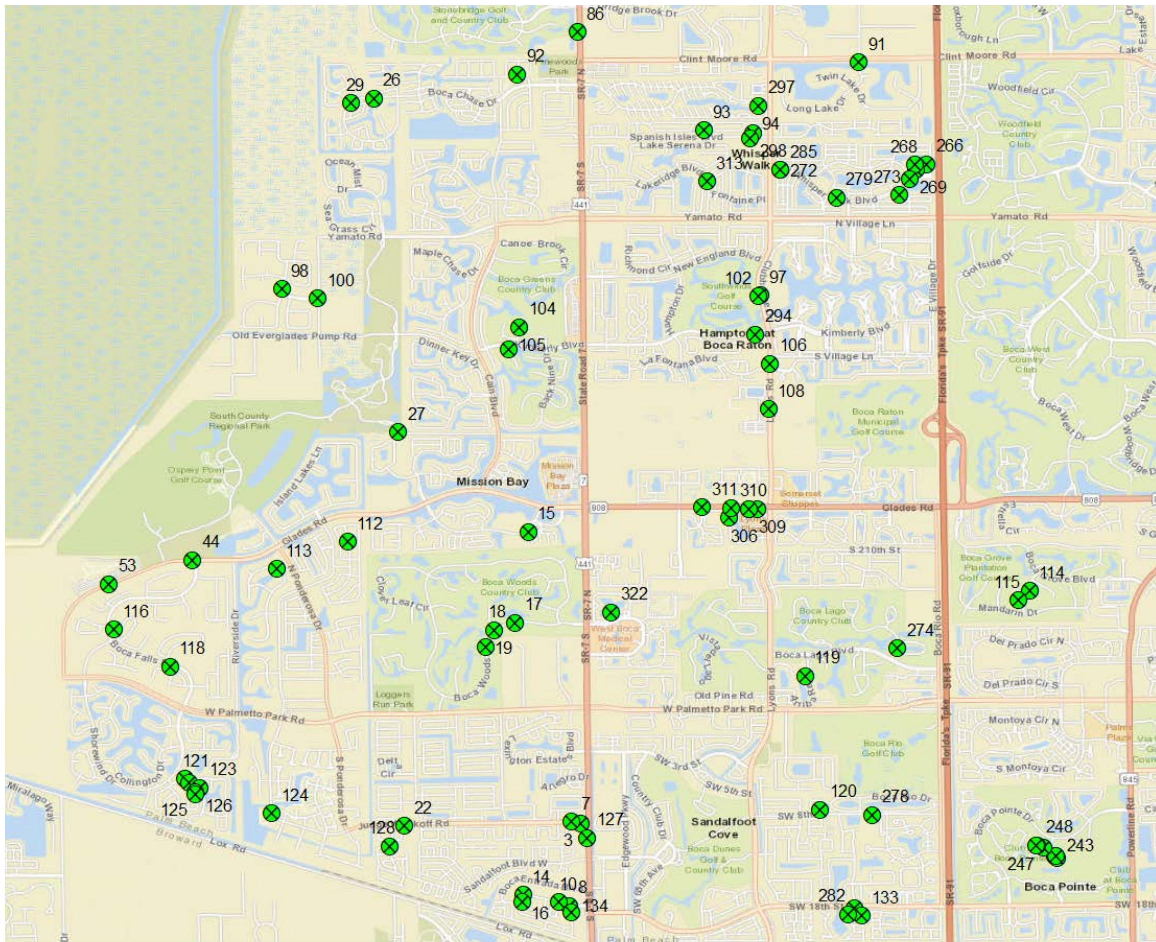


Figure 41 System 9 Sampling Locations

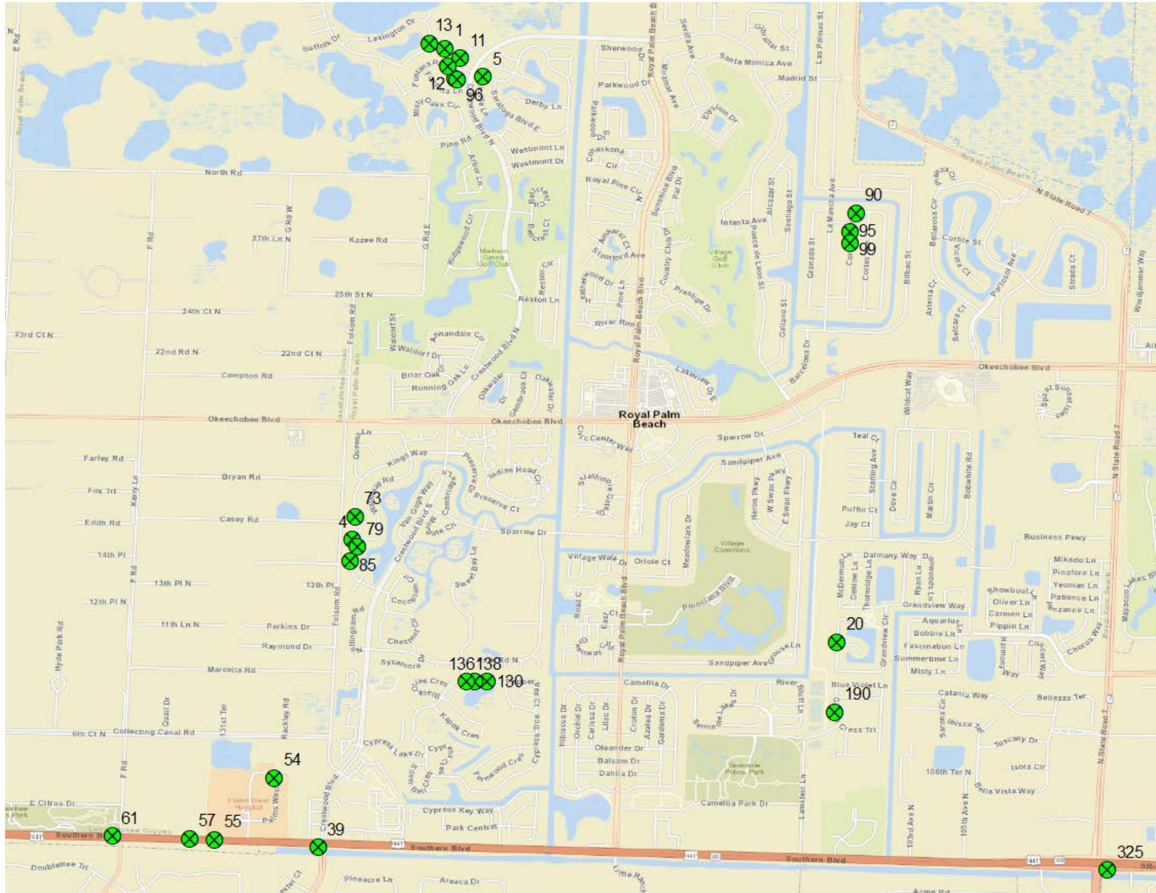


Figure 42 System 8 West Sampling Locations

References

American Water Works Association. (1992). *Standard for Disinfecting Water Mains.* Denver, CO: AWWA.

American Water Works Association. (2012). *Computer Modeling of Water DIstribution Systems* (3 ed.). (L. Robinson, J. A. Edwards, & L. D. Willnow, Eds.) Denver, CO, US: American Water Works Association.

American Water Works Association. (2013). *Nitrification - Prevention and Control in Drinking Water* (2 ed.). (J. Routt, & J. Skadsen, Eds.) Denver, CO, US: American Water Works Association.

APHA, AWWA, and WEF. (2005). *Standard Methods for the Examination of Water and Wastewater* (21st ed.). Washington DC: American Public Health Association.

Asbestos.comLLC. (2019, April 10). *Types of Asbestos*. Retrieved from Asbestos.com: https://www.asbestos.com/asbestos/types/

Bachmann, R. T., & Edyyean, R. J. (2006). Biofouling: An Historic and Contemporary Review of its Causes, COnsequences, and Control in Drinking Water Distribution Systems. *Biofilms, 2*, 197.

Bagh, L. K., Albrechtsen, H. J., Arvin, E., & Ovesen, K. (2002). Biofilm Formation in a Hot Water System. *Water Science Technology, 46*, 95.

Batte, M., Mathieu, L., Laurent, P., & Prevost, M. (2003). Influence of Phosphate and Disinfection on teh Composition of Biofilms Produced from Drinking Water, as

measured by Fluorescence in situ Hybridization. *Canadian Journal of Microbiology, 49*, 741.

Besner, M.-C., Gauthier, V., Barbeau, B., Millette, R., Chapleau, R., & Prevost, M. (2001, July). Understanding Distribution System Water Quality. *Journal AWWA*, 101-114.

Besner, M.-C., Gauthier, V., Servais, P., & Camper, A. (2002, August). Explaining the Occurance of Coliforms in Distribution Systems. *Journal AWWA, 94*(9), 95-109.

Blenkinsopp, S. A., & Costerton, J. w. (1991). Understanding Bacterial Biofilms. *Trends in Biotechnology, 9*(1), 138-143.

Bloetscher , F., Bullock, R. J., Fergen, R. E., Witt, G. M., & Fries, G. D. (2002). Selecting the Proper Grade of Stainless Steel is a Major Concern in Membrane Treatment Projects. *Florida Resource Journal*.

Bloetscher, F. (2017). *Review of Nitrification and Distribution System Water Quality.* Hollywood: Public Utility Management and Planning Services, Inc.

Bloetscher, F. (2019). *Infrastructure Management.* Plantatation: In Press.

Bloetscher, F., & Plummer, J. D. (2011). Evaluating the Significance of Certian Pharmaceuticals and Emerging Pathogens in Raw Water Supplies. *Envrionmental Practice, 13*(9).

Bloetscher, F., Bullock, R. J., Fergen, R. E., Witt, G. M., & Fries, G. D. (2002). Revisiting the Selection of Stainless Steel Grades in Water and Wastewater Treatement Envrionments. *Water Engineering and Management, 149*(5, 6, and 7), 36-44, 12-15, and 12-26.

Camper, A. K., Brastrup, K., Sandvig, A., Clement, J., Spencer, C., & Capuzzi, A. J. (2003, July). Effect of Distribution System Materials on Bacterial Growth. *Journal AWWA, 95*(7), 107-121.

Cantor, A. F., Bushman, J. B., Glodoski, M. S., Keifer, E., Bersch, R., & Wallencamp, H. (2006). Copper Pipe Failure by Microbiologically Influenced Corrosion. *Material Performance, 45*, 38.

Cartier, C., Besner, M.-C., Barbeau, B., Lavioe, J., Desjardins, R., & Prevost, M. (2009, July). Evaluating Aerobic Endospores as Indicators or Intrusion in Distribution Systems. *Journal AWWA, 101*(7), 46-57.

Cast Iron Soil Pipe Institute. (2006). *Cast Iron Soil Pipe and Fittings Handbook* (12th ed.). Chattanooga, TN: Cast Iron Soil Pipe Institute.

Cates, J., & Lavinder, S. (1999). Improving Chloramine Residuals and Minimizing Nitrification. *Florida Water Resources Journal, 2*, 26-28.

Chapelle, F. H. (1993). *Groundwater Microbiology and Geochemistry.* New York: Wiley and Sons.

Chowdhury, S., Campagne, P., & McLellan, P. J. (2010, June). Factorial Analysis of Trihalomethanes Formation in Drinking Water. *Water Envrionment Research, 82*, 1-11.

Cloete, T. E. (2003). Resistance Mechhanisms of Bacteria to Antimicrobial Compounds. *International Biodeterioration and Biogradation, 51*, 277-282.

Crozes, G. F., & Cushing, R. S. (2000). *Evaluatins Biological Regrowth in Distribution Systems.* Denver: American Water Works Association Research Foundation.

Deininger, R. A., Clark, R. M., Hess, A. F., & Bernstam, E. V. (1992). Animation and Visulization of Water Quality in Distribution Systems. *Journal AWWA*, 48-52.

Dillion, C. P. (1995). *Corrosion Resistance of Stainless Steels.* New York: Marcel Dekker, Inc.

Donlan, R. M. (2002, September). Biofilms: Microbial Life on Surfaces. *Emerging Infectious Diseases, 8*(9), 881-890.

D'Souza, C. D., & Kumar, M. (2010, July). Comparison of ANN Models for Predicting Water Quality in Distribution Systems. *Journal AWWA, 102*(7), 92-106.

Dunteman, G. H. (1989). *Principal Components Analysis* (07-069 ed.). (M. S. Lewis-Black, Ed.) Newbury Park, Ca, USA: Sage Publicaations.

Eboigbodin, K. E., Serh, A., & Biggs, C. A. (2008, October). A Review of Biofilms in Domestic Plumbing. *Journal AWWA, 100*(10), 131-138.

Eboigbodin, K. E., Seth, A., & Biggs, C. A. (2008). A Review of Biofilms in Domestic Plumbing. *Journal AWWA, 100*, 131.

Elton, A., Brammer, L. F., & Tansley, N. S. (1995, July). Water Quality Modeling in Distribution Networks. *Journal AWWA*, 44-52.

Emde, K. E., Smith, D. W., & Fancey, R. (1992). Inital Investigation of Microbially Influenced Corrosion (MIC) in Low Temperature Water Distribution Systems. *Water Resource Journal, 26*, 169.

Fass, S., Block, J. C., Boualam, M., Gautheir, V., Gatel, D., Cavard, J., . . . Lahoussine, V. (2003). Release of Organic Matter in a Discountinously Chlorinated Drinking Water Network. *Water Resource Journal, 37*, 493.

Geldreich, E. E. (1996). *Microbial Quality of Water Supply in Distribution Systems.* Boca Raton: CRC Press.

Gujer, W., & Jenkins, D. (1974). A Nitrification Model for Contact Stabilization Activated Sludge Process. *Water Resource Journal, 9*(5), 5.

Hall-Stoodley, L., & Stoodley, P. (2005). Biofilm Formation and Dispersal and the Transmission of Human Pathogens. *Trends in Microbiology, 13*, 7.

Herson, D. S., Marshall, D. R., Baker, K. H., & Victoreen, H. T. (1991, July). Association of Microorganisms with Surfaces in Distribution Systems. *Journal AWWA*, 103-106.

Heydorn, A., Ersboll, B., Kato, J., Hentzer, M., Parsek, M. R., Tolker-Neilsen, T., . . . Molin, S. (2002, April). Statistical Analysis of Pseudomonas Aeruginosa Biofilm Development: Impact to Mutations in Genees Involved in Twitching Motility, Cell-to-Cell Signaling, and Stationary-Phase Sigma Factor Expression. *Applied and Envrionmental Microbiology, 68*(4), 2008-2016. Retrieved February 22, 2019, from http://aem.asm.org

Hiltebrand, D. J., & Theiss, R. J. (2002, December). Distribution Systems Consider Water Quality adn Hydralic Quality. *Opflow*, 6-21.

Holt, D., Todd, R. D., Delanoue, A., & Colbourne, J. S. (1995). A Study of Nitrite Formulaiton and COntrol in Chloraminated Distribution Systems. *Water Quality Technology Conference.* Denver: American Water Works Associaiton.

Hwang, C. (2010, November). What's Inside the Hidden World of Biofilm? *Opflow*, 16-18.

IOS. (2012, May 29). *Creating a Scree Plot*. (I. O. Software, Producer) Retrieved from http://www.improvedoutcomes.com/docs/WebSiteDocs/PCA/Creating_a_Scree_Plot.htm

Jolliffe, I. T. (1972). Discarding Variables in a Principal Component Analysis. II : Real Data. *Journal of the Royal Statistical Society (Applied Statistics), 22*(1), 13. doi:10.2307/2346488

Kaiser, H. F. (1960). *The Application of Electronic Computers to Factor Analysis.* University of Illinois. Chicago: Educational and Psychological .

Kim, J.-o., & Mueller, C. W. (1978). *Introduction to Factor Analysis What it is and how to use it* (07-013 ed.). Newbury Park, CA, USA: Sage University Paper Series on Quaantitative Applications in Social Sciences.

Kirkmeyer, G. J., Friedman, M., Martel, K. D., Noran, P. F., & Smith, D. (2001). Practical Guidelines for Maintaining Distribution System Water Quality. *Journal AWWA, 93*, 62.

Kirkmeyer, G., Richards, W., & Smith, C. D. (1994). *An Assessment of Water Distribution Systems ans Association Research Needs.* Denver: American Water Works Association Research Foundation.

Lansey, K. E., & Boulos, P. F. (2005). *Comprehensive Handbook on Water Quality Analysis for Distribution Systems* (1 ed.). Pasadena, CA, US: MWH Soft.

LeChevallier, M. W., Lowry, C. D., & Lee, R. G. (1990). Disinfecting Biofilms in a Model Distribution System. *American Water Works Associaiton Journal*, 87-99.

LeChevallier, M. W., Lowry, C. D., & Lee, R. G. (1990). Disinfection of Biofilms in a Model Distribution System. *Journal Amernican Waater Works Association, 82*, 87.

Lee, D. G., & Kim, S. J. (2003). Bacterial Species in Biofilms Cultivated from the end of Seoul Waater Distribution System. *Applied Microbiology, 95*, 317.

Lin, Y.-s. E., Vidic, R. D., Stout, J. E., & Yu, V. L. (1998, September). Legionella in Water Distribution Systems. *Journal AWWA, 90*(9), 112-122.

Liu, S., Taylor, J. S., Randall, A. A., & Dietz, J. D. (2005, October). Nitrification Modeling in Chloraminated Distribution Systems. *Journal AWWA, 97*(10), 98-108.

Mahmood, F., Pimblett, J., Grace, N., & Utne, B. (1999). Combining Multiple Water Sources and Disinfectants: Options for Water Quality Compatibility in Distribution Systems. *AWWA Water Quality Technology Congerence* (p. 15). Tampa: AWWA.

Mains, C. (2008). Biofilm Control in Distribution Systems. *Technical Brief, 8*, 1.

Martin, B., & Ries, T. (2014, June). Getting Optimized - Partnership for Safe Water Guidelines - Flushing Maintians Distribution System Water Quality. *Opflow*, 8-9.

Mays, L. W. (1999). *Water Distribution System Handbook* (1st ed.). (L. Hager, Ed.) New York, New York: McGraw-Hill.

McLaughlan, R. (2002). *Managing Waater Well Deterioration: IAH International Contributions to Hydrogeology.* CRC Press.

Mercer, K. (2011, March). Question of the Month - Advice for Small Systems - How Old is too Old for Distribution System Water? *Opflow*, 6-7.

Mesotehlioma Justic Network. (2019, May 22). *History of Asbestos*. (S. Kidd, Editor) Retrieved June 1, 2019, from Asbsetos.net: https://www.asbestos.net/asbestos/history/

Miller, R., Friedman, M., Koci, D., & Moore, R. (2013, September). Biofilm Control - Develop an Effective Strategy to Preserve Quality, Ensure Compliance. *Opflow*, 10-13.

Murga, R., Forster, T. S., Brown, E., Pruckler, J. M., Fields, B. S., & Donlan, R. M. (2001). Role of Biofilms in the Survival of Legionella Pneumophila in a Model Potable-Waater System. *Microbiology, 147*, 3121.

Muylwyk, Q., Smith, A. L., & MacDonald, J. A. (1999). Implications on Disinfection Regime when Joining Water Systems: A Case Study of Blending Chlorinated and Chloraminated Water. *AWWA Water Quality Technology Conference.* Tampa: AWWA.

Nicklow, J. W., & Boulos, P. F. (2007). *Essential Water and Wastewater Calculations for Engineers and Operators* (1 ed.). Broomsfield, CO, US: MWH Soft.

Niquette, P., Servais, P., & Savoir, R. (2000). Impacts of Pipe Materials on Densities of Fixed bacterial Biomass in Drinking Water Distribution System. *Water Resource Journal, 34*, 1952.

Norton, C. D., & LeChevallier, M. W. (2000). A Piolet Study of Bacteriological Population Changes Through Potable Water Treatement and Distribution. *Applied Envrionmental Mocrobiology, 66*, 268.

Ollos, P. J., Huck, P. M., & Slawson, R. M. (2003, January). Factors Affecting Biofilm Accumulation in Model Distribution Systems. *Journal AWWA, 95*(1), 87-97.

Pintar, K. D., Anderson, W. B., Smith, E. F., & Huck, P. M. (2005, July). Assessment of a Distribution System Nitrification Critical Threshold Concept. *Journal AWWA, 97*(7), 116-129.

Pleitez, F. (2012). *Predicting Removal Efficiency of Reverse Osmosis Membranes with REspect to Emerging Substances Using a Discriminant Function Analysis.* Boca Raton: Florida Atlantic University.

Rayburn, C., Blaha, F., Anderson, K., & Rotert, K. (2011, July). Research and Information Collection with an Eye on Distribution Systems. *Journal AWWA, 103*(7), 50-56.

Reilly, J. K., & Kippin, J. S. (1983). Relationaship of Bacterial Counts with Turbidity and Free Chlorine in Two Distribution Systems. *Journal American Water Works Association, 75*, 309.

Ribas, F., Perramon, J., Terradillos, A., Frias, J., & Lucena, F. (2000). The Pseucomonas Group as an Indicator of Potential Regrowth in Water Distribution Systems. *Journal of Applied Microbiology, 88*, 704-710.

Rodrigues, P., Esteves da Silva, J. C., & Antunes, M. G. (2007). Factorial Analysis of the Trihalomethanes Formation in Water Disinfection Using Chlorine. *Analytia Chimica Acta, 595*, 266-274.

Ruda, T. (1997, August). Microbial Regrowth and Distribution System Management. *Opflow, 23*(8), 4-6.

Sanks, R. L. (2005). *Water Transport.* Montana State University, Department of Civil Engineering. Oxford: Encylopedia of Life Support Systems.

Satory, D. P., & Holmes, P. (1997). Chlorine Sensitivity of Envrionmental, Distributions System, and Biofilm Coliforms. *Water Science Technologies, 35*, 289.

Schmeisser, C., Stockigt, C., Raasch, C., Wingender, J., Timmis, K. N., Wenderoth, D. F., . . . Streit, W. R. (2003). Metagenome Survey of Biofilms in Drinking Water Networks. *Applied Envrionmental Microbiology, 69*, 7298.

Schrantz, K. A., Pressman, J. G., & Wahman, D. G. (2013). Simulater Distribution Nitrification: Nitrification Index Evaluation and Viable AOB. *Journal AWWA, 105*(46), 242-254.

Schrempp, T., Goold, R., Kwan, B., & Sarai, D. (1994). Effect of Mechanical Cleaning and Free REsidual Chlorine on teh Chlorine Demand and Nitification Process in a Chloraminated System. *AWWA Water Quality Technology Conference.* San Francisco: AWWA.

September, S. M., Els, F. A., Venter, S. N., & Brozel, V. S. (2007). Prevalence of Bacterial Pathogens in Biofilms of Drinking Water Distribution Systems. *Journal of Water Health, 5*, 219.

Skadsen, J. (1993, July). Nitrification in a Distribution System. *Journal AWWA*, 95-103.

Stedmon, C. A., & Bro, R. (2008). Characterizing Dissolved Organic Matter Fluorescence with Parallel Factor Analysis: A Tutorial. *American Society of Limnology adn Oceanography*, 572-579.

Trussell, R. R. (1999, January). Safeguarding Distribution System Integrity. *Journal AWWA, 91*(1), 46-54.

U.S. Environmental Protection Agency. (2002). *Health Risks from Microbial Growth and Biofilms in Drinking Water Distribution Systems.* Office of Ground Water and Drinking Water. Washington: U.S. Environmental Protection Agency.

Uni-Bell PVC Pipe Association. (2012). *Handbook of PVC Pipe Design and Construction.* New York: Industrial Press.

US Envrionmental Protection Agency . (2002). *Nitrification.* Office of Ground Water and Drinking Water. Washington: US Envrionmental Protection Agency .

US Envrionmental Protection Agency. (1989). *Drinking Water; National Primary Drinking Water Regulation; Filtration, Disinfection; Turbidity, Giardia Lamblia, Viruses, Legionella, and Hetertophic Bacteria; Final Rule.* Washington DC: Federal Register.

US Envrionmental Protection Agency. (1992). *Control of Biofilm Growth in Drinking Water Distribution Systems.* Office of Research and Development. Washington DC: US Envrionmental Protection Agency.

US Envrionmental Protection Agency. (1999). *Legionella. Human Health Criteria Document.* Office of Water, Office of Science and Technology. Washington DC: US Envrionmental Protection Agency.

US Envrionmental Protection Agency. (2002). *Deteriorating Buried Infrastructure Management Challenges and Strategies.* Office of Water, Office of Ground Water and Drinking Water Distribution Issue Paper. Washington DC: US Envrionmental Protection Agency.

US Envrionmental Protection Agency. (2002). *The Clean Water and Drinking Water Infrastructure Gaps Analysis.* Office of Waater. Washington DC: US Envrionmental Protection Agency.

Valentine, R. L., Ozekin, K., & Vikesland, P. J. (1998). *Chloramine Decomposition in Distribution System and Model Waters.* Denver: AWWA Research Foundation.

Van Der Kooij, D., & Veenendaal, H. R. (2001). Biomass Production Potential of Materials in COntact with Drinking Water: Methods and Practical Importance. *Water Science Technology Water Supply, 1*, 39.

Vanhaeecke, E., Remon, J. P., Moors, M., Raes, F., DeRudder, D., & Van Peteghem, A. (1990). Kinetics of Pseudomonas Aeruginosa Adhesion to 304 to 316L Stainless Steel: Roll of Cell Surface Hydrophobicity. *Applied and Envrionmental Microbiology, 56*, 788-795.

Verstraete, W., & Alexander, M. (1973). Heterotrophic Nitrification in Samples of Natural Ecosystems. *Envrionmental Science Technologies, 7*(39).

Videla, H. A. (1996). *Manual of Biocorrosion.* Boca Raton: CRC Press.

Wahman, D. G., Maestre, J. P., & Speitel Jr., G. E. (2016, July). Monochloramine Cometabolism by Nitrifying Biofilm Relevant to Drinking Water. *Journal AWWA, 108*(7), 362-373.

Walker , J. T., & Morales, M. (1997). Evaluation of Chlorine Dioxide (CLO2) for the Control of Biofilms. *Water Science Technology, 35*, 319.

Walski, T. M. (2006, March). A History of Water Distribution. *Journal AWWA, 98*(3), 110-121.

Watson, S. W., Valos, F. W., & Waterbury, J. B. (1981). The Family Nitrobacteraceae. In the Prokaryotes. (M. P. Starr, Ed.) *Springer-Verlag*.

Wilczak, A. (2001). Chloramine Decay Rate: Factors and Research Needs. *AWWA Annual Conference Procedings.* Washington DC: AWWA.

Wolfe, R. L., & Lieu, N. I. (2001). Nitrifying Bacteria in Drinking Water. In J. W. Sons, & G. Bitton (Ed.), *Encyclopedia of Envrionmental Microbiology.* New York, New York: John Wiley and Sons.

Wolfe, R. L., Leiu, N. I., Izaguirre, G., & Means, E. (1990). Ammonia Oxidizing Bacteria in a Chloraminated Distribution System: Seasonal Occurance, Distribution, and Disinfeection REsistance. *Applied Envrionmental Microbiology, 56*(2), 451-462.

Woolschlager, J. E., Rittmann, B. E., Piriou, P., & Schwartz, B. (2001). Developing an Effective Strategy to Control Nitrifier Growth Using the Comprehensive Disinfection and Water Quality Model. *World Water and Envrionmental Resources Congress.* Renton: ASCE.

World Health Organization. (2002). *The World Health Report 2002.* Geneva: World Health Organization.

World Health Organization. (2018). *World Health Statistics 2018: Monitoring Health for the SDG's, Sustainable Development Goals.* Geneva: World Health Organization.

XL Stat. (2017, January 1). *XLSTAT 2012.1*. Retrieved from XLSTAT by Addinsoft: https://www.xlstat.com/en/news/xlstat-2012-1

Zhang, M., Semmens, M. J., Schuler, D., & Hozalski, R. M. (2002, September). Biostability and Microbiological Quality in an Chloraminated Distribution System. *Journal AWWA, 94*(9), 112-122.

Zhang, Y., Griffen, A., & Edwards, M. (2010). Effect of Nitrification on Corrosion of Galvanized Iron, Copper, and Concrete. *Journal AWWA, 102*, 83.