DEVELOPMENT OF CONDITION PREDICTION MODELS

FOR SANITARY SEWER PIPES

by

MOHAMMADREZA MALEK MOHAMMADI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2019

Dedication

This work is dedicated to my parents, Manouchehr Malek Mohammadi and Afsaneh Behzadi, my beloved wife, Nazanin, and my brother Mehran Malek Mohammadi for their endless support, encouragement, patience and unconditional love.

Acknowledgements

What a journey I have been taking until this point when I write on this very page. It is impossible for me to survive this journey alone. First and foremost, I thank God for blessing me with this opportunity to pursue my dreams. For always giving me the best options, the courage to choose, funneled me through the consequences, but let me enjoy the rewards.

I would like to express my most sincere gratitude and appreciation to my academic advisor, and mentor Dr. Mohammad Najafi, P.E., F. ASCE, and Director of the Center for Underground Infrastructure Research and Education (CUIRE). I was fortunate to assist Dr. Najafi in several courses and research projects over the last four years and I am indebted for the opportunities he provided me throughout my graduate studies that helped me improve myself both academically and professionally. Dr. Najafi has always been a leader, a great motivator and was always supportive throughout all phases of my research. His insight has significantly assisted me in choosing my career path, presently and in the future.

I would like to thank my committee members; Dr. Sharareh Kermanshachi, Dr. Mostafa Ghandehari, and Dr. Edmund Prater for their valuable comments, suggestions and contributions to my dissertation. I greatly appreciate the support provided by Ms. Jamie Riley and Ms. Jessica Gruber, wastewater engineers in City of Tampa. Without their assistance in providing data and insightful comments, this study would have encountered more obstacles. I would like to acknowledge Mr. Chris Young, Dr. Ellen McDonald, and Mr. Dexter May for providing me the opportunity to work in Alan Plummer Associates, Inc. and apply my research in the wastewater industry during my last semester.

My heartfelt appreciation goes to my parents, Manouchehr Malek Mohammadi and Afsaneh Behzadi, who have always supported me in every step of my life. Their constant devotion and encouragement were among the most significant factors that helped me complete this challenging journey. I would like to thank my brothers, Mehran and Reza, for being supportive and understanding in times of hardships that I have experienced over the last four years. I also thank my parents-in-law, Mohammad Salehabadi, and Fereshteh Vosoughi for their hearted words and supports.

Last but not least, I would like to express my love and deepest appreciation to my wife, Nazanin, for her endless support, encouragement, understanding and patience. Her presence has and will always be the most important source of motivation as I embark on new challenges in life. Life is a long and challenging journey, but your presence makes it worthwhile. I know together we can make it to the end.

July 18, 2019

Abstract

DEVELOPMENT OF CONDITION PREDICTION MODELS

FOR SANITARY SEWER PIPES

Mohammadreza Malek Mohammadi, PhD

The University of Texas at Arlington, 2019

Supervising Professor: Dr. Mohammad Najafi

Utility managers and owners have challenges when addressing appropriate intervals for inspection of gravity sanitary sewer pipelines and other underground pipeline systems. Frequent inspection of sewer network is not cost-effective due to large inventory of pipes and high cost of inspections, such as using closed-circuit television (CCTV) surveys. Therefore, it would be more beneficial to first predict critical sewers most likely needing maintenance and then perform inspections to optimize use of their limited budgets and target pipelines most in need of repairs, rehabilitation or renewal. Development of sewer condition prediction models is extremely vital for utilities to evaluate the short-term and long-term behavior of their pipe network considering different uncertainties. However, providing a prediction model is difficult due to lack of adequate datasets. The primary objective of this dissertation is to develop prediction models that can forecast future conditions of sanitary sewer pipes. The outcomes of the models can be used to prioritize inspection and renewal needs of sanitary sewer pipes for polyvinyl chloride (PVC) and vitrified clay pipes (VCP). In addition, this dissertation identifies significant factors that affect deterioration of sanitary sewers. To achieve these objectives, three different statistical and artificial intelligence models, namely logistic regression, gradient boosting tree and K-nearest neighbors were developed in successive steps. Data collected from

City of Tampa (Florida) was used to demonstrate the applicability of the developed models. Thirteen independent variables including pipes age, material, diameter, flow rate, pipe segment length, depth, slope, soil type, pH, and sulfate content, and water table, soil hydraulic group and soil corrosivity were used to build these prediction models. The results of this dissertation show that performance of all three developed models were acceptable; however gradient boosting tree achieved a higher accuracy during validation process. Additionally, pipe age, length, diameter, material and water table are found to be significant variables influencing deterioration of sanitary sewers. This dissertation contributes to body of knowledge by developing condition prediction models that can be used as part of a comprehensive asset management system of sanitary sewers.

Table of Contents

List of illustrations

List of Tables

Chapter 1 Introduction and Background

1.1 Introduction

The U.S. underground pipeline systems includes thousands of miles and form a significant part of the total infrastructure (Najafi and Gokhale, 2005). Sanitary sewers as a part of the wastewater infrastructure systems, are designed to collect sewage from domestic, industrial, and commercial users and convey to treatment plants. Most sewer systems are gravity sewers, which transfer the flow based on a slope. There are over 800,000 miles of public sewer pipes and 500,000 miles of private sewer laterals in the United States. Approximately 240 million Americans are connected to 14,748 treatment plants for wastewater treatment. By 2032, it is estimated that 56 million more people will use centralized treatment plants (ASCE, 2017).

Majority of the U.S. wastewater infrastructure is more than 100 years old and the combination of aging, chemical and environmental factors cause at least 23,000 to 75,000 sanitary sewer overflows per year (EPA, 2004). The latest infrastructure report card, published by American Society of Civil Engineering (ASCE) in 2017, states a “D plus” grade for the wastewater infrastructure. ASCE indicated that water and wastewater systems in the U.S. are clearly aging and to keep-up with needs, a capital funding gap of $150 billion is needed by 2025 (ASCE, 2017). Furthermore, the U.S. population is increasing and shifting geographically. This requires investment for new infrastructure and maintaining existing infrastructure in areas of decreasing population with limited budgets (EPA, 2007).

According to AWWA (2012), some municipalities and agencies spend a relatively smaller investment for sewer rehabilitation rather than expanding sewer systems to meet growth and treatment plant upgrades. The risk of inflow and infiltration, sanitary sewer overflows, and sinkholes are increased by inadequate maintenance and deficient asset management practices. The consequence of not maintaining sanitary sewer systems may threaten human health as well as causing property damage and expensive emergency repairs (Kumar et al., 2018).

As stated earlier, sewer pipes constitute a major portion of wastewater systems, as they form the pathway between points of wastewater generation and treatment plants. As sewer system becomes older, the structural and operational performance may degrade. The aging of sewer pipes increases the failure rates and deteriorated pipes can result in social, environmental and economic impacts (Opila, 2011).

Maintenance and rehabilitation strategies are important factors to keep the performance of the system at an acceptable level of service and to provide cost-effective solutions for avoiding unforeseen failures. In the past, repair or rehabilitation of sewer pipes were only done once a pipe collapsed or failed. However, the current trend is to maintain and manage pipe systems before failure time. To achieve this goal, municipalities and utilities have begun to implement asset management systems. Infrastructure asset management is a comprehensive and cost-effective tool to maintain pipeline system at desired conditions. An effective asset management plan can develop various strategies to help utility owners and municipalities to understand the timing and associated costs of maintenance, rehabilitation or replacement of deteriorated pipes.

Deterioration of sewer pipes is very complex process and several factors affect the condition of pipes simultaneously. Sewer pipes are covered and buried in urban areas and it is very difficult to identify the pipes with high potential of failure. It is obvious that monitoring and inspection of all sewer pipes is almost impossible due to limited budget, time and assessment technologies. Therefore, more attention is needed to develop pipe deterioration models that can predict the current and future condition of sewer pipelines. This dissertation discusses the different statistical and artificial intelligence models used to predict condition states of sanitary sewer pipes. Furthermore, the influence factors that affect the condition states of sanitary sewer pipes will be reviewed.

1.2 Research Needs

A wide variety of pipe deterioration and condition prediction models were developed to forecast long term behavior of sewer pipes, but according to following literatures, there is still a high demand to implement more advanced models with higher accuracy and details. One of the most important limitation of current sewer prediction models has been unavailability of enough data to train and validate reliable models. Several authors suggested that condition prediction models for sewer pipes are required to be improved from different perspectives as described below:

- Kulandaivel (2004) suggested improving neural network model for deterioration of sewer pipes by considering more historical input variables, such as surface load, groundwater, bedding conditions, soil corrosion and stability and sewer location.
- Tran (2007) recommended that different case studies should be used to develop sewer deterioration models to verify findings of previous studies. Also, more investigation can improve results of previous neural network models by considering extra input variables.
- Chughtai (2008) suggested using more predictors, such as soil conditions, and seismic factors for developing condition deterioration models for sewer pipes. Also, application of other prediction models should be investigated in future studies.
- Park (2009) indicated that not much works regarding the deterioration mechanism for the sewer pipes have been conducted and more research is needed to identify the parameters that affect the deterioration of sewer pipes.

- Syachrani (2010) once again suggested more comprehensive models can be developed by incorporating additional location related attributes such as soil type, water table, etc.
- Salman (2010) recommended improving deterioration models by consideration of more variables, such as soil type, groundwater level, and initial quality of construction.
- Mashford et al. (2011) recommended a detail comparison of support vector machine and artificial neural network models.
- Opila (2011) indicated that additional development of the condition prediction models would result in more accurate failure predictions. Other prediction models may provide more accurate result.
- Sousa et al. (2014) suggested to employ more advanced deterioration models and compare the results with machine learning and neural network models.
- Atique (2016) stated that more studies can be done on different variables of soil data such as dry/wet condition of soil, chloride level, sulfate level of soil, and their effects on pipe deterioration.
- Bakry et al. (2016) recommended gathering more data to investigate more influential factors that affect the deterioration of sewer pipes.
- Kabir et al. (2018) suggested that the developed sewer structural condition prediction models can be further improved by analyzing the effects of other independent variables such as sewer function, groundwater level, soil type, road class, and initial quality of construction.
- Laakso et al. (2018) indicated that future research is needed to show how pipe condition depends on predictor variables.

1.3 Research Objectives

The primary objective of this study is to develop condition prediction models that can forecast condition of sanitary sewer pipes based on historical inspection database. The condition score of individual sanitary sewer pipes and the probability of pipe being in each condition level can be estimated through development of prediction models.

The secondary objective of this dissertation is to identify significant factors that affect the deterioration of sewer pipes. As presented before, deterioration of pipelines is very complex process and several factors affect the deterioration of pipes simultaneously. Therefore, by identifying these factors the design and installation of sewer pipes can be improved by optimizing performance of sewer system. For example, if the slope of pipe is a significant factor, it is possible to consider an appropriate slope at design phase to decrease the rate of deterioration. Similarly, identifying influence factors helps agencies and municipalities to collect less data points during inspection.

The third objective of this research is to compare the performance of different modeling techniques, such as statistical and artificial intelligence models for predicting the condition levels of sanitary sewers. In general, it is not possible to claim that one model is always better than the other for condition prediction of pipes, but the performance and accuracy of different models can be investigated based on the data and methodology used to develop the prediction models.

1.4 Scope of Work

The scope of this dissertation is limited to use of condition scoring of sanitary sewer pipes obtained from closed circuit television (CCTV) inspection for modeling the deterioration of pipe systems. Condition of individual sewer pipes can be categorized based on Pipeline Assessment and Certification Program (PACP) developed by the

National Association of Sewer Service Companies (NASSCO). Table 1-1 presents the scope of this dissertation.

Table 1-1 Scope of Work

Included	Not Included
Sanitary sewer pipes	Stormwater pipes
Gravity sewer pipes	Force main sewer pipes
Inspected pipes based on PACP guidelines	Inspected pipes based on other guidelines
Polyvinyl Chloride Pipes (PVC), Vitrified Clay Pipes (VCP)	Cast Iron (CAS), Ductile Iron Pipe (DIP), Reinforced Concrete Pipes (RCP) and unknown pipes
Sanitary sewer pipes without any repair or rehabilitation history	Pipe segments that have history of lining and repairs

1.5 Research Methodology

The deterioration models developed in this study are used to predict the condition rating of individual sewer pipes by considering the physical attributes of the pipes and various environmental factors. These influential factors are possibly contributing to the deterioration of sewer pipes over time. The following steps are carried in this methodology to achieve the expected outcome of the research as shown in Figure 1-1 as well.

- Step 1: Problem definition
- Step 2: Comprehensive literature review
- Step 3: Data collection
- Step 4: Data analysis
- Step 5: Development of deterioration models
- Step 6: Model validation

- Step 7: Identifying significant factors
- Step 8: Comparing statistical and artificial intelligence models
- Step 9: Select the best model

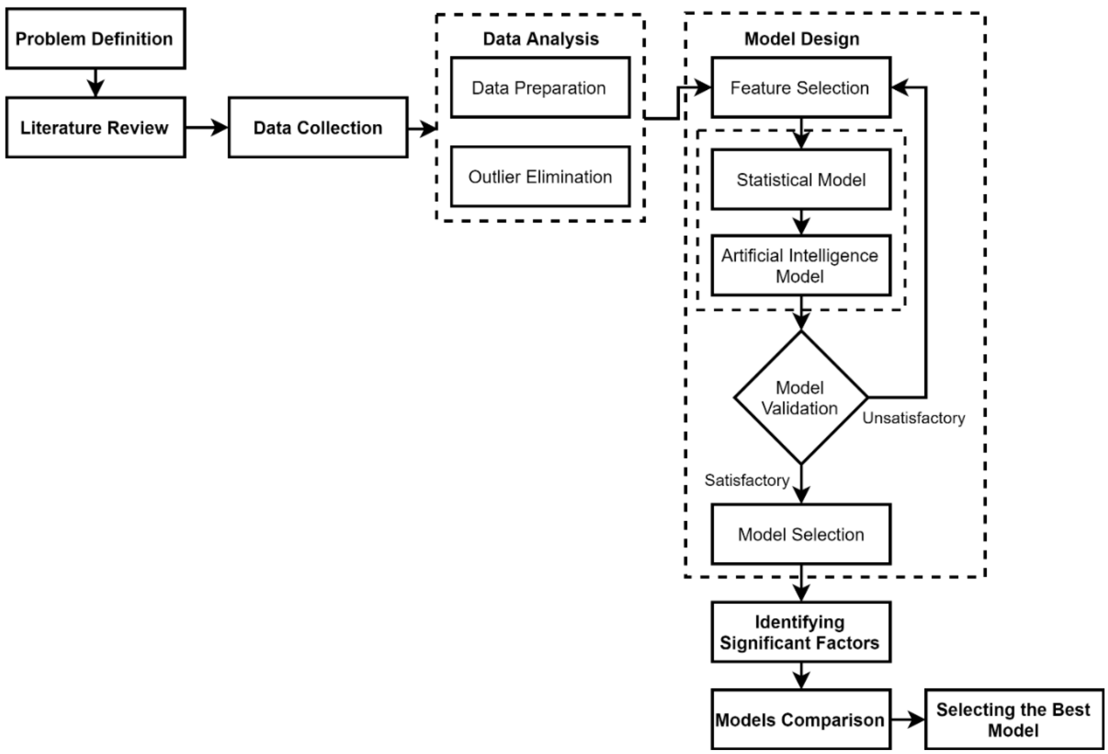


Figure 1-1 Research Methodology

1.6 Expected Outcome

The outcomes of this study are outlined below:

- Logistic regression model is used in this study to investigate the deterioration of sanitary sewer pipes statistically. The result of logistic regression reflects: 1) predicted condition rating of sewer pipes; 2) probability of pipes being in each condition scale; 3) significant factors influencing deterioration of sewer pipes; 4) sewer pipes deterioration curve.

- Artificial intelligence models, such as, gradient boosting trees and k-nearest neighbors (k-NN) are developed in this study to evaluate deterioration of sewer pipes. The outcomes of these models show the predicted condition rating and important variables that affect deterioration of sewer pipes.
- The performance comparison of statistical and artificial intelligence models presents the best approach to predict deterioration of sewer pipes based on database used in this study.

1.7 Hypothesis

With enough data, it is expected that both statistical and artificial intelligence models can be utilized to predict condition levels of sewer pipes. It is expected that that pipe age and manhole to manhole length are significant parameters affecting deterioration of sewer pipes based on dataset used in this dissertation.

1.8 Organization of Dissertation

The research results are presented in this dissertation, divided into the following chapters:

- Chapter one presents background information about condition of sanitary sewer pipes and the importance of sewer inspection and maintenance strategies. Research needs, objectives, scope of work, methodology, expected outcomes and contribution to the body of knowledge are also presented.
- Chapter two provides a comprehensive review of literatures on history and types of sewer systems, asset management, condition assessment of sewer pipes, sewer inspection methods, factors affecting condition of sewer pipes and condition prediction models.

- Chapter three discusses the structure and detail of statistical and artificial intelligence models used in this dissertation. Additionally, various data analysis and validation techniques are presented in this chapter.
- Chapter four presents detail information of case study used in this dissertation to develop condition prediction models. In this chapter, description of all dependent and independent variables is provided along with descriptive and correlation analysis.
- Chapter five discusses the procedure of developing logistic regression, gradient boosting tree and K-nearest neighbors' models. The validation results of the models and influence of significant variables are also provided in this chapter.
- Chapter six describes the detail of model validation and identifying the influence variables affecting deterioration of sanitary sewer pipes. The results of this study will discuss in this chapter.
- Chapter seven presents the summary and conclusions of the research. Limitations and recommendations for further research are also included.

1.9 Chapter Summary

This chapter discussed background information about condition of sanitary sewer pipes and the importance of sewer inspection and maintenance strategies. Research needs, objectives, scope of work, methodology, expected outcome and contribution to the body of knowledge also were presented in this chapter.

Chapter 2 Literature Review

2.1 Background and Overview

The large expected future funding gap and the aging sewer infrastructure systems lead into the need for efficient use of available funds (ASCE, 2017). Thus, asset management has payed considerable attention in recent years to provide an acceptable level of service at a minimum cost. Municipalities and utility districts use asset management program to determine the current states of assets, level of service, critical assets, minimum life cycle cost and long-term funding plan.

The primary components of any asset management program include the identification, location, and condition of assets. Pipeline condition assessment provides the critical information about physical and operational condition of pipes to estimate remaining service life and long-term performance of infrastructure pipe systems. Pipe condition assessment can be determined through standard coding systems and collected information from inspection process (EPA, 2009).

A variety of tools and techniques are available today to detect and predict the condition of sewer pipes. However, the average rate of pipeline rehabilitation and renewal is not adequate to control quality demands and frequently deteriorating systems (EPA, 2010a). Developing a comprehensive asset management program can result a systematic decision-making approach to identify critical assets and rehabilitate or replace the pipes before failure time.

Vast majority of agencies responsible for wastewater collection are public and administered under a municipal or regional government structure. The maintenance and replacement of wastewater collection system has historically been underfunded (EPA, 2010a). Condition assessment is a time consuming and expensive part of asset management program and due to limited budget, water and wastewater agencies always try to find an alternative solution to decrease the costs and time of condition assessment procedure.

As pipes are inspected on multiple occasions over time, and as inspection techniques improve, condition prediction models will be able to incorporate recorded defects and predict how pipes with certain defects will behave. Additionally, condition prediction models can predict failures for various levels of available information. If a pipe network has both inspected and uninspected pipes, prediction models will be able to predict current and future condition of all pipes in the network (Opila, 2011). Therefore, developing a comprehensive condition prediction model is very helpful for utility agencies and municipalities to assess the current and future condition of sewer pipes.

2.2 Sewer System in the United States

2.2.1 History

In the seventeenth century, there was no conveyance system to collect the raw sewage. Because of low density population, the lack of sewage system did not create sanitation problem at that time. Sewage system were more common in Europe, and Asia since they had more experience to construct it. During the 1800s, demand for more effective sanitary system increased with growing the population in the United States (Burian et al., 2000). After rapid urbanization between 1840 and 1880, municipalities began to build sewer systems to protect public health and preventing the flood (Melosi, 2000).

Sewers constructed before the 1850s were not planned, designed, or constructed by skilled engineers and consequently, the goal of solving sanitation problem was not achieved by overall public or private sewers constructed in the early nineteenth century. The unplanned and uncontrolled drainage of wastewater from privy vaults and cesspools polluted soils and groundwater, and that occasionally led to contaminated drinking water and disease epidemics (Burian et al., 2000). Therefore, the American municipalities decided to find an alternative solution for sewage system.

In the late 1850s, the first comprehensive sewer system in the United States were constructed in Chicago and Brooklyn (Burian et al., 2000). Sanitary condition and disease epidemics extremely improved by construction of sewer system. Extensive construction of urban sewer systems did not start until the 1880s.

In the United States, municipalities installed two types of sanitary sewer systems which are Combined Sewer Systems (CSS) and Separate Sanitary Sewer and Storm Sewer System (SSS) (EPA, 2004).

2.2.2 Combined Sewer System (CSS)

In combined sewer system, a single pipe is used to transport domestic, commercial and industrial wastewater, and storm water to a selected disposal location. The first comprehensively plan and designed combined sewer system was constructed in Hamburg, Germany. The development of the combined sanitary sewer in the United States with considering planned network and large diameter sewers began during the late nineteenth century (Burian et al., 2000).

Municipalities that needed both sanitary and storm sewers tended to construct combined sewer system since it was less expensive for two conveying system. In general, combined sanitary sewers are constructed more in large cities to control unexpected flood (EPA, 2004). Figure 2-1 illustrates a typical combined sewer system flows during wet and dry weather conditions.

.

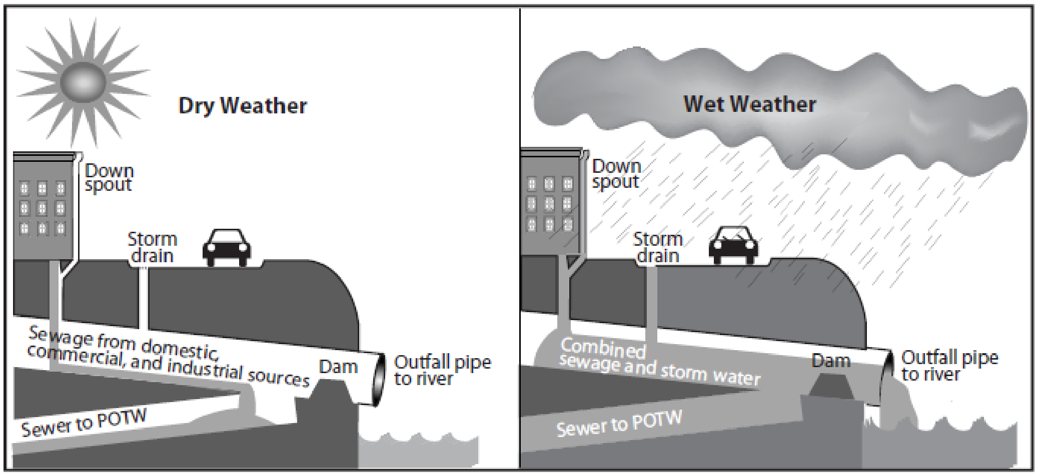


Figure 2-1 Combined Sewer System

(EPA, 2004)

2.2.3 Separate Sanitary Sewer and Storm Sewer System (SSS)

The concept of separate sanitary sewer and storm sewer system is to manage storm water and sanitary wastewater separately. In this method, two separate pipes are used to convey domestic, commercial, and industrial wastewater, and storm water to a selected disposal location. Separate sanitary sewer was less expensive for municipalities that desired only a wastewater collection system. Unlike combined sewer systems, the separate system was not constructed to collect the large amount of water from wet weather events (EPA, 2004).

In the late nineteenth century, most of sewer system constructed in the United States were combined because: 1) there was no evidence to prove the success of separate sewer system in Europe; 2) municipalities believed that the combined sewer system is less expensive; and 3) the agriculture use of separate sewer system was unknown for engineers (Burian et al., 2000). Figure 2-2 illustrates a typical separate sanitary sewer system during wet and dry weather conditions.

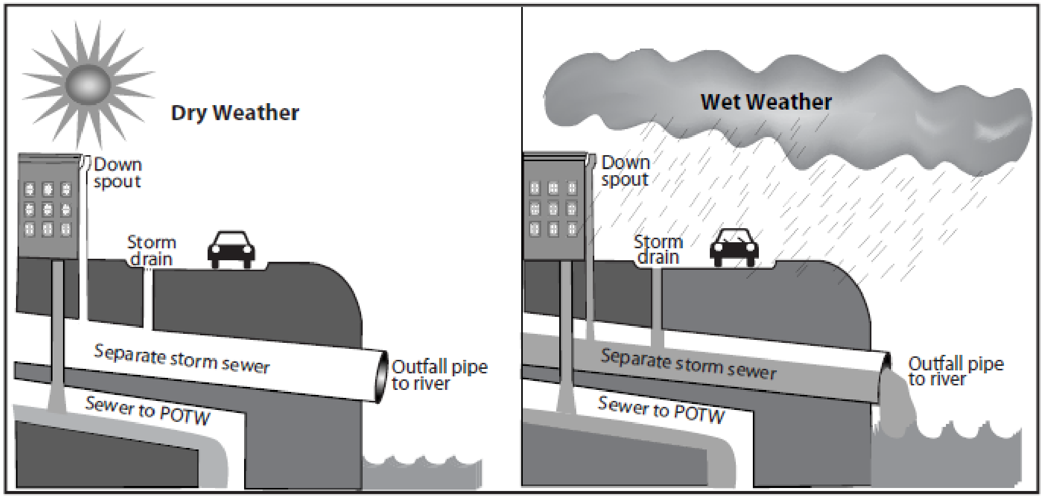


Figure 2-2 Separate Sanitary Sewer System

(EPA, 2004)

2.2.4 Types of Sewer Pipes

Several different pipe materials are available to construct sanitary sewer systems, and each pipe with a unique characteristic is used in different conditions. Until 1850, bricks were used to construct sewer systems and at the middle of nineteenth century vitrified clay pipes were used more. Concrete pipes were used at the beginning of the twentieth century and other pipes, such as, polyvinyl chloride, fiberglass, high-density polyethylene, ductile iron, steel and reinforced concrete were used gradually after that (Kulandaivel, 2004).

Gravity lines, force mains, and service laterals are the most common types of pipe using in wastewater systems. A gravity line is a sewer pipe that is operating based on initial designed slope. In the force main sewers, a pump generates the pressure and convey the swage through the pipe. And, service laterals are the pipes that transfer wastewater from buildings to the sanitary lanes (EPA, 2010b). According to EPA (2010b) sanitary and wastewater sewer systems are generally constructed by ferrous pipes, concrete pipes, ceramic-based pipes and plastic pipes as presented in Table 2-1.

Table 2-1 Sanitary Sewer Pipe Material

(EPA, 2010b)

Pipe Types	Pipe Material
Ferrous Pipe	ductile iron, cast iron, and steel
Concrete Pipe	reinforced concrete pipe (RCP) and prestressed concrete cylinder pipe (PCCP)
Ceramic-based Pipe	brick and vitrified clay pipe (VCP)
Plastic Pipe	polyvinyl chloride (PVC) and high-density polyethylene (HDPE)

2.3 Asset Management

Asset management in the water and wastewater industry is an adapted concept from many success implementations in other industries such as transportation and building infrastructure management. In the early 1990s in Australia and New Zealand asset management was introduced before developing in other countries involving Canada, England and the United States. In the United States, Federal Highway Admiration introduced the infrastructure asset management in the early 1990s and Asset Management Primer was published by FHWA in 1999. This was the first published asset management in the U.S. that entirely covers asset management procedures and after that many other agencies were convinced to implement it. In the early 2000s, asset management started to be developed in the water and wastewater industry and the Environment Protection Agency (EPA) played an important role to provide and support asset management practices (Syachrani, 2010). While, the asset management program is relatively new concept for the water and wastewater industry, it is rapidly developing to reach wide acceptance in the U.S. and elsewhere in the world (Schulting and Alegre, 2007).

According to EPA, asset management is a continuous procedure that leads the acquisition, use, and disposal of infrastructure assets to optimize service delivery and minimize costs over the asset’s entire life. For wastewater management utilities, asset management can be defined as an inclusive plan to manage infrastructure capital assets to minimize the total cost of owning and operating them, while delivering a satisfactory level of service (EPA, 2002). Based on EPA report (2002), among public utility agencies in the U.S., the concept of infrastructure asset management is most widely used in the transportation area to protect and maximize investments in highway, rail, and airport infrastructure assets. Implementation of an infrastructure asset management are varying from one agency to the other depending on their available fund, needs and abilities (Vanier, 2001; EPA, 2002; IIMM, 2006).

An infrastructure management system mainly involves seven different components. The actual structure may vary in different methods, but the basic concepts are similar (Park, 2009). The main components of Infrastructure management system framework are presented in Figure 2-3.

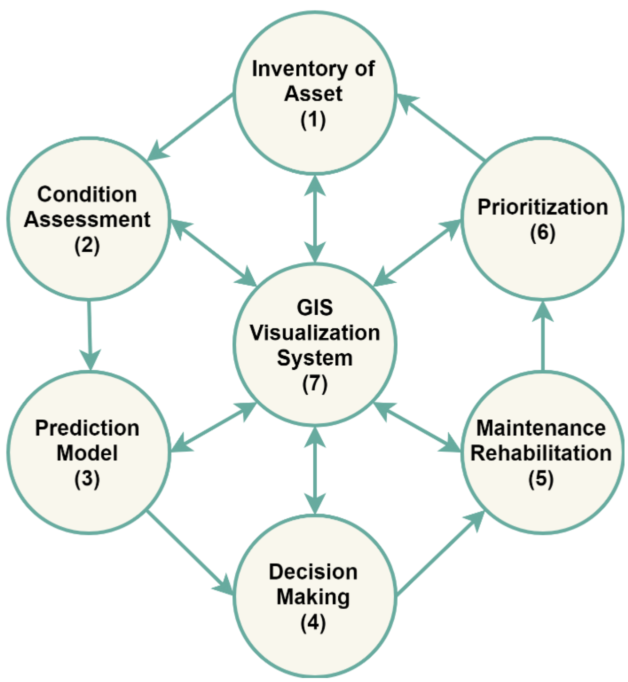


Figure 2-3 Infrastructure Management System Framework

(Park, 2009)

As Figure 2-3 illustrates, step one of infrastructural asset management process begins with the development of an asset inventory which contains information of all inspected assets in the network. The International Infrastructure Management Manual (IIMM, 2006) recommended this step as a first approach to build an infrastructure asset management system. As part of inventory process, data collection can play and important role. An inclusive record of asset, such as age, location, material, depth, length and other important information must be collected by utilities or municipalities in water and wastewater industry.

The second step is evaluating the physical, operational and economical condition of the asset. After collecting the data, the information must be analyzed and then rank based on condition coding systems. Usually a scale of 0-100 or 1-9 are applied to evaluate the condition of bridges and pavement, while, a 1-5 grading system is used for sewer system (Park, 2009). In wastewater industry, most of agencies and utilities use a scale of 1 to 3 (WSAA, 2002) or 1 to 5 to assess the condition of sanitary sewer or storm water pipes. The condition assessment procedures and methods are presented more in further sections. The next step is building prediction models to forecast the future condition of asset. In this step historical data are used to predict the future performance of the asset, due to preventing any unexpected collapse or failure. Infrastructure systems are critical for daily activities and forecasting their future condition and their remaining useful life is essential for utilities and municipalities.

Decision making process in step 4 is an infrastructure management system which provides long term plans to maintain the asset and optimize resource allocation. The results of condition assessment and prediction models lead the agencies to organize a decision-making plan for current and future condition of the asset. In decision-making process several aspects, such as available fund, regulations, method of rehabilitation or replacement and other important factors must be reflected. The step 5 in an infrastructure management system is maintenance and rehabilitation of the asset, based on result of decision-making process. For example, in water and wastewater industry, agencies can make decisions to rehabilitate or replace the damaged pipe by using trenchless technology or conventional open-cut methods. And, finally all the previous procedures assist the government to prioritize the assets for future investments as shown in step 6.

In the current practice of asset management, all the infrastructure management steps are merged together with a Geographic Information Systems (GIS) shown in step 7. By using the GIS system, huge amount of data in different layers can be managed and update easily. Application of the monitoring data in condition assessment of the new projects are highly important. Monitoring data can be utilized to verify the effectiveness of the new methods and projects (Sterpi et al. 2017, 2018).

2.4 Condition Assessment of Sewers

2.4.1 Introduction

In the United States, millions of gallons of human and industrial waste are conveyed into wastewater treatment plan through underground sewer systems every day. This process takes place underground (out of sight) as maintaining wastewater collection systems is always one of the critical challenges of governments. As the most municipal sewer systems are at least 60 years old, many communities and utilities are paying more attention to assess the condition of their underground pipes and associated infrastructures (EPA, 2015).

Condition assessment is one of the essential components of infrastructure asset management program. According to EPA (2007), condition assessment is analyzing the data and information collected from direct inspection, in-direct monitoring and reporting to determine the structural, operational and performance status of infrastructure assets. In the other worlds, the term “condition assessment” relates to evaluating the existing physical condition, identifying the deterioration pattern, and determining the potential of collapse or failure of an asset (NRC, 2004).

The main concept of sewer condition assessment is to compare the current structural and operational condition of a sewer pipeline to a new or like new pipe. The result of comparison is a numerical grade for asset which present the existing condition of underground sewer pipelines. The existing sewer pipelines are considered in sewer condition assessment program to set a milestone for giving maintenance priorities to different pipelines depending on the risks associated with their breakdown (Khazraeializadeh, 2012). Figure 2-4 presents the condition assessment algorithm suggested by McDonald and Zhao (2001).

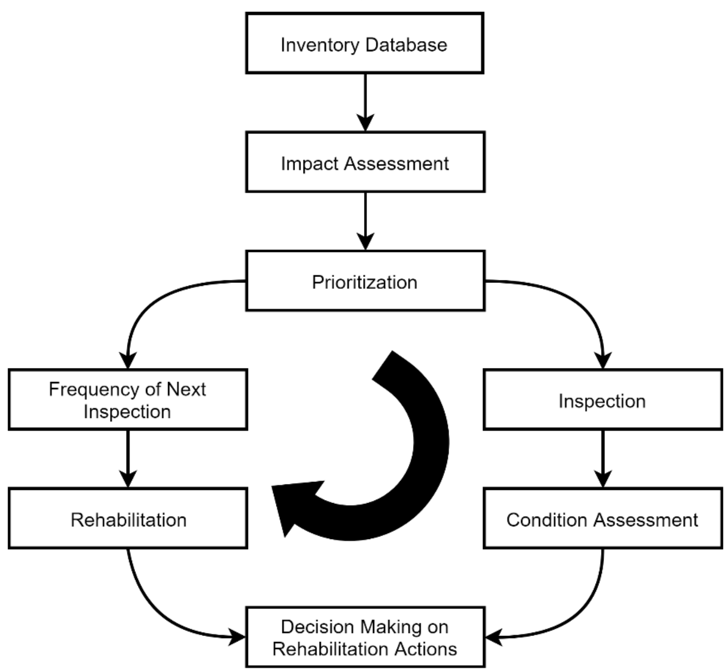


Figure 2-4 Condition Assessment Algorithm

(McDonald and Zhao, 2001)

2.4.2 Condition Rating Methods of Sewer Pipes

Sewer defect coding is essential for the worldwide sewer rehabilitation industry to discover the critical information about the underground infrastructure (Thornhill and Wildbore, 2005). The historical background of the sewer condition assessment protocols goes back to1977 when the Water Research Centre (WRc) in United Kingdom started a five year research project to implement a methodology to assess the condition of sewer pipes based on a general coding system. The world first Sewerage Rehabilitation Manual (SRM) was published in 1980 by WRc and this standard later became the main reference to develop more sewer condition assessment protocols (Chughtai and Zayed, 2001; Rahman and Vanier, 2004). The condition rating is used to objectively evaluate the current condition of sewer pipes. Structural condition and operational condition are two common pipe condition categories (Chughtai and Zayed, 2008). Structural condition evaluates the pipe defects, the physical strength of a pipe and the capability of the pipe to resist external loads, and operational condition indicates the ability of the pipe to meet its service requirements. The result of structural conditions can be used to determine the necessity of pipe rehabilitation or replacement while the operational condition of a pipe indicates the need for cleaning and maintenance (Opila, 2011).

Numerous methodologies have been developed to score the condition of buried sewer pipes in different countries, such as, WRc in United Kingdom, PACP in the U.S., NRC in Canada, and WSAA in Australia (Moteleb, 2010). In general, condition prediction scales are classified by discrete or finite scale values of relatively limited ranges (Baur and Herz, 2002). For example, a 1 to 5 classification scale is used to assess the condition of sanitary sewer pipes in PACP and WSAA methods with 1 as an acceptable condition, and 5 as a poor condition. The details of most common condition rating methods are presented in following sections.

2.4.2.1 PACP Condition Grading Method

Pipeline Assessment and Certification Program (PACP) is the North American Standard for pipeline defect identification and assessment to identify the pipe condition and manage the sewer pipe networks. In 2001, National Association of Sewer Service Companies (NASSCO) developed the PACP in partnership with Water Research Center (WRc) to assess the condition of sewer pipes. The goal of PACP is to create a comprehensive database to correctly identify, plan, prioritize, manage and renovate the sewer pipe assets based on condition evaluation.

Pipe defects and features can be classified into five categories by NASSCO coding system. The defect classification involves; (1) continuous defects, (2) structural defects, (3) operational and maintenance, (4) construction features, and (5) miscellaneous features coding (EPA, 2015). For each type of defect, the numeric codes are used to rank the severity of the pipe defect and capital letters define the type of defect as shown in Table 2-2. For example, “FC” represents a circumferential fracture, “SCP” shows the surface chemical attack and “X” presents the pipe collapse.

Grades are assigned based on several factors, such as, significance of the defect, extent of damage, and percent of restriction to flow capacity or the amount of wall loss due to deterioration. The final condition rating is defined from two major categories which are structural and operation and maintenance (O&M). The below list presents the grades and definitions of grades respectively (NASSCO, 2015):

5 - Most significant defect grades

4 - Significant defect grade

3 - Moderate defect grade

2 - Minor to moderate defect grade

1 - Minor defect grade

Table 2-2 PACP Structural and Operational Defects Codes Sample

(NASSCO, 2015)

Family	Group	Descriptor	Code	Structural Grade	O&M Grade
Structural	Crack (C)	Circumferential (C)	CC	1	
		Longitudinal (L)	CL	2	
		Multiple (M)	CM	3	
		Spiral (S)	CS	2	
	Fracture (F)	Circumferential (C)	FC	2	
		Longitudinal (L)	FL	3	
		Multiple (M)	FM	4	
		Spiral (S)	FS	3	
	Collapse (X)	Pipe (P)	XP	5	
		Brick (B)	XB	5	
	Weld Failure (WF)	Circumferential (C)	WFC	2	
		Longitudinal (L)	WFL	2	
		Multiple (M)	WFM	3	
		Spiral (S)	WFS	2	
O&M	Infiltration (I)	Weeper (W)	IW		2
		Dripper (D)	ID		3
		Runner (R)	IR		4
		Gusher (G)	IG		5
	Deposits Attached (DA)	Encrustation (E)	DAE		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Grease (G)	DAGS		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Ragging (R)	DAR		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Other (Z)	DAZ		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
	Obstacles/Obstructions (OB)	Brick or Masonry (B)	OBB		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Pipe Material in Invert (M)	OBM		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Object Protruding Thru Wall (I)	OBI		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Rocks (R)	OBR		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Built into Structure (S)	OBS		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		Construction Debris (N)	OBN		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5
		External Pipe or Cable in Sewer (P)	OBP		<=10% - 2, <=20% - 3, <=30% - 4, >30% - 5

PACP assess the condition of pipes on a scale of 1 to 5 based on the result obtained from CCTV inspections and operator judgments. Condition 1 determines the pipe is in excellent condition and condition 5 specifies the pipe has failed or is likely to fail. Pipe with condition rating of 5 needs immediate action for rehabilitation or replacement. Table 2-3 provides the PACP condition rating, from the PACP manual.

Table 2-3 PACP Defect Grades

(NASSCO, 2015)

Condition Grade	Description	Time to Failure
5 Immediate Attention	Defects requiring immediate attention	Pipe has failed or is likely to fail within the next five years
4 Poor	Severe defects that will become Grade 5 defects within the foreseeable future	Pipe will probably fail in 5- 10 years
3 Fair	Moderate defects that will continue to deteriorate	Pipe may fail in 10-20 years
2 Good	Defects that have not begun to deteriorate	Pipe unlikely to fail for at least 20 years
1 Excellent	Minor defects	Failure unlikely in the foreseeable future

The outcome of PACP condition grading system is completely dependent on the quality of the defect coding and any error during detection of defects affects the result of final grades. The PACP condition grading system ranks the pipe segments based on severity of the observed defect and conditions. Quick rating, segment grade score, overall pipe rating and pipe rating index are three different ways to express the condition of sewer pipe segments.

The PACP quick rating is a four-character score which explains the number of occurrences for the two highest severity grades. The first and third digits of four-character index express the highest severity grades occurring along the pipe. The second and forth digits indicate the total number of these defects and alphabetic characters are used if the total number exceeds 9. For example, a quick rating of 5219 indicates 2 defects with grade of 5 and 9 defects with grade of 1 were observed along the pipe. In other example, a pipe with nineteen grade 4 and twenty grade 2 would receive the quick rating of 4B2C. Figure 2-5, briefly shows the detail of quick rating index.

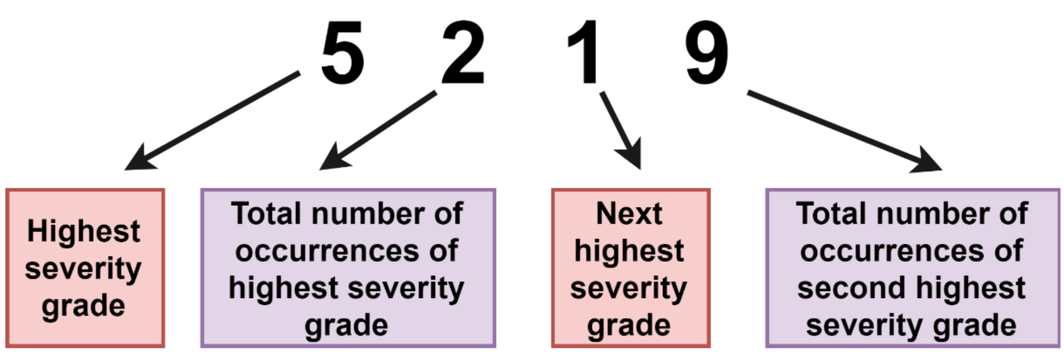


Figure 2-5 PACP Quick Rating Index

(NASSCO, 2015)

Segment grade scores (SG) are calculated by multiplying each condition grade by its number of occurrences. Therefore, each pipe segment has individual grade score for each of the five condition grades. For example, if a pipe has 5 structural defects of grade 5, 2 defects of grade 3 and 6 defect of grade 2, the segment grade scores are respectively SG5 = 25, SG3 = 6 and SG2 = 12. Segment grades scores are calculated for both structural and operational defects.

Overall pipe rating (OR) is obtained from summation of the five individual segment grade scores. For instance, the overall pipe rating is 43 for structural defects in previous example. Structural and O&M defect grades are used separately to calculate the overall pipe index for each pipe segment. Table 2-4 provides an example of overall pipe rating and segment grade score calculation.

Table 2-4 Overall Pipe Rating and Segment Grade Score

Condition Grade	Defects		Segment Grade	
	Structural	O&M	Structural	O&M
5	5	0	25	0
4	0	0	0	0
3	2	2	6	6
2	6	4	12	8
1	0	0	0	0
Total Defects=	13	6		
		Overall Rating=	43	14

And finally, the pipe rating index (RI) provides the overall defect severity along the pipe segment by dividing the overall pipe rating by the total number of defects. The pipe rating index are calculating separately for structural and O&M conditions. For example, in previous case the $RI_{structural}$ is 3.3 and $RI_{O\&M}$ is 2.3. Condition 1 determines the pipe is in excellent condition and condition 5 specifies the pipe has failed or is likely to fail and immediate action is needed to rehab or replace it. As explained before, In the United States most of municipalities and agencies use the PACP methodology to assess the condition of sewer pipes.

2.4.2.2 WRc Condition Grading Method

The historical background of the development of sewer condition assessment protocols goes back to1977 when the Water Research Centre (WRc) in United Kingdom started a five year research project to implement a methodology to assess the condition of sewer pipes based on a general coding system. The world first Sewerage Rating Manual (SRM) was published by WRc in 1983 to assess the condition of individual pipes using the

Closed-Circuit Television (CCTV). Since 1983, the WRc standard has been revised five time and the fifth edition was published on 2013 (WRc, 2004).

According to WRc, two major structural and operational categories are used to determine the defects and evaluate pipe condition. In this protocol deduct structural or operational values are assigned for various defect categories ranging from 1 to 165 and an overall sewer condition grade is identified for the whole pipe segment using scale 1 to 5 (WRc, 2004). Structural and operational deduct values are assigned from the defect codification. The impact of the defect on the service life and performance of the sewer pipe are determined by defect weights. Structural defects define the physical condition of a sewer pipe. The structural defect scores depend on severity of defects and type of pipe material (Chughtai, 2008). Table 2-5 show some common defect scores in concrete pipes.

Table 2-5 Overall Pipe Rating and Segment Grade Score

(Adapted from Chughtai 2008)

Defect	Detail	Score	Unit
Joint Opening	Slight	0.1	Per Joint
	Medium	0.5	Per Joint
	Large	2	Per Joint
Crack	Circumferential	1	Per Crack
	Longitudinal	2	Per Crack
	Multiple	5	Each
Fracture	Circumferential	8	Per Crack
	Longitudinal	15	Per Crack
	Multiple	40	Each
Deformation	5%	10	Each
	10%	30	Each
	15%	60	Each
	20%	90	Each
	25%	125	Each
	30% or more	165	Each
Hole	<1/4 Circumferential	80	Each
	>1/4 Circumferential	165	Each
Broken Pipe		80	Each
Collapsed Pipe		165	Each

The peak score shows the greatest worst defects in each pipeline and it is the maximum defect score for any one meter length of pipe. The defect scores are calculated and based on peak defect score or deduct value, a single condition grade for the structural or operational condition of pipe is considered. Table 2-6 provides severity condition grades for WRc protocol for both structural and operational condition (Chughtai, 2011).

Table 2-6 Severity Condition Grades for WRc Protocol

(Chughtai 2011)

Condition Grade	Description	Peak structural defect score	Peak operational defect score
1	Acceptable condition	< 10	< 1
2	Minimal collapse risk but potential for further deterioration	10–39	1–1.9
3	Collapse unlikely but further deterioration likely	40–79	2–4.9
4	Collapse likely in near future	80–164	5–9.9
5	Collapse imminent or collapsed	165 and higher	> 10

2.5 Sewer Inspection Methods

2.5.1 Introduction

Collecting pipe data and data analysis are two required processes to perform a condition rating score for pipe infrastructure. Inspection is the first step of condition assessment plan to collect pipe characteristic data, such as, physical attributes (pipe diameter, material, depth, length, age, etc.), environmental attributes (soil type, corrosivity, groundwater level, pipe bedding, temperature, etc.), and operational attributes (internal pressure, velocity, operational and maintenance procedures, etc.) (Opila, 2011). Different inspection and monitoring methods can be used to collect pipe information.

The primary purpose of an inspection is to evaluate the current condition of an asset, and to detect structural and operational (hydraulic) problems along the pipe segment. A detailed work plan is needed to outline the assets that should be inspected, time of inspection and the required technologies to perform inspection. Ideally, an inspection would occur at the end of pipe service life, but the condition of buried pipeline is unknown before performing any condition assessment (EPA, 2009). In general, cracks, fractures, abrasions, joint problems, and corrosion are the most important factors cause the structural problems. Operational problems occur due to several factors, such as, obstacles, infiltration, inflow, sedimentation and root intrusion (Salman, 2010). Inspection and monitoring of asset performance plays a significant role in pipe condition assessment.

It is very costly to inspect every linear foot of a sewer system, especially when a little prior inspection history is available for sewer network, so a comprehensive condition assessment plan is needed to focus on critical pipes to establish inspection methodology. Inspection timing and frequency is another important factor to minimize the inspection cost and the likelihood of sewer failure. The selected inspection technique depends on type of asset and the methodology to scale sewer pipe condition. CCTV is the most common method to inspect sewer pipes for structural and operational defects, however, a variety of technologies are available to inspect wastewater pipeline networks.

2.5.2 Inspection Technologies

The integration of inspection and condition grading systems helps forecast the future condition or remaining useful life of sewer pipes by developing pipe deterioration models. According to EPA (2009), inspection technologies for wastewater systems can be classified in following categories:

- Camera
- Acoustic
- Electrical/electromagnetic
- Laser
- Innovative technologies

Some of the most common methodologies are briefly described in following sections. Table 2-7 provides a summary of typical applications for each technology.

Table 2-7 Inspection Technology Overview

(EPA 2009)

Technology		Sewer type			Pipe material	Pipe diameter	Defects detected			
		Gravity	Force main	Lateral			Internal condition	Pipe wall	Leakage	Pipe support
Camera	Digital cameras	•			Any	6-in. to 60-in.	•	•	•	
	Zoom cameras	•			Any	> 6-in.	•	•	•	
	Push-camera			•	Any	1-in. to 12-in.	•	•	•	
Acoustic	In-line leak detectors	•	•		Any	≥ 4-in.			•	
	Acoustic monitoring systems		•		PCCP	≥ 18-in.		•		
	Sonar/ultrasonic	•	•		Any	≥ 2-in.	•	•		
Electrical/ electromagnetic	Electrical leak location	•	•	•	Nonferrous	≥ 3-in.			•	
	Remote field eddy current	•	•	•	Ferrous, PCCP	≥ 2-in.		•	•	
	Magnetic flux leakage	•	•	•	Ferrous	2-in. to 56-in.		•		
Laser	Laser profiling	•	•		Any	4-in. to 160-in.	•	•		
Innovative technologies	Gamma-gamma logging	•	•	•	Concrete	Not yet defined				•
	Ground penetrating radar	•	•	•	Any	Not yet defined			•	•
	Infrared thermograph	•	•	•	Any	Not yet defined			•	•
	Micro-deflection	•			Brick	Not yet defined		•		•
	Impact echo/SASW	•			Brick/ Concrete	> 6-ft		•		

2.5.2.1 Camera Inspection

Closed Circuit Television (CCTV) is the most common inspection method in camera inspection category to evaluate condition of pipes by generating video records. CCTV inspection have been used by agencies and municipalities since 1950s (ISTT, 1990). CCTV inspection is one of the most widely used inspection methods allow agencies to obtain and store more accurate and detailed information. In general, video cameras are used to perform a visual recording of inside condition of a pipeline. The visual inspection of sanitary sewer lines enables a CCTV operator to recognize specific defect along the pipeline and make it possible to inspect too small or hazardous pipelines. Different equipment, such, as pushrod cameras or remote-control robot crawlers are used to convey the camera through the pipeline. The primary disadvantages of CCTV inspection method are the limitation of detecting the pipe surface above the water line, the restriction to provide any structural data on the pipe wall integrity and surrounding soil around the pipe (EPA, 2009).

CCTV inspections can identify numerous types of defects involving cracks, infiltration, inflow, tree roots, collapse, obstacles, protruding laterals, offset joints and presence of grease (WEF/ASCE, 2009). According to EPA (2009), data obtained from CCTV inspection includes:

- Evidence of sediment, debris, roots, etc.
- Evidence of pipe sags and deflections
- Off-set joints
- Pipe cracks
- Leaks
- Location and condition of service connections

The quality of defect identification and accuracy of CCTV highly dependent on operator experience, picture quality and flow level (Salman, 2010; EPA, 2009; Allouche and Freure, 2002; Chae and Abraham, 2001). CCTV operator must have formal training and certification to be able to detect different problems and defects along the pipe. CCTV inspection is a cost-effective method providing required information to assess the pipe condition. Many existing inspection technologies just provide data on the structural condition of pipe or soil surrounding the pipe, while CCTV provides visual data on leak, location of service lateral, cracks, off-set joints and other important defects. It is the reason that CCTV is an important inspection tool for agencies or municipalities to evaluate condition of wastewater pipe systems (EPA, 2009). In general, cost of CCTV inspection increases with sewer depth, because of additional set-up time and cable length from the surface to the sewer. Figure 2-6 shows a CCTV camera inside a gravity pipe.

Figure 2-6 CCTV Inspection Technology

(EPA, 2010b)

2.5.2.2 Acoustic Technologies

Acoustic technology is an inspection method by producing vibrations or sound waves to determine pipe condition. In general, acoustic technologies are used to inspect water mains, therefor, these type of technologies can be used to assess the condition of force mains in wastewater industry. According to EPA (2009), acoustic technologies can be classified into three categories:

- Leak detectors
- Acoustic monitoring systems
- Sonar, or ultrasonic systems

Leak detectors are used to assess the condition of water or sewer pipelines by detecting sounds or vibrations produced by leaks. Variety of tools including hand-held listening device, underwater microphones, geophones, leak noise correlators, and in-line devices are used in this method to assess the condition of the pipes. Acoustic monitoring systems are used to evaluate the condition of PCCP pipes which are generally subjected to failure due to internal or external corrosion. Identifying the acoustic signals produced by broken wire inside the pipes is the methodology of defect detection in acoustic monitoring system. Acoustic Emission Testing (AET) and Sound-Print® are two common technologies to provide continues acoustic monitoring of PCCP (EPA, 2009).

Sonar or ultrasonic system was established in 1906 and for the first time, WRc used this technology to inspect pipelines in 1987 (EPA, 2009). In this method, very high frequency ultrasonic sound waves are sent through the surface of the pipeline and then reflected waves are analyzed to identify the defects along the pipe. Due to performing a comprehensive pipe inspection, ultrasonic system equipment can be combined with CCTV or other inspection tools (Najafi 2005). In this method it is possible to identify defects located below the flow line (Salman, 2010).

One benefit of sonar system is that it can be positioned in pressurized force mains without taking them out of service (EPA, 2010b). Figure 2-7 shows the results of a combined sonar and CCTV scan. Deviations in wall thickness are indicated by horizontal bars, red and orange areas show greater corrosion and the amount of sediments at the bottom of the pipe are shown in cross section view.

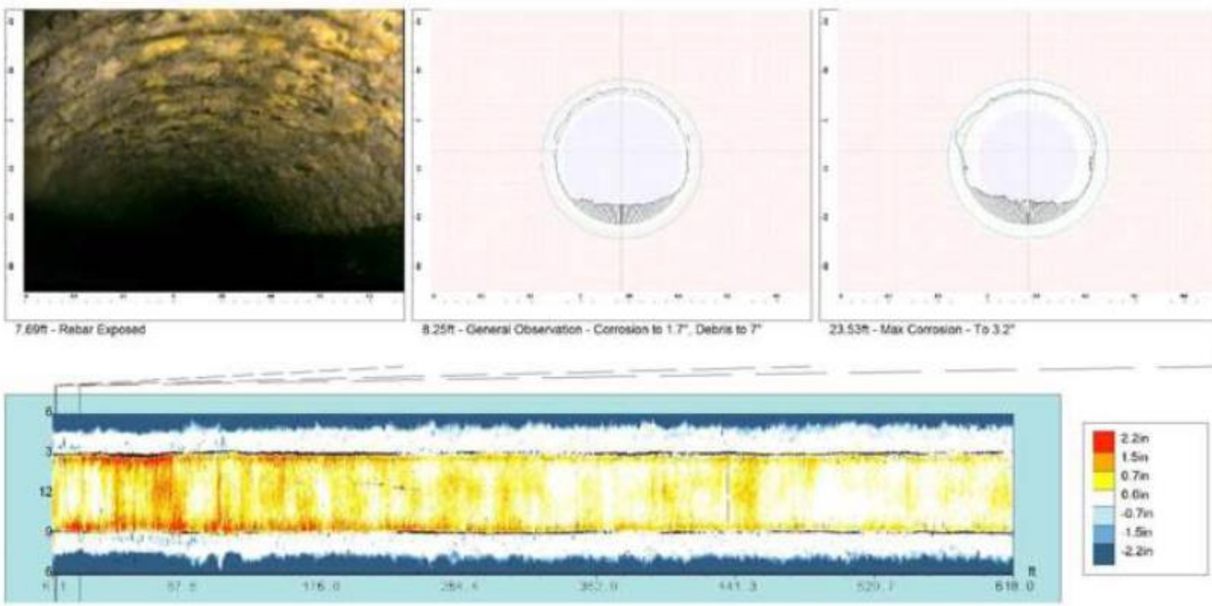


Figure 2-7 Combined Sonar and CCTV Results of a 42-in. RCP Pipe

(EPA 2010b)

According to WEF/ASCE (2009), data obtained from ultrasonic inspection involves:

- Existing condition of the pipeline
- The amount of debris
- The capacity of pipe after cleaning

Table 2-8 summarize different classification of acoustic technologies:

Table 2-8 Acoustic Technologies Summary

(EPA 2009)

Summary	Leak detectors	Acoustic monitoring systems	Sonar/ultrasonic systems
Sewer type	Force mains, gravity sewers	Force mains	Force mains, gravity sewers
Material	Any	PCCP	Any
Pipe size	≥ 4-in.	≥ 18-in.	≥ 4-in.
Defects detected	Leaks	Broken pre-stressed wires	Pipe wall deflections, corrosion, pits, voids, and cracks, debris
Original application	Leak detection in pressurized water lines	Monitoring PCCP water lines	Maritime use
Status	Commercially available for sewer inspection	Commercially available for sewer inspection	Commercially available for sewer inspection
Advantages	Can detect very small leaks	Useful as a screening technique prior to more detailed inspection	Suitable for pipes of any material and a wide range of diameters
Disadvantages	Requires minimum flow to be carried through pipe	Only detects general distress	Only inspects pipe below the waterline

2.5.2.3 Electrical and Electromagnetic Methods

Several electrical and electromagnetic methods including electrical leak location, eddy current testing (ECT), remote field eddy current (RFEC), and magnetic flux leakage (MFL) are utilized to evaluate the condition of pipes. Electrical leak location was developed in 1981 and it is one of the most widely used techniques for detecting leaks in geomembrane liners (EPA, 2009). This method can be used to assess the condition of non-ferrous force mains, gravity and lateral swears greater than 3 in. in diameter. The advantage of electrical leak location is that this method is available for service lateral, however, as a disadvantage, the gravity pipes must be filled prior to inspection.

Eddy current testing and remote field eddy current technology are used to evaluate the condition of ferrous force mains, gravity and service lateral sewer systems. In these methods variety of defects, such as, metal loss, cracks, leaks, broken wire, graphitization, and wall thickness can be detected. As a disadvantage typically post processing of data by vendors is needed before inspecting the pipe. Magnetic flux leakage detection technique is widely used to assess the condition of oil and gas pipelines. The MFL technique for pipe inspection was developed in 1965 and it can be used to inspect ferrous force mains, gravity and service lateral sewer pipes. Metal loss, circumferential and longitudinal cracks are some types of defects which can be detected by MFL technique. Magnetic flux leakage detection technique has not been widely used for assessment of sewer pipes (EPA, 2009).

2.5.2.4 Laser-Based Inspection

Laser based inspection creates a profile of the pipe wall by analyzing pipe shape and detect the defects on the pipe surface (Najafi, 2005). In this method a laser is used to generate a line of light around the pipe wall. The laser light assesses the shape of the sewer to detect any changes to the shape or pipe size, which may be caused by deflection, corrosion or siltation. This inspection method can only be performed when the pipe is out of service and in dry condition. Similar sonar system, this method can be combined with other inspection techniques, such as CCTV or sonar (EPA, 2009).

According to Salman (2010) and WEF/ASCE (2009), laser-based inspection can be used to identify following defects:

- Shape and cross-sectional area of the pipe
- Defects on the pipe wall surface
- Debris
- Capacity before/after cleaning
- Quality of the lining work

Table 2-9 provides a summary description of the laser profiling technology.

Table 2-9 Laser Profiling Technologies Summary

(EPA 2009)

Summary	
Sewer type	Force mains, gravity sewers
Material	Any
Pipe size	Product dependent
Defects detected	Deformations, siltation, corrosion
Original application	Earlier use in large diameter tunnels and caverns
Status	Commercially available for sewer inspection
Advantages	Provides better data quality then CCTV alone, can be used to create 3D models of pipelines
Disadvantages	Can only detect defects above the water line.

2.6 Sewer Pipe Deterioration Mechanisms

Like all infrastructure, the condition of pipelines deteriorates gradually over time. Combinations of corrosion, soil movements and traffic loads lead to deterioration of pipes (Jalalediny Korky et al., 2019). Deterioration of pipe has several economic and social impacts and development of annual replacement plans for critical pipes are essential for municipalities and pipe network owners. Pipe systems require continuous inspection and maintenance, and several factors increase risk of pipe deterioration. The mechanisms of sewer pipe deterioration can be classified into structural, hydraulic and operational failure which are briefly explained below (Najafi and Gokhale, 2005; EPA, 2009; Opila, 2011).

- Structural failure: this type of failure is caused by any kind of defects on pipe wall that reduce the structural integrity of pipe segment. Similarly, the soil surrounding the pipe has essential role to failure time of pipelines. In general, cracks, internal and external corrosion, pipe deflection, misaligned joints, and brakes are the most common type of defects associated with structural failure. The soil supports the pipe and transmit subjected live load and dead load to the bedding which acts as a foundation. Loss of bedding material can lead to pipe deflection, deformation, and defects on the pipe wall. Therefore, loss of soil in both bedding and cover portion, and increasing traffic load can be other causes of structural failure of pipes. According to EPA (2009), different pipe material has different degree of failure and table 2-10 shows typical failure modes in sewer systems.

Table 2-10 Failure Modes for Various Types of Pipe Material

(EPA ,2009)

Pipe Material	Failure Modes
Ferrous Pipe (Ductile Iron, Cast Iron, Steel)	• Internal or external corrosion are the primary failure mode for metal pipes
Concrete Pipe (RCP, PCCP)	• Corrosion is often a main factor in the structural failure of concrete pipes when the concrete break up at the result of corroded reinforcing steel inside the pipe
Ceramic-based pipe (Brick, Vitrified Clay Pipe)	• Collapse caused by weakened mortar is one of the main reasons of brick pipes failure • Loss of surrounding soil into the pipe is the other important mod of failure for ceramic based pipes
Plastic Pipe (Polyvinyl Chloride (PVC), High-density Polyethylene (HDPE))	• Environmental stress cracking is the primary mode of plastic pipe failure • Leaking joints can also be

- Operational failure: it is the most common failure in wastewater collection systems and generally occurs by a physical cause. The operational failure can be resolved during a maintenance procedure and normally does not affect the structural integrity of the pipe. Several type of defects, such as, debris, infiltration, root intrusion, sediment accumulation, obstruction and grease build-up fall within operational failure category (Opila 2011, EPA 2009).

- Hydraulic capacity failure: in general, hydraulic capacity failure occurs when demand is higher than pipe capacity. In other words, the pipe segment does not have adequate capacity to convey wastewater, without having any structural or operational problem. Hydraulic capacity failure may be the result of infiltration/inflow (I/I), where the groundwater and storm water enter the sewer system through connections, cracks and defects. Other factors including pipe deformation and inadequate slope along the pipe increase the risk of hydrophilic capacity failure. Inadequate pipe slope can be due to loss of pipe bedding or insufficient construction and design. Hydrophilic capacity failure is often a sign of other type of structural defects such as cracks, broken pipe, leaks and other factors.

Next section covers different physical, operational and environmental factors that affect condition of sewer pipes.

2.7 Factors Affecting Condition of Sewer Pipes

2.7.1 Introduction

Deterioration of pipe is a very complex process and several factors influence the service life of pipe networks (Atique, 2016; Opila, 2011; Lindner 2008, Chughtai, 2008; Yan and Vairamoorthy, 2003). Davies et al. (2001) provided a comprehensive review of previous studies on the factors that influence structural deterioration of rigid pipes and categorized them into three groups of pipe construction, operational and environmental factors. According to Jalalediny Korky et al. (2017) selecting proper construction and equipment can affect performance of underground buried infrastructures. In other study, Al Barqawi and Zayed (2006) classified these factors into three categories; physical, environmental, and operational for water pipes, as shown in Table 2-11.

Table 2-11 Factors Affecting Water Pipes Deterioration

(Al Barqawi and Zayed, 2006)

Physical factors	Environmental factors	Operational factors
Pipe age	Climate	Backflow potential
Pipe diameter	Disturbances	Flow velocity
Pipe installation	Groundwater	Leakage
Pipe lining and coating	Pipe bedding	
Pipe manufacture	Pipe location	
Pipe material	Seismic activity	
Pipe wall Thickness	Soil type	
Type of joints	Stray electrical currents	
	Trench backfill	

Typically, agencies and municipalities have enough physical data, but environmental and operational factors are often unavailable (Salman, 2010). According to Kley and Caradot (2013), it is important to identify the factors that influence deterioration of sewer pipes due to following reasons:

- Data collection is a very expensive process during condition assessment and gathering all the pipe information is not a cost-effective approach. Identification of significant factors decreases the number of required features and reduces data collection costs.
- High prediction accuracy can be achieved when more significant factors are used in the model.

Table 2-12 provides summary of variables used to develop condition prediction models in previous studies.

Table 2-12 Factors Affecting Sewer Pipes Deterioration

Physical Factors	Environmental Factors	Operational Factors
End invert elevation	Backfill type	Blockages
Installation method	Bedding material	Burst history
Joint type	Ground movement	Debris
Pipe length	Groundwater level	Flow velocity
Pipe shape	pH	Hydraulic condition
Pipe slope	Road type	Infiltration/exfiltration
Pipe age	Root interference	Previous maintenance
Pipe depth	Soil corrosivity	Sediment level
Pipe material	Soil fracture potential	Sewer Type
Pipe size	Soil moisture	
Start invert elevation	Soil type	
	Soil sulfate level	
	Traffic characteristics	
	Vehicle flow	

2.7.2 Physical Factors

2.7.2.1 Pipe Age

Pipe age is normally the difference between the pipe installation year and the date of inspection. The age of pipe begins to start after the minute the sewer pipe is installed (Kulandaivel, 2004). Aging is one of the most important factors during pipe deterioration. Bathtub curve is a plot that determines rate of pipe failure depending on the age of the pipe. As showed in Figure 2-8, bathtub curve involves three distinct phases. The first phase is the early life period with a high failure rates that shows the failures right after installation. In this phase, the failures can be occurred because of human factors, pipe damage during construction and installation, and inappropriate pipe material. The second phase shows useful life of the pipe and the frequency of failure rate is very low and almost constant in this phase. Several random phenomena such as extreme heavy loading, earth movement, settlement or third-party interference can be the result of failures in second phase. And, in the third phase (wear-out life) the frequency of failure is high due to pipe deterioration and aging (Singh and Adachi, 2013).

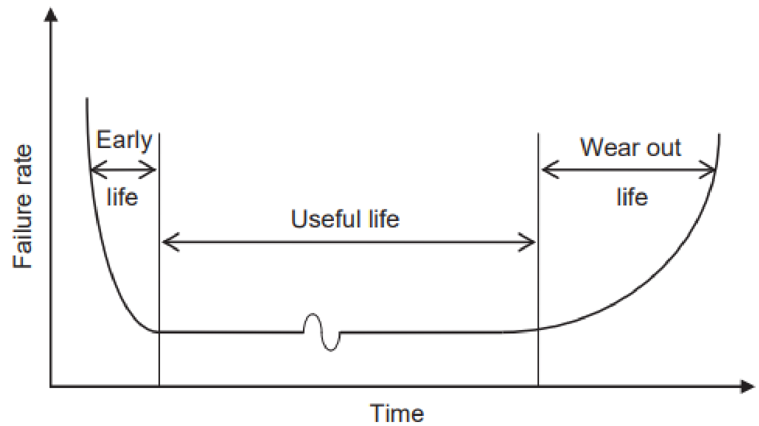


Figure 2-8 The Theoretical Bathtub Curve of Buried Pipe

(Singh and Adachi, 2013)

Most of the condition prediction models developed in previous studies proved that pipe age has strong relationship with deterioration of sewer pipes. Ariaratnam et al. (2001) stated that pipe age significantly affects deterioration of sewer pipes due to the consequence of pipe aging process. Jeong et al. (2005) and Ana et al. (2009) achieved similar result and indicated that the deterioration rate is lower during the early years of pipe service life and higher during the later years. Khan et al. (2010) specified that deterioration of sewer pipes does not start right after the installation process but arises after a certain period of time. Lubini and Fuamba (2011) demonstrated that with aging the sewer pipes root intrusion keeps growing and pipe roughness is increased gradually. Since, there is a direct relationship between pipe roughness and friction factor, hydraulic performance of the pipe will be dropped and the likelihood of pipe deterioration increases. Pipe age was found significant in more prediction models developed by Salman and Salem (2012), Kabir et al. (2018) and Laakso et al. (2018).

In contrast, Davies et al. (2001) found that the pipe age is not a significant variable in sewer pipe deterioration model, however their dataset did not include age of each individual sewer pipes and property age was the reference value of pipe age. Tran et al. (2006) found similar result and indicated that pipe age is not significant factor since structural deterioration of sewer pipes depends on combined effect of various factors.

2.7.2.2 Pipe Material

Sewer pipes constructed with different material have different reaction to the environmental factors, such as soil type, water table, etc. (Salman, 2010). For example, concrete pipes are highly resistant to abrasion and clay pipes act very well against acids. Plastic pipes, such as PVC or HDPE, resist to acidic and alkaline wastes, however they can suffer excessive deformations under loading (Singh and Adachi, 2013). Pipe material can be used as an independent variable during development of condition prediction models and it is possible to identify whether this variable is significant or insignificant through the results of the model. Davies et al. (2001) identified that pipe material is a significant variable and there is a direct relationship between deterioration of sewer pipes and pipe material. Micevski et al. (2002) described that concrete pipes are stronger and more durable than clay pipes based on the results of their Markov model. Ana et al. (2009) indicated that concrete pipes showed better behavior in the model than bricks and clay pipes. One probable reason for the difference in aging behaviors of pips is the production procedure of the pipes. Concrete pipes are typically constructed offsite, in a controlled environment condition and resulting high quality and integrity. While, the brick pipes usually are constructed in situ and different environmental condition and poor workmanship affect the quality of the pipes.

Pipe material was also significant in the model developed by Lubini and Fuamba (2011). They found that reinforced concrete pipes are the most resistant to deterioration than other pipes due to reinforcing steel that makes the conduit strong enough to prevent structural deterioration. Bakry et al. (2016) demonstrated that vitrified clay pipes behave better than asbestos cement and reinforced concrete pipes in their model. In prediction model developed by Laakso et al. (2018), concert and polyethylene high-density pipes were found significant. A possible explanation for the different behavior of pipe material in their study was that, the quality of certain batches of polyethylene high-density pipes were deficient. In contrast, Jeong et al. (2005) stated that pipe material was not a significant variable in their study. According to their report, one probable reason could be the class imbalance and low number of data that they used to develop the prediction model. In general, deterioration behavior of pipes could be predicted better if separate models generated for different pipe material.

2.7.2.3 Pipe Diameter

Numerous studies investigated the relationship between sewer size and deterioration of the pipes, and the result is contradictory. Some condition prediction models identified that sewer deterioration rate decreases with increasing the diameter and in contrast some other studies found that smaller diameter pipes have more failure. Pipe diameter was found a significant factor affecting deterioration of sewer pipes in several prediction models. Ariaratnam et al. (2001) indicated that when pipe diameter increases the likelihood of a pipe being in a deficient condition decrease. Davies et al. (2001) discussed that the risk of rigid sewer pipes being in poor condition decreases significantly with increasing diameter and larger sewers are at a lower risk than small ones. They mentioned this result may be to the fact that the structural design of rigid sewer pipes is restricted to the cross section of the pipes and the ring or crushing stress experienced.

Micevski et al. (2002) found that deterioration of smaller pipes was greater than the larger pipes. A probable explanation could be that the pipe designers underestimate the required depth of cover and loading traffics for the smaller pipes. Tran et al. (2009) indicated that larger pipes are often buried deeper and more appropriate design and construction crew are used to install them, therefore larger diameter pipes are more resistant to deterioration based on the result of developed prediction model. Lubini and Fuamba (2011) determined that with the occurrence of obstacles in the conduit, segments with larger diameter still enable to convey wastewater and small diameters are more likely to deteriorate due to lose of hydraulic flow. Salman and Salem (2012) and Bakry et al. (2016) found same result and in their prediction model, larger pipes behave better than smaller pipes.

On the contrary, Jeong et al. (2005) stated that larger pipes are more likely to deteriorate, since they have more surface area exposed to sewage and surrounding soil areas. The larger pipes are more at risk of damage because they are heavy and bulk, and it is difficult to install them accurately. Khan et al. (2010) found a dual behavior in the variation of pipe diameter and condition levels of sewer pipes. According to the results of their prediction model, smaller diameter pipes are more stable as compared to the larger pipes. No adverse effect was found in condition of pipes for diameter up to 24 inches. While, sewer pipes larger than 24 inches had lower deterioration rate than smaller pipes. Laakso et al. (2018) determined that pipes with 12- and 60-inches diameter were in better condition due to more carefully supervision during design and installation phases.

Tran et al. (2006) and Ana et al. (2009) found that pipe diameter is insignificant variable in their model, however they demonstrated that larger pipes are usually buried deeply and have lower deterioration rate than smaller one.

2.7.2.4 Pipe Length

Practically in all sewer pipe inventories, length of pipes is stored as manhole to manhole length of pipe segments, since CCTV is the most common tool for inspecting the sewers. Typically, longer sewer pipes have higher deterioration rate because the probability of occurring defects is more in longer pipes. However, in previous studies a dual behavior was found in the effect of pipe length to deterioration rate. Davies et al. (2001) indicated that risk of pipe being in poor condition decreased when individual pipe sections had longer length (more than 5 feet). Pipe joints are the main source of infiltration in pipelines and it can result in the movement of soil into a sewer and lack of support leading to structural instability. Longer individual pipe section means that the number of joints per manhole to manhole length of sewer pipes is reduced and then the risk of infiltration and joint defects also reduced. Jeong et al. (2005) indicated that longer sewers pipes are less likely to deteriorate than shorter one. A probable explanation could be that longer pipes have fewer bends in which less debris and fewer blockages or damages occur along the pipe length.

Ana et al. (2009) discussed that risk of pipe deterioration increases when sewer pipes are longer. This result could be attributed to the fact that longer pipes have more points and areas of possible failure specially in joints. Joint defect is one of the common defects in sewer systems and increases the probability of failure. Additionally, longer pipes are more vulnerable to have blockage and sediment deposition which facilitate the deterioration of sewer pipes. Khan et al. (2010) found a dual behavior in the condition of pipes with respect to changes in pipe length. Pipe segments smaller than 230 ft have no effect on the condition of sewer pipes. While pipes longer than 230 ft increase the rate of deterioration due to density of end joints which are source of break, infiltration and exfiltration. Salman and Salem (2012) determined that longer pipes behave better in sewer network because as the length of the pipe increases, the level of exposure to deteriorating factors also increases. Laakso et al. (2018) identified that sewer pipes beyond 131 ft in length deteriorate faster than other pipes in network. This outcome can be explained by the higher potential of defects and bending stress in longer pipes. Additionally, lateral connections are a potential cause of structural damage and longer pipes have more lateral connections.

2.7.2.5 Pipe Slope

Sewer pipes with flat slopes have lower velocities and then wastewaters remain longer time inside the pipe. The longer the wastewater remains in the sewer pipes the more probable is generation of hydrogen sulfide gas inside the sewer. Hydrogen sulfide can be converted to sulfuric acid and attacks the cementitious pipes, such as, concrete and mortar, and increases the rate of corrosion inside the sewer pipes (Ana et al., 2009; Ayoub et al., 2004). Similarly, Baur and Herz (2002) indicated that sediment deposition and clogging occur more into the pipes with flat slops and risk of deterioration is higher in these pipes. Jeong et al. (2005) indicated that when the pipe slope is steeper, the probability of deterioration is higher in sewer segments. Faster flow rate and lower stability are the probable cause of higher deterioration rate based on the results of this study. Tran et al. (2006) have similar finding and suggested that pipes with steeper slope are vulnerable to more defects due to void in the soil, soil movement and pipe joint defects.

Prediction model developed by Salman and Salem (2012) revealed the significance of pipe slope and according to their result steeper pipes are more likely to deteriorate due to stability issues and high flow rate. Laakso et al. (2018) identified that negative and very low slope was the most harmful condition for sewer pipes based on the result of their prediction model. Negative slopes and extremely low slopes cause inadequate rinsing, which can lead to debris accumulation and blockages. Pipe slope was found insignificant variable based on the results of Tran et al. (2006), Ana et al. (2009), Sousa et al. (2014) and Kabir et al. (2018).

2.7.2.6 Pipe Depth

Several factors such as soil type, water table, pipe material, pipe diameter and regulations must be considered to identify the appropriate depth of sewer pipes. The results of investigating the effect of depth on deterioration of sewer pipes is contradictory in different prediction models. Khan et al. (2010) indicated that pipe depth is a significant variable in their prediction model and any increase in depth has a negative effect on sewer pipe condition level. The rational reason for this behavior could be the greater dead load over the pipes and also higher probability of ground water table.

In contrast, pipe depth was insignificant variable in prediction model developed by Davies et al. (2001). This is not to say that sewer depth does not affect deterioration of pipes when considered on its own, but in data analysis based on the features of pipe datasets, there may not be a direct relationship between pipe depth and condition level of sewer pipes. Tran et al. (2006) and Ana et al. (2009) reported that pipe depth was insignificant in their prediction models. Generally, shallowly buried pipes would be subjected to more defects and higher deterioration rate due to surface load, illegal connections and tree root intrusion. Additionally, more cover depth above the pipes decreases the effect of surface factors such as road traffic and road maintenance or construction activities. Salman and Salem (2012) found same result and among the eight independent variables used in their model, pipe depth was the only insignificant variable in the model. Laakso et al. (2018) demonstrated that installation depth between 6 and 10 ft had correlation with poor condition level in their study and they recommended a minimum depth of 5 ft due to the frost in the winter.

2.7.2.7 Pipe Shape

The effect of pipe shape has been investigated only in few studies. Ana et al. (2009) demonstrated that the pipe shape is not a significant factor to examine deterioration of sewer pipes. However, Modica (2007) indicated the circular sewer pipes are stronger and show higher structural performance. Baur and Herz (2002) found deterioration rate of egg-shaped sewers are significantly slower in comparison to circular sewers.

2.7.3 Environmental Factors

2.7.3.1 Sewer Location

A sewer pipe obviously can be affected by applied load from the surface. Land use and traffic above the sewer pipe affect the magnitude of surface loading carried to the pipe. It is very difficult to measure or estimate the magnitude or frequency of surface loads because they vary in time (Kley and Caradot, 2013). The sewer pipe can be subjected on large one-time loads, such as, surface construction, ground utility construction, landslide and earthquakes, or a small cyclic load with hourly, daily or seasonal frequency, such as, bus stop, traffic and maintenance activities (Ashoori et al., 2017; Marlow et al., 2009).

The influence of road type on deterioration of sewer pipes just investigated in a few studies. Davies et al. (2001) determined that sewer located under rural main roads and sidewalks were at a significantly lower risk of being in poor condition than those pipes located under urban main roads. The main reason of this difference may be the more significant traffic loading in urban areas. Pipe location was also significant based on the result of prediction model generated by Tran et al. (2009). They stated that location of pipes determines depth of cover and when the cover is large the structural deterioration could be low due to the less amount of load on pipes. Salman and Salem (2012) demonstrated that pipe segments under local streets and alleys are less likely to deteriorate than pipe segments located under gardens or any type of roadways. One probable reason can be application of better design and installation process for pipes under the roads. Bakry et al. (2016) indicated that deterioration of sewer pipes is higher when they are located under industrial zones, while they have better behavior when serve a residential zone. It is obvious that sewage carried by sewer pipes in industrial zones have different characteristics and cause faster deterioration mechanism.

In contrast Tran et al. (2006) indicated that pipe location is not a significant variable in their prediction model. The effect of any critical environments such as coastline and industrial zones were not satisfactory to be considered as an influence variable. Ana at al. (2009) achieved similar results and the deterioration rates of sewer under light traffic and main roads were not significantly different in their model.

2.7.3.2 Groundwater Level

Groundwater is the water found underground in the cracks and spaces in the soil, sand and rock. The availability of ground water at or above sewer pipelines may cause water flowing through the pipe, increasing the structural defects, formation of void and loss of sewer support. In cohesive soil, raising the groundwater level may cause a reduction in the soil cohesive strength and growing the void around the pipe. Consequently, supporting soil can be washed (loosed) easily and the pipe is more likely to collapse in this condition. Typically, sewers located in areas subjected to high groundwater are more likely at a risk of failure than sewers located in an area where the groundwater level is below sewer level. Davies et al. (2001) described that the availability of groundwater around the pipe causes the loss of soil support and infiltration defect. Additionally, formation of void and lack of proper support around the pipe lead to sewer structural problems. Periodic water table in a cohesive soil may result in a reduction of soil strength and the possibility of soil being washed into the sewer. Malek Mohammadi et al. (2019) found groundwater level is a significant variable based on prediction model developed for City of Tampa. They indicated that groundwater increases the amount of load on pipe and the risk of soil movement and infiltration. Typically, groundwater level is unavailable in pipeline inventories and it was used as a variable jus in few prediction models. More research is needed to evaluate the effect of groundwater level on condition of sanitary sewer pipes.

2.7.3.3 Soil Type

Type of soil is one of the important factors can affect the ground loos and stability of the sewer pipeline. For example, when soil is subjected to stress, it may behave different due to various swell or shrinkage factors (Davies et al., 2001). Different types of soil have different reactions with pipe material, groundwater, and other pipe attributes or environmental factors (Kaushal, and Guleria, 2015). The fracture of soil was investigated by Davies et al. (2001) and, the result showed that sewers buried in a soil with a very high fracture potential, has significantly higher resistance to deterioration and failure. Typically, fracture is very high in clay soils. Wirahadikusumah et al. (2001) stated that surrounding soil is a significant factor affecting the deterioration of sewer pipes. According to this study, sewer defect size, hydraulic conditions and soil properties are the main factors affecting the rate of ground loss. The sewer pipe can be moved when loss of soil support occurs around the pipe. The loss of ground or soil support causes void formation around the pipe and therefore the sewer pipe is more likely to collapse or deform in this condition. Soil type also was found a significant variable according to the result of Micevski et al. (2002) Markov model. They revealed that pipes in alluvial soils are deteriorated faster than those in podzolic soils. Alluvial soils are deposited from a saline environment and have much more corrosive properties. While, podzolic soils are formed through the weathering of rocks. The significance level of soil type might be the result of the different formation of these soil types. In contrast, soil type was insignificant in prediction model developed by Laakso et al. (2018). They mentioned the quality of soil data was not sufficient enough in their dataset and more study is needed to evaluate the effect of soil type on deterioration of sewer pipes.

2.7.3.4 Soil Corrosivity

Soil corrosivity is a soil characteristic that increases the probability of external corrosion on pipe surface. The rate of corrosion is highly influenced by the characteristic of the pipe material and surrounding soil around the pipe (Kaushal et al., 2018; Yajima, 2015). Davies et al. (2001) demonstrated that sewers buried into a high corrosive soil were at a significantly higher risk of deterioration. Mashford et al. (2011), showed that the soil corrosivity is a factor contributing to the deterioration of the pipe by increasing the corrosion. Just a few studies have evaluated the effect of soil corrosivity on deterioration of sewer pipelines.

2.7.3.5 Soil Erosion

According to Tan and Moore (2007), development of erosion void around the pipe causes pipe damage, due to water entrance through joints and fractures. Deterioration of soil support around the pipe is the most critical factor leading to structural damage and sewer pipe deterioration (Law and Moore, 2007). Spasojevic et al. (2007) studied the effect of soil erosion on condition of culverts and the result indicated that the voids can be filled by soils moving down from the springlines and the result is losing of ground support around the pipe. Moore (2008) stated that soil erosion decreases lateral ground support to the sewer, and bending moment increases eventually fracture the sewer.

Soil erosion is very important factor of ground loss and the influence of this feature has not been evaluated yet in any condition prediction models.

2.7.3.6 Soil pH

The soil pH is considered as the most important factor affecting underground corrosion. Almost, all the studies in the field of underground corrosion indicated that the pH of the soil increases the corrosion rate of buried pipes (Wasim et al., 2018). Hou et al. (2016) conducted a comprehensive research to evaluate the effect of soil pH on pipes made with different material. Based on the research outcome, cast iron pipes are more likely to be corroded in the same corrosive environments compared to steel pipes. The effect of soil pH was investigated more in water pipe systems. For instance, Rajani and Maker (2000) and Doyle et al. (2003) used soil pH as a feature to predict the remaining useful life of water pipeline. The outcome showed that the pH was not a significant factor to generate the model. Based on their results, pH alone is not a good indicator to predict the condition of pipes there is no positive relationship between pH and corrosion rate. In general, there is no direct relationship between pH and corrosion rate (Wasim et al., 2018). Range of pH can be described as alkaline (pH>7), natural (pH=7) and acidic (pH<7).

2.7.4 Operational Factors

2.7.4.1 Sewer Type

Sewer pipes can be classified into separate sanitary and storm sewer systems and combined sewer systems. In combined sewer system, a single pipe is used to transport domestic, commercial and industrial wastewater, and storm water to a selected disposal location. The concept of separate sanitary sewer and storm sewer system is to manage storm water and sanitary wastewater separately. In this method, two separate pipes are used to convey domestic, commercial, and industrial wastewater, and storm water to a selected disposal location. O'Reilly et al. (1989) found that the rate of deterioration is higher in combined sewer systems than sanitary sewers. They argued that generally combined sewer systems are constructed shallower than separate systems and the flow fluctuation is higher in combined systems. Davies et al. (2001) and Baur and Herz (2002) showed a lower deterioration rate in combined sewers due to more planning and engineering effort during construction of combined sewers. Ariaratnam et al. (2001) demonstrated that waste type is a significant variable and sanitary sewers are found to have the greatest effect on pipe deficiency, followed by storm and then combined sewers. Salman and Salem (2012) claimed that sanitary sewer pipes are more resistant to deterioration than combined sewers. This outcome can be explained by the higher potential of soil loss, infiltration and exfiltration in combined sewers due to the high flow velocity during rainfall events.

2.7.4.2 Sewer Velocity

A minimum velocity is required inside the sewer pipelines to prevent any settlement of solids and particles along the pipe and should at least occur once in a day. Otherwise, the settlement of material leads to obstruction of free flow and finally causing the complete blockage. The effect of velocity on deterioration of pipe was investigated by Koo and Ariaratnam (2006) and the outcome reflected that the velocity is not a significant factor to assess the performance of gravity sewer pipes. No more research was found with considering the velocity as an independent variable to build deterioration model.

2.7.4.3 Sewer Hydraulic Condition

Tran et al. (2006), built a deterioration model including sewer hydraulic condition as an input variable. The hydraulic condition was divided into three: good, fair and poor categories to predict the condition of sewer pipes. The result showed that the hydraulic condition is highly influence the deterioration of sewer pipelines. They achieved similar result by improving the model in 2008 and argued that there is a direct relationship between structural deterioration with hydraulic deterioration (Tran et al., 2008). However, Micevski et al. (2002) indicated that the hydraulic condition is not a significant factor to analyze deterioration of sewer pipes.

2.7.4.4 Sewer Maintenance

Davies et al. (2001) discussed that use of inappropriate maintenance method accelerates the deterioration rate of sewer pipelines. For example, high water pressure during pipe cleaning is one of the concerns regards increasing the defects along the pipe. In other example, sewer flushing technic may cause damage to the pipe wall during cleaning process (Najafi, 2016; Najafi and Gokhale, 2005).

2.8 Condition Prediction Models for Sewer Pipelines

2.8.1 Introduction

Condition prediction models can be used to forecast condition rating of sewer pipes by using information obtained from inspection databases. Prediction models can perform an essential role to generate a comprehensive prioritization plan as provide valuable information to forecast short-term and long-term behavior of sewer pipes. In general, utility companies and municipalities can forecast the future condition of their assets by generating deterioration models to identify the pipes that require maintenance, rehabilitation and replacement. The primary objective of sewer condition prediction models is to apply an appropriate mathematical technique to estimate future condition states of sewer pipes. Additionally, condition prediction models are capable to identify significant factors affecting deterioration of the pipes.

Current condition of sewer pipes is often assessed through inspection techniques, however, understanding the future condition of pipe systems needs a comprehensive deterioration model. Most of the sewer prediction models are developed by pipe data obtained from CCTV inspection to forecast the failure time of pipes based on condition rating standards.

Deterioration models for sewer pipelines are classified into different categories. Deterministic, probabilistic, statistical, physical and artificial intelligence models are the most common techniques used in previous studies (Altarabsheh, 2016; Kley and Caradot, 2013; Tran, 2007; Morcous and Lounis, 2004; Yang, 2004). Tran (2007) suggested deterministic and statistical models as a model-driven type and artificial intelligence-based models as a data-driven type. Typically, the structures of data model-driven are defined by the expert, while, the sample data demonstrates the structure of models in data-driven type.

A deterministic model is the inherent lack of randomness or stochastically and often used for phenomenon where relationships between components are certain (Tran, 2007). A deterministic model assumes that the output can be exactly predicted by given input variables. The basic explanation of a statistical model is a random variable X, which represents a quantity whose outcome is uncertain. In statistical models, the probabilistic nature of historical data is used to describe the model output as a random variable. In any statistical analysis, estimates are "best guesses" based on the condition of given historical data (Coles, 2011).

Artificial intelligence can be defined as “the study of mental faculties through the use of computational models” (Charniak and McDermott, 1985). In artificial intelligence models, the dependent variables are classified from a set of independent variables by learning from the available data. These models are appropriate to estimate ordinal condition ratings or nonlinear deterioration behaviors.

With growth of Artificial intelligence and statistical models, physical models are not very common to predict deterioration of sewer pipes due to their complexity. Furthermore, deterministic, probabilistic and statistical models can be in a same group based on their statistical nature. Thus, existing sewer deterioration models can be classified into two groups of statistical and artificial intelligence models as shown in Figure 2-9.

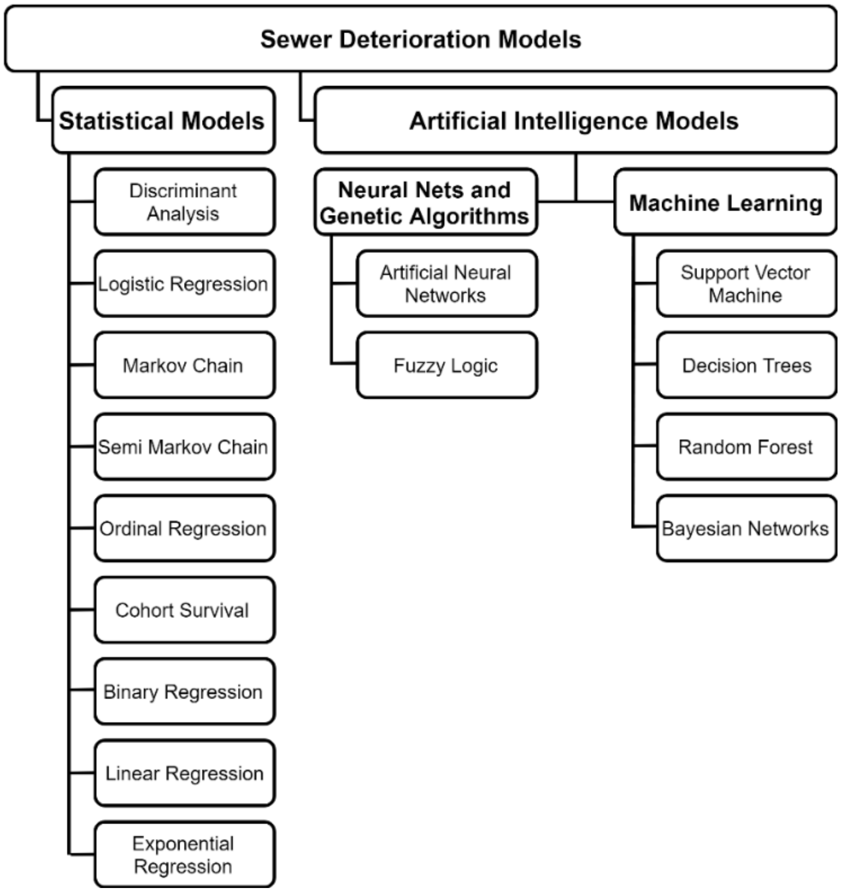


Figure 2-9 Classification of Sewer Deterioration Models

2.8.2 Statistical Models

The basic explanation of a statistical model is a random variable X, which represents a quantity whose outcome is uncertain. In statistical models, the probabilistic nature of historical data is used to describe the model output as a random variable. In any statistical analysis, estimates are "best guesses" based on the condition of given historical data (Coles, 2011). Dasu and Johnson (2003) indicated that parametric density function is used in statistical models to measure the errors and identify probabilistic relationships between dependent and independent variables. The results and outcomes of statistical models can be presented in probability values and they are more applicable to predict the current and future condition of sewer pipelines rather than deterministic models which provide quantitative results (Tran, 2007).

According to Tran (2007), predicting the ordinal data type and considering the probabilistic nature of the underlying deterioration process can be the advantages of statistical models. While, the sensitivity of statistical models to noisy data and the methodologies to measure the errors are disadvantages of these models. The sensitivity analysis by employing a Monte Carlo simulation, a powerful statistical analysis tool that is commonly used in both engineering and non-engineering fields and can assess the sensitivity of the output of the analysis with respect to each input variable (Habibzadeh-Bigdarvish et al., 2019). Numerous statistical models, such as, logistic regression, Markov chain, ordinal regression and cohort survival model were used to predict the condition of sewer pipelines in previous studies.

2.8.2.1 Linear Regression Models

The simplest linear regression model involves only one independent variable and the dependent variable can be predicted based on their relationship. The regression model states that true mean of the dependent variable changes at a constant rate as the value of independent variable increases or decreases. Therefore, the equation of a straight line shows the function relationship between the true mean of Y_i and X_i as shown in Eq. 2.1 (Rawlings, 1989).

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad \text{Eq. 2.1}$$

Where: i = facility index;

Y_i = dependent variable for facility i;

β_0 and β_1 = parameters to be estimated;

X_i = independent variable;

ϵ_i = random error term

Chughtai and Zayed (2007a, 2007b, and 2008) used the multiple regression technique to predict the deterioration mechanisms of sewer pipelines. Various factors, such as, pipe material, depth, length, age, diameter, bedding, road type and slope were considered as independent variables to build the model. Best subset analysis was used to select important variables in this paper. The significance of the variables was investigated by different statistical test including F-test, t-test, and residual analysis, lack of fit test and Durbin-Watson test. Four regression models were developed to predict the condition of concrete, asbestos, cement, and PVC pipes. The result showed 72 to 88% accuracy and they suggested inspection priority should be given to the pipes with extremely steep bed slopes.

Gedam et al. (2016) presented a condition prediction model for sewer pipeline by developing a linear regression model. Various factors, such as, pipe age, diameter, material and depth were contributed to build the model. The analysis revealed that the developed model can be used to assess the condition of sewer pipelines.

Bakry et al. (2016a, 2016b) used regression analysis technique to develop a condition prediction model for sewer pipes which rehabilitated before by CIPP method. The data was obtained from closed-circuit television (CCTV) inspection reports of Quebec CIPP rehabilitations. Various physical, operational and environmental factors were used to generate the models. The regression models were validated using coefficient of multiple determinations and the result revealed range between 80 to 97%. In addition, the accuracy of the models was determined by calculating mean absolute error and root mean square error. Linear deterioration curves were developed in this paper by examining the effect of increasing the age while changing the dependent variables.

In general, linear regression model is too simplistic to display the probabilistic nature of pipe deterioration and it is not an appropriate model to predict the discrete condition values (Madanat and Ibrahim 1995; Morcous et al., 2002; Tran, 2007, Moteleb, 2010). In addition, pipe deterioration is a complex process and the linear regression may not be an effective model to find the relationship between the independent variables and condition rating (Salman, 2010).

2.8.2.2 Exponential Regression Models

Exponential regression is a nonlinear model that goes beyond the simple summarization of the relationships displayed in a set of data. In nonlinear models, at least one of the expectation functions depends on at least one of the parameters. Nonlinear models are more realistic because the response can be better fitted with fewer parameters (Rawlings, 1989). Nonlinear models are more flexible than linear models and can be more appropriate than the use of transformations (Chatterjee and Simonoff, 2013). Eq. 2.2 presents the mathematical relationship between independent variables and the outcome.

$$Y_i = e^{\beta_1 + \beta_2 + \epsilon_i} \qquad \text{Eq. 2.2}$$

Mailhot et al. (2000) developed a predictive modelling strategy to determine the structural state of sewer networks. In this model only, age was considered to generate the prediction model. Wirahadikusumah et al. (2001), used a nonlinear optimization model to develop deterioration model for sewer pipeline. Pipe material, depth, groundwater level, and type of soil were used to generate the model. The result indicated that municipalities and sanitary districts need to inspect their assets routinely.

2.8.2.3 Markov Chains

The Markov chain was developed by Andrei Markov in 1906 as a discrete-time stochastic process. A Markov chain is a mathematical model of a random phenomenon over a unit of time to predict the future based on the present values and regardless of the past effects. The time can be discrete, continuous or ordered set (Konstantopoulos, 2009). Instead of deterministic objects, Markov chain deals with random variables. The Markov chain-based deterioration model assumes that conditional probability does not change over time and for all states *i* and *j* and all *t*, probability is independent of time as shown in Eq. 2.3 (Jeong et al., 2005).

$$P(X_{t+1} = j \mid X_t = i) = p_{ij} \qquad \text{Eq. 2.3}$$

Where, P_{ij} is the transition probability that given the system in state *i* at time *t*, it will be at state *j* at time (t+1). Generally, the transition probability matrix (m×m matrix) are used to calculate the transition probabilities. The transaction probability matrix is given in Eq. 2.4 and 2.5.

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \qquad \text{Eq. 2.4}$$

$$\sum_{j=1}^{m} P(X_{t+1} = j \mid (X_t = i)) = 1 \qquad \text{Eq. 2.5}$$

Then the probability of being in different states at time t+1, can be estimated by total probability theorem as shown in Eq. 2.6.

$$P_j^{t+1} = \sum_{j=1}^{j} P_{ij} \times P_i^t \qquad \text{Eq. 2.6}$$

where P_i^t is the probability of being in state i in year t (Micevski et al., 2002). Once the probability matrix is identified, the future condition of pipes can be easily obtained by Markov model.

For example, consider a set of pipe state condition, C = {C1, C2, C3, C4, C5}. When a sewer pipe is in condition 1 a series of probabilities P11, P12, P13, P14 and P15 determine the condition state of pipe in the next period. The deterioration process starts in one of the states and moves from one to another. If the sewer pipe is currently in condition C3, it moves to condition C4 in the next step with a probability of P34. This probability is called transition probability and only taking account the current condition of pipe without considering the historical data and previous conditions. Therefore, for sewer pipe with 5 scales condition, a transition matrix P can be developed as shown below:

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} \\ 0 & P_{22} & P_{23} & P_{24} & P_{25} \\ 0 & 0 & P_{33} & P_{34} & P_{35} \\ 0 & 0 & 0 & P_{44} & P_{45} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Extensive studies have been carried out to predict the deterioration of sewer pipelines by developing Markov chain models. Wirahadikusumah et al. (2001) used Markov-chains-based models in combination with nonlinear optimization for generating infrastructure management modeling for sewer pipes. In this study, a frequency analysis technique was used to develop transition probabilities of Markov deterioration model for large combined sewers in Indianapolis. The sewer database was divided into sixteen group and simple linear regression was developed to identify relationship between time and condition of pipes. The transition matrix was generated by assuming that the condition of sewer pipe moves to poorer condition or stays at current condition. It means pipe in condition 4 cannot improve and move to condition 2. And finally, a nonlinear optimization technique was used to minimize the sum of absolute difference between regression result and Markov chain estimations. The outcome of this study was deterioration curve for sewer pipes to illustrate the changes in condition states while the pipe is aging.

Micevski et al. (2002) developed a Markov model for the structural deterioration of storm water pipes. The pipe dataset was randomly categorized into two separate dataset and Bayesian techniques was used to identify the parameters of Markov model. The Metropolis-Hastings which is a member of the family of Markov chain Monte Carlo (MCMC) was used to calibrate the model. The validation of the model was performed through hypothesis testing to determine if the Markov model is appropriate for storm water pipe deterioration. The result indicated that the Markov model is consistent (at the 5% significant level) and can be used for storm water pipe deterioration. In addition, pipe diameter, construction material, soil type and exposure classification were found as significant variables that influence deterioration of pipes.

Kleiner et al. (2004) used a fuzzy rule-based, non-homogeneous Markov process to model the deterioration of buried pipes. The Markov procedures applied at each time step in two stages and a non-linear regression was used to train the model. The model could not be validated due to the lack of appropriate data for validation.

Jeong et al. (2005) used probit model-based approach to develop Markov chain deterioration model for wastewater infrastructure system. The model was generated by inspection database obtained from city of San Diego. Various factors, such as, pipe age, length, size, material and slope were considered as input variables. The result of study indicated that ordered probit model is an appropriate method to generate Markov chain model. They suggested using more input variables, such as, pipe depth, soil condition, groundwater level and sewage overflows, could be more effective to generate the deterioration models.

An ordered probit model was generated by Baik et al. (2006) to estimate the transition probabilities for a Markov chain-based deterioration for wastewater systems. The outcome reflected that the ordered probit model provides better result comparing with nonlinear optimization-based approach, however, it is necessary to have multiple time periods data for developing more accurate model. Sinha and McKim (2007) modeled deterioration of sewer pipes by Markov chain and a polynomial regression was used to determine the probability values of transition matrix.

Le Gat (2008) developed a mixed multi-state deterioration process by non-homogeneous Markov chains process to model the deterioration of urban drainage infrastructures. GompitZ analysis method was used to estimate the parameters of the time dependent transition probabilities through maximum marginal likelihood estimation. The GompertZ model considered a set of pipelines as a set of generic objects that are different based on their covariate values. The dataset was divided into different categories based on pipe diameter, sewer type and installation period. Cross validation method was used to split the data randomly for test and validation process. The result of this study indicated that a statistical model like GopmpitZ cannot predict the exact condition of a given pipe and only condition probabilities can be estimated. Another problem in applying GopmpitZ methodology is that calibration of this method is very difficult, and risk of misclassification is very high if population of pipes is not sufficient in database.

Scheidegger et al. (2011) developed a network condition simulator (NetCoS) to provide a synthetic population of sewer pipes based on historical inspection database. This model can be used to benchmark deterioration models and select an appropriate data management strategy. A semi-Markov chain technique was used to model deterioration of sewer pipes and transition probabilities. The deterioration of sewer pipes was defined by a set of survival function in this study. A survival function described condition states of sewer pipe based on age-dependent probabilities. Then semi-Markov chain computed the probabilities of changing the condition of pipes. The strength of NetCoS is that it is not limited to certain type of distributions and it is very flexible to generate more complex data. However, the main problem of this model is that it is not possible to validate the model by real-life data.

2.8.2.4 Logistic Regression

Logistic regressions are used to analyze the relationship between multiple independent variables and a categorical dependent variable. In logistic regression the probability of occurrence of an event is estimated by fitting data to a logistic curve. Logistic regression models can be classified in three groups of binary logistic regression, multinomial logistic regression and ordinal logistic regression (Park, 2013). Binary logistic regression is typically used when the response variable involves two categories (success or failure) and in the case of more than two response variable, multinomial logistic regression is applicable. For example, a binary logistic regression for deterioration of pipeline has two response variables of 0 and 1. If the outcome is equal to 0, pipe is in poor condition and in contrast response variable of 1 indicates that the pipe is in good condition.

For a binary response variable Y and a single explanatory variable X, let $\pi(\mathrm{X}) = \mathrm{P}(\mathrm{Y} = 1 \mid \mathrm{X} = \mathrm{x}) = 1 - \mathrm{P}(\mathrm{Y} = 0 \mid \mathrm{X} = \mathrm{x})$, the logistic regression model has linear form for the logit of this probability as shown in Eq. 2.7 (Agresti, 2007).

$$\operatorname{logit}[\pi(X)] = \log\left(\frac{\pi(X)}{1-\pi(X)}\right) = \alpha + \beta x \qquad \text{Eq. 2.7}$$

Eq. 2.8 presents the formula for the probability $\pi(\mathrm{X})$, using the exponential function $(\exp(\alpha + \beta x) = e^{\alpha+\beta x})$.

$$\pi(X) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \qquad \text{Eq. 2.8}$$

And the Eq. 2.9 presents the multiple logistic regression formula when multiple explanatory variables are used in the model (Agresti, 2007).

$$\log\left[\frac{\pi}{1-\pi}\right] = \log\left[\frac{P(Y = 1 \mid X_1, X_{2,} \dots, X_p)}{1 - P(Y = 1 \mid X_1, X_{2,} \dots, X_p)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p \mathrm{X}_p$$

$$= \alpha + \sum_{j=1}^{p} \beta_j \mathrm{X}_j \qquad \text{Eq. 2.9}$$

where:

X_1, X_2, …, X_p are independent variables

α is the intercept parameter for category *i*

β is the regression coefficients

And the probability than Y=1 can be measured using an exponential transformation as shown in Eq. 2.10.

$$\mathrm{P}(\mathrm{Y} = 1 \mid X_1, X_{2,} \dots, X_p) = \frac{e^{\alpha + \sum_{j=1}^{p} \beta_j \mathrm{X}_j}}{1 + e^{\alpha + \sum_{j=1}^{p} \beta_j \mathrm{X}_j}} \qquad \text{Eq. 2.10}$$

An important parameter in logistic regression is odds ratio that measures the relationship between explanatory and response variables as shown in Eq. 2.11.

$$\frac{\pi(X)}{1 - \pi(X)} = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x} \qquad \text{Eq. 2.11}$$

Multinomial logistic regression is used when multiple levels of categorical response variables are in the model. Eq. 2.12 shows the multinomial logistic regression formula.

$$\log\left[\frac{\pi}{1 - \pi}\right] = \log\left[\frac{P(Y = i \mid X_1, X_{2,} \dots, X_p)}{1 - P(Y = k \mid X_1, X_{2,} \dots, X_p)}\right] = \alpha + \beta_{i1} X_1 + \beta_{i2} X_2 + \dots + \beta_{ip} \mathrm{X}_p$$

$$= \sum_{j=1}^{p} \beta_{ij} \mathrm{X}_j \qquad \text{Eq. 2.12}$$

where:

i = 1, 2, …, K-1 correspond to categories of the dependent variable

$X_1, X_2, \dots, X_p$ are independent variables

α is the intercept parameter for category *i*

β is the regression coefficients associated with dependent category *i*

Logistic regression is widely used to model the deterioration of sewer pipelines. Davies et al. (2001) developed a logistic regression model to predict the structural condition of rigid sewer pipes. The main objective of this study was to identify influenced factors affecting deterioration of sewer pipes. Numerous factors, such as, Pipe length, debris, pipe size, sewer type, soil fracture potential, soil corrosivity, sewer location, groundwater level, pipe material and bus flow were used to develop the model. The condition of sewer pipes was divided into two categories of good and poor condition and the logistic transformation was used to estimate the probabilities. Stepwise forward and backward methods and binary logistic regression were employed in this study to select appropriate dependent variables. The result indicated that, pipe material, diameter, length, sewer type, location, groundwater and soil corrosivity are the influence factors that affect deterioration of sewer pipes. The main weakness of this study was that there is no information regarding validation and accuracy of the model. Additionally, only p-test was used to determine the significance of the dependent variables.

Ariaratnam et al. (2001) used logistic regression to predict condition states of sewer pipes by considering pipe age, depth, material, diameter and service types as independent variables. A linear regression variable selection method was used to specify the suitable independent variables in the model. Significance of the variables in this study was examined by Wald Test and likelihood-ratio test. The likelihood-ratio test revealed that pipe age, diameter and sewer types are the significant variables in the model. A sensitivity analysis was performed to validate the logistic regression model. However, sensitivity analysis is not enough to determine the performance of logistic regression model.

Koo and Ariaratnam (2006) generated a logistic regression model to predict the deterioration of sewer infrastructure systems. The data obtained from city of Phoenix, Arizona, was used to develop binary logistic regression and pipe age, maximum velocity and cumulative flow were considered as input variables to generate the model. Expert judgment was used to select pipe age, maximum velocity and cumulative flow as dependent variables in the model. They divided the dependent variables into three separate groups with combination of 27 sub-classes. P-test, Wald Test and likelihood-ratio test were used to assess the significance of the variables in the model. The result reflected that maximum velocity is not a significant factor in the model. The performance of logistic regression was not validated in this study.

Ana et al. (2009) investigated the influence of sewer physical properties on the structural deterioration of the sewer pipelines using logistic regression. Pipe age, size, depth, length, slope, shape, material, sewer type, construction period, and location were the factors considered in this study. They used the backward stepwise regression method for selection the predictor variables. The significance of the dependent variables was assessed by carrying out Wald Test and likelihood-ratio test. They also investigated the interaction effects of independent variables. For example, length of sewer pipes may be found insignificant in deterioration model but may become significant when combined with another independent variable. Sewer age, material and length were found significant in this study and no validation method was used to validate the result of logistic regression.

Tran et al. (2009) used multiple logistic regression to develop a model for predicting the structural condition of individual pipes. The predictive performances of model was compared using CCTV data collected for a local government authority in Melbourne, Australia. The independent variable used in this model were pipe size, age, depth, slope, trees, hydraulic condition, road type, and soil type. Maximum likelihood calibration model was used to calibrate the logistic regression model. The result indicated that other models such as neural network are more suitable for modeling the structural deterioration of sewer pipelines.

Lubini and Fuamba (2011) developed a logistic regression model for deterioration timeline of sewer systems. This model was applied to a case study in Quebec City, Canada and pipe age, diameter, material, length and slope were the contributing factors to generate the model. Several statistical tests such as overall model test, strength of association, likelihood-ratio test and Wald Test were used to assess the significance of independent variables. A deterioration curve was developed in this study for maintenance and operational planning. However, the performance and accuracy of the logistic regression model was not validated.

Salman and Salem (2012), employed three statistical models including ordinal regression, multinomial logistic regression and binary logistic regression to model the deterioration of wastewater collection lines. Several factors, such as, pipe size, length, slope, age, depth, material, sewer function and road class were used to calibrate the models. Five different ordinal regression were generated, and likelihood-ratio test was used to determine the relation of dependent and independent variables. The result indicated that none of the ordinal regression models satisfied the odds assumptions. Also, developed multinomial logistic regression obtained just 52% accuracy. Binary logistic regression was the only model that could predict condition of sewer pipes with 66% accuracy. This study provided different deterioration curves and equations which are useful to understand behavior of individual pipes in network. Moreover, logistic regression models were validated by confusion matrix and real data. The result of binary logistic regression revealed that pipe size, length, slope, age, material and sewer type are the significant factors in the model.

Logistic regression model was used by Sousa et al. (2014), to assess structural deterioration of sewer pipelines. Pipe material, diameter, length, age, depth and slop were the independent variables for generating the model. Moreover, in this study other techniques, such as, support vector machine and artificial neural networks were used to build the deterioration model. Based on research outcomes, the logistic regression provided the lowest correlation during the modeling. Furthermore, the authors indicated that due to overlapping among the models, it is not possible to select the best model in this study.

Kabir et al. (2018) developed a Bayesian logistic regression model to predict the structural condition of sewer pipelines. 12,728 sewer mains of the wastewater network of the city of Calgary, Canada, were selected to generate the model. Pipe age, material, diameter, length, slope, depth, rim elevation, up invert, and down invert were used to build the model. In this study Bayesian model averaging technique was used to identify significant variables and the condition of sewer pipes were predicted by logistic regression. P-test, Wald Test, likelihood-ratio test and Durbin-Watson test were employed to determine the significance of the independent variables. The condition states of sewer pipes were divided into two categories including good and poor conditions. The performance of the model was validated through confusion matrix. The main weakness of this model is that the pipe data was grouped base on pipe material and the model could not predict condition of pipe by considering all pipe material.

2.8.3 Artificial Intelligence Models

2.8.3.1 Introduction

The first artificial intelligence (AI) work was implemented by Warren McCulloch and Walter Pitts in 1943. knowledge of the basic physiology and function of neurons in the brain, propositional logic, and Turing's theory of computation, were three sources of introducing first artificial intelligence work (Russell and Norvig2010). Artificial intelligence can be defined as "the study of mental faculties through the use of computational models" (Charniak and McDermott, 1985). In other definition, AI is "The art of creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1990).

According to Luger (2009), the artificial intelligence can be decomposed into several categories as describes in below items:

- Game playing

- Automated reasoning and theorem proving
- Expert systems
- Natural language understanding and semantics
- Modeling human performance
- Planning and robotics
- Languages and environments for AI
- Machine learning
- Alternative representations: neural nets and genetic algorithms
- AI and philosophy

In artificial intelligence models, the dependent variables are classified from a set of independent variables by learning from the available data. These models are appropriate to estimate ordinal condition ratings or nonlinear deterioration behavior, however, as a disadvantage, the large amount of data is needed to generate artificial intelligence models (Scheidegger et al., 2011). Artificial intelligence models are capable to handle complex problems and in recent years extensive studies have been done to model deterioration of infrastructures using neural nets and machine learning methodologies.

2.8.3.2 Neural Nets and Genetic Algorithms

The objective of developing neural nets and genetic algorithms is to provide a model which works parallel the structure of neurons in the human brain (Luger,2009). A biological metaphor (human brain) was the reason of inspiration to invent both neural nets and genetic algorithms. These computing models include several interconnected unites or nodes that work similar the brain and the power of the model is highly dependent on the structure of nodes connections (Koehn, 1994).

The brain consists of 10^{11} neurons and the structure of brain is more complex than simple computer models. There are several neural net models that can be used for different functions and applications (Hagan et al., 2016). Among the neural net and genetic algorithm techniques, fuzzy set theory and neural networks (NNs) were used for modelling the deterioration of infrastructure facilities, (Flintsch and Chen 2004; Kleiner *et al.* 2004, Tran, 2007).

Artificial Neural Networks is one of the models used to predict the deterioration of sewer pipelines. Najafi and Kulandaivel (2005) employed artificial neural network to develop a prediction model. Pipe length, size, material, age, depth, slope and type of sewer were the variables used in this model. Backpropagation algorithm was used to train the data. The research concluded that application of neural network is feasible to develop condition prediction model for pipelines, however, the model accuracy is highly dependent on larger and more inclusive sample size.

The probabilistic neural network was used by Tran et al. (2006) to model structural deterioration of stormwater pipes. This study used a data set provided by the City of Greater Dandenong in Victoria, Australia, and approximately 650 data points was used to build the model. Pipe diameter, age, depth, slope, location, number of trees, hydraulic condition, soil type and soil moisture were the input variables considering in this model. The result showed that the probabilistic neural network works better than the discriminant models to predict deterioration of pipelines.

Tran et al. (2007) developed neural network deterioration model to predict serviceability condition of buried stormwater pipes. Markov Chain Monte Carlo simulation was used in this study to calibrate the model. Also, the ranking performance of neural network compared to multiple discrimination analysis model. Various independent variables, such as, pipe age, size, depth, slope, number of trees, road type, soil type, and moisture were used in this model. The research outcome reflected that the performance of neural network calibrated with Markov chain is better than neural network calibrated with backpropagation method.

In 2009, Tran et al. (2009) compared the performance of probabilistic neural network and multiple logistic regression models to develop prediction model for individual stormwater pipes. The predictive performances of model was compared using CCTV data collected for a local government authority in Melbourne, Australia. The independent variable used in this model were pipe size, age, depth, slope, trees, hydraulic condition, road type, and soil type. The maximum likelihood method was used to calibrate logistic regression model and neural network was calibrated using a Genetic Algorithm (GA). The result stated that neural network model is more suitable for modeling the structural deterioration of individual wastewater pipes.

Khan et al. (2010) developed a structural condition prediction model to investigate the importance and influence of certain characteristics of sewer pipes. Back propagation and probabilistic neural networks were used in this study to express condition rating of the pipes. The municipality of Pierrefonds, Quebec, provided the data used to develop this model. Pipe material, diameter, depth, bedding material, length and age were used to build the model. The developed models indicated that neural network is capable to prioritize inspection and rehabilitation plans for existing sewer mains.

Sousa et al. (2014), investigated the efficiency of artificial intelligence tools, such as, neural network, support vector machine and logistic regression for predicting sewer structural performance. Pipe material, diameter, length, age, depth and slope were the independent variables for generating the model. The research outcomes reflected that the different methods provided similar overall result, while the logistic regression providing the lowest correlations and the artificial neural networks the highest. Furthermore, due to overlapping among the models, it was not possible to select the best model in this study.

Hawari et al. (2016) developed a simulation-based condition assessment model for sewer pipeline using integrated fuzzy analytical network process (FANP). A weighted scoring system was used to determine the condition rating of sewer pipes and FANP determined weight of the factors affecting assessment of the pipelines. The result proved that the developed model is suited to assess the condition of sewer pipelines and it can be a useful tool for decision makers and municipalities.

Gheytaspour at al. (2018) developed a neural network model to forecast oxygen demand in wastewater treatment plants. Due to the increasing concerns over environmental effects of treatment plants considering the poor operation, fluctuations in process variables and problems of linear analyses, algorithms developed using artificial intelligence methods such as artificial neural networks have attracted a great deal of attention. In this research, first using regression analysis, the parameters of biological oxygen demand, chemical oxygen demand, and pH of the input wastewater were chosen as input parameter among other different parameters. Next, using error analysis, the best topology of neural networks was chosen for prediction. The results revealed that multilayer perception network with the sigmoid tangent training function, with one hidden layer in the input and output as well as 10 training nodes with regression coefficient of 0.92 is the best choice. The regression coefficients obtained from the predictions indicate that neural networked are well able to predict the performance of the wastewater treatment plant.

2.8.3.3 Machine Learning

In 1959, Arthur Samuel defined machine learning as a “Field of study that gives computers the ability to learn without being explicitly programmed” (Simon, 2015). Machine learning can learn directly from examples and experiences in the form of data, by exploring different prediction constructions and algorithms (Bishop, 2016). Typically, the predictive strength of machine learning models is used in industrial situations, especially when there is need to have a vision of future data (prediction) based on historical data.

Machine learning can be classified in three broad categories based on the nature of learning as described below (Bishop, 2016):

- Supervised learning: in supervised learning models, the training data includes examples of input variables with their corresponding output variables.
- Unsupervised learning: application in which the training data comprises a set of input variables without any corresponding output variables.
- Reinforcement learning: same as unsupervised learning, the output variables are not given in the model and the targets should be predicted by trial and error.

Another classification of machine learning can be based on the desired output of the modeling systems. Below items define these categories:

- Classification: the outputs are divided into two or more classes and typically supervised learning are used to model this class.
- Regression: in this category the outputs are continuous rather than discrete and a supervised problem.
- Clustering: in clustering category, a set of inputs are classified into different groups. Unlike classification and regression, this is an unsupervised task.
- Density estimation: the distribution of inputs is found in some space in this category.
- Dimensionality reduction: simplifying the inputs by mapping them into a lower-dimensional space.

The trend of using machine learning is growing very fast in different industries. various machine learning models, such as, support vector machine (SVM), decision trees, random forest and Bayesian regressions have been used in wastewater industry, to predict deterioration of sewer pipelines.

Mashford et al. (2011) employed support vector machine to predict condition grade of sewer pipelines. The predictive performances of model was developed using CCTV data collected from wastewater collection network in Adelaide, South Australia. The condition rating of 1 (good condition) to 5 (very poor condition) was used to assess the condition of sewer pipes. Pipe diameter, age, road type, slope, start invert, end invert, material, soil type, soil corrosivity, grade, angle, sulfate soil and groundwater level were used as input variables. The result of study showed that the support vector machine has very good predictive performance with 91% accuracy and can be used as a new approach to model deterioration of sewer pipes. The authors stated that the limitation of this study was lack of available condition data.

Syachrani et al. (2013) provided a decision tree-based model to study deterioration of buried wastewater pipeline. The combination of visual representation and sound statistical background was used to build the model. Same data set was used to compare the decision tree with conventional regression and neural network models. The databased used in this study were collected from Johnson County Wastewater (JCW) in Kansas. Pipe age, diameter, length, slope, number of trees and pipe defects were used as input variables in this model. The outcome reflected that; decision tree achieved more accuracy to predict the real age of sewer pipes.

Support vector machine was employed by Sousa et al. (2014) to investigate the efficiency of artificial intelligence tools, such as, neural network, support vector machine and logistic regression for predicting sewer structural performance. Pipe material, diameter, length, age, depth and slope were the independent variables for generating the model. The research outcomes reflected that the different methods provided similar overall result, with the logistic regression providing the lowest correlations and the artificial neural networks the highest. Furthermore, due to overlapping among the models, it is not possible to select the best model in this study.

Harvey and Mcbean (2014a) used random forests model to predict the structural condition of individual sanitary sewer pipes. The sewer database was collected from city of Guelph, Ontario, Canada. Several factors, such as, pipe age, material, length, diameter, service type, slope, up elevation, down elevation, depth, land use and road type were employed to develop the model. The research outcomes indicated that the random forest models are capable to predict the condition of individual sewer pipes by an excellent area under the ROC curve of 0.81. Using random forest prediction models has the potential to reduce the cost and time of projects and this strategy can be used to estimate the condition of uninspected sewer pipelines.

Harvey and Mcbean (2014b) published another paper about application of support vector machine and decision tree models for planning inspections of sewer pipelines. Similar to the previous research, data collected from city of Guelph, Ontario, Canada, was used to prepare the model. Pipe material, age, type of sewer, diameter, length, slope, down elevation, depth and road coverage were the input features of this model. The results stated that the support vector machine achieved 76% accuracy to predict the condition rating of sewer pipes. Although, decision trees were found to be a useful tool for planning prioritization and planning future inspection of sewer pipes.

Laakso et al. (2018) employed random forest and binary logistic regression to predict condition rating of sewer pipeline. Although, the factors that affecting the deterioration of pipes were investigated in this study. The databased used in this research were collected from southern Finland. European standards (EN- 13508-2) was used to assess the condition of sewer pipes. Score 0 indicated “no defect” and score 4 “serious defect." Various predictors, such as, pipe age, diameter, material, slope, depth, length, soil type, road class, distance to tree, intersection with stormwater or water supply pipes and annual sewage flow were used to generate the model. The accuracy of the models was 62% and 67% for binary logistic regression and random forest respectively. The result of the study indicated that both logistic regression and random forest models can be used to predict future condition of sewer pipelines.

In recent years, several sewer condition prediction models were developed and Table 2.13 shows detail of selected studies. Furthermore, Table 2.14 presents different variables included in prediction models.

Table 2-13 Sewer Condition Prediction Models

Authors	Year	Model	Number of Data	Condition Assessment Standard	Condition Rating Output
Davies et al.	2001	• Logistic regression	12,000	WRc	**0**: 1, 2, 3,4 **1**: 5
Ariaratnam et al.	2001	• Logistic regression	748	WRc	**0**: 1, 2, 3 **1**: 4,5
Wirahadikusumah et al.	2001	• Markov chain	-	Other	1,2,3,4,5
Micevski et al.	2002	• Markov chain	497	SEWRAT	1,2,3,4,5
Najafi and Kulandaivel	2005	• Neural network	-	PACP	1,2,3,4,5
Tran et al.	2006	• Neural network	583	WSAA	1,2,3
Koo and Ariaratnam	2006	• Logistic regression	579	PACP	**0**: 1, 2, 3 **1**: 4,5
Tran et al.	2007	• Neural network • Multiple discrimination analysis	150	WSAA	1,2,3
Chughtai and Zayed	2008	• Linear regression	-	WRc	1,2,3,4,5
Gat	2008	• Markov chain	5,262	DWA	1,2,3,4,5
Ana et al.	2009	• Logistic regression	1,316	NEN3399	**0**: 1, 2, 3 **1**: 4,5
Tran et al.	2009	• Neural network • Ordered probit model	417	WSAA	1,2,3
Khan et al.	2010	• Neural network	200	WRc	1,2,3,4,5
Mashford et al.	2011	• Support vector machine	1,441	Other	1,2,3,4,5
Lubini and Fuamba	2011	• Logistic regression	459	PACP	1,2,3
Salman and Salem	2012	• Ordinal regression • Logistic regression • Binary regression	11,373	PACP	**0**: 1, 2, 3 **1**: 4,5
Sousa et al.	2014	• Neural network • Support vector machine • Logistic regression	745	PACP	**0**: 1, 2, 3 **1**: 4,5
Harvey and McBean	2014	• Random forest • Decision Tree • Support vector machine	1,825	WRc	**0**: 1, 2, 3 **1**: 4,5
Bakry et al.	2016	• Multiple regression	84	PACP	1,2,3,4,5
Gedam et al.	2016	• Linear regression	155	PACP	1,2,3,4,5
Kabir et al.	2018	• Bayesian logistic regression	12,728	PACP	**0**: 1, 2, 3 **1**: 4,5
Laakso et al.	2018	• Random forest • Binary logistic regression	6,700	EN-13508-2	**0**: 0, 1, 2 **1**: 3,4

Table 2-14 Variables Included in Sewer Condition Prediction Models

($\color{green}\checkmark$: significant factors, $\color{red}\checkmark$: insignificant factors, $\checkmark$: not indicated)

Authors	Year	Independent Variables																			
		Age	Material	Diameter	Depth	Length	Slope	Sewer type	Location	Up invert	Down invert	Soil type	Bedding type	Groundwater	Corrosivity	Road type	No. Trees	Traffic	Flow	Hydrohalic	Other factors
Davies et al.	2001	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$		$\color{green}\checkmark$	$\color{green}\checkmark$					$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$					$\checkmark$
Ariaratnam et al.	2001	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$			$\color{green}\checkmark$													
Wirahadikusumah et al.	2001		$\color{green}\checkmark$		$\color{green}\checkmark$							$\color{green}\checkmark$		$\color{green}\checkmark$							
Micevski et al.	2002		$\checkmark$	$\checkmark$								$\checkmark$									
Najafi and Kulandaivel	2005	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$													
Jeong et al.	2005	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$		$\color{green}\checkmark$	$\color{green}\checkmark$														
Tran et al.	2006	$\color{red}\checkmark$		$\color{red}\checkmark$	$\color{red}\checkmark$		$\color{red}\checkmark$		$\color{red}\checkmark$			$\color{red}\checkmark$					$\color{red}\checkmark$			$\color{green}\checkmark$	$\checkmark$
Koo and Ariaratnam	2006	$\color{green}\checkmark$																	$\color{green}\checkmark$		$\checkmark$
Tran et al.	2007	$\color{red}\checkmark$		$\color{green}\checkmark$	$\color{red}\checkmark$		$\color{green}\checkmark$		$\color{red}\checkmark$			$\color{red}\checkmark$					$\color{green}\checkmark$				$\checkmark$
Chughtai and Zayed	2008	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$						$\color{green}\checkmark$			$\color{green}\checkmark$					
Gat	2008	$\checkmark$		$\checkmark$				$\checkmark$													
Ana et al.	2009	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{red}\checkmark$	$\color{red}\checkmark$												
Tran et al.	2009	$\color{red}\checkmark$		$\color{green}\checkmark$	$\color{red}\checkmark$		$\color{red}\checkmark$		$\color{green}\checkmark$			$\color{red}\checkmark$					$\color{red}\checkmark$			$\color{green}\checkmark$	$\checkmark$
Khan et al.	2010	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$							$\color{green}\checkmark$								
Mashford et al.	2011	$\checkmark$	$\checkmark$	$\checkmark$			$\checkmark$			$\checkmark$	$\checkmark$			$\checkmark$	$\checkmark$	$\checkmark$					$\checkmark$
Lubini and Fuamba	2011	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$		$\color{red}\checkmark$	$\color{red}\checkmark$														
Salman and Salem	2012	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$								$\color{green}\checkmark$					
Syachrani et al.	2013	$\checkmark$		$\checkmark$		$\checkmark$	$\checkmark$										$\checkmark$				$\checkmark$
Sousa et al.	2014	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$														
Harvey and McBean	2014	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$					$\checkmark$					$\checkmark$
Bakry et al.	2016	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$		$\checkmark$				$\checkmark$		$\checkmark$				$\checkmark$			$\checkmark$
Gedam et al.	2016	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$																
Bakry et al.	2016	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$			$\color{green}\checkmark$								$\color{red}\checkmark$		$\color{green}\checkmark$			
Hawari et al.	2016	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$			$\checkmark$			$\checkmark$	$\checkmark$	$\checkmark$							$\checkmark$
Kabir et al.	2018	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{red}\checkmark$			$\color{red}\checkmark$	$\color{red}\checkmark$										$\checkmark$
Laakso et al.	2018	$\color{green}\checkmark$	$\color{red}\checkmark$	$\color{red}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$	$\color{green}\checkmark$							$\color{red}\checkmark$	$\color{green}\checkmark$		$\color{green}\checkmark$		$\checkmark$

2.9 Chapter Summary

As described in previous sections, the deterioration of pipe is very complex process and only one factor cannot be the cause of pipe deterioration. Moreover, the wastewater agencies and municipalities are typically under budget to assess the condition of all pipes in the network periodically. Thus, an alternative solution must be used to reduce the inspection cost and to provide a comprehensive plan regarding prioritization and appropriate scheduling for inspection. Numerous deterioration models and several factors that affect deterioration of sewer pipes were presented in this chapter. However, condition prediction models for individual sewer pipes have not been fully examined yet and the result of most studies reflected that it is possible to assess future condition and behavior of sewer pipeline through new data analysis approaches. To this end, the objective of this dissertation is to model deterioration of individual sewer pipelines and investigate the factors that influence structural and operational condition of sewer pipes.

Chapter 3 Model Selection and Justification

3.1 Introduction

Deterioration models can be used to predict condition rating of sewer pipes by using information obtained from inspection databases. Prediction models can perform an essential role to generate a comprehensive prioritization plan as provide valuable information to forecast short-term and long-term behavior of sewer pipes. In general, utility companies and municipalities can forecast the future condition of their assets by generating deterioration models to identify the pipes that require maintenance, rehabilitation and replacement. The primary objective of sewer condition prediction models is to apply an appropriate mathematical technique to estimate future condition states of sewer pipes. Additionally, condition prediction models are capable to identify significant factors affecting deterioration of the pipes.

The existing sewer deterioration models can be classified into two groups of statistical and artificial intelligence models. The basic explanation of a statistical model is a random variable X, which represents a quantity whose outcome is uncertain. In statistical models, the probabilistic nature of historical data is used to describe the model output as a random variable. In any statistical analysis, estimates are "best guesses" based on the condition of given historical data (Coles, 2011).

Artificial intelligence can be defined as “the study of mental faculties through the use of computational models” (Charniak and McDermott, 1985). In artificial intelligence models, the dependent variables are classified from a set of independent variables by learning from the available data. These models are appropriate to estimate ordinal condition ratings or nonlinear deterioration behaviors.

The objective of this dissertation is to develop statistical and artificial intelligence models to predict future condition of sanitary sewer pipes. In this chapter detail and important properties of the selected models are presented.

3.2 Model Selection

One of the most important processes of any statistical analysis is model selection; because many factors can influence the result of regression models. Selection of deterioration models for sewer pipes depends on various factors, such as, the available information data, type and number of independent variables, and type of dependent variables. As explained before, the condition prediction scales are classified by discrete values for sewer pipes. Therefore, it is essential to select a predictive model with the capability to forecast categorical dependent variables.

Statistical models and artificial intelligence models were investigated comprehensively in literature review chapter. Some models such as, Markov chains, survival functions and simulation methods are appropriate to forecast condition of pipe networks or groups of pipes (Salman, 2010). Since, the objective of this study is to predict condition states of individual sewer pipes, group-based models were excluded from further investigation. Additionally, the condition states of sewer pipes are typically described as discrete or categorical values; therefore, linear and exponential regressions are not suitable to predict categorical variables since they minimize the sum of squared distances between predicted and actual condition ratings.

In this dissertation, the most appropriate models were selected based on the following reasons:

- Performance of the model to predict categorical dependent variables
- The capability of the model to be trained by nominal variables such as pipe material and soil type

- And, the results generated by the model

Logistic regression is the statistical model developed in this dissertation. The logistic regression model is the most frequently used regression model for the analysis dataset with two or more discrete outcome variables (Hosmer et al., 2013). Logistic regressions are used to analyze the relationship between multiple independent variables and a categorical dependent variable. Both numerical and nominal independent variables can be used to build a logistic regression model. And, significant factors that affect deterioration of sewer pipes can be identified by development of this model.

Gradient boosting is the second model developed in this study as artificial intelligence model. Boosting is one of the most powerful learning techniques presented in past twenty years and it is originally designed for classification problems. Gradient boosting is a machine learning technique for prediction and simulation with combining weak learners into a single strong learner (Hastie, 2017). This model can predict future condition of individual sewer pipes and determine the important independent variables. In this section, detail of logistic regression and gradient boosting models will be presented.

The third model developed in this dissertation is K-Nearest Neighbors (KNN) to predict the condition of sewer pipes. Nearest neighbors method works based on identifying the labels of K-nearest patterns in data space and predict the dependent variable based on the distance of the data points. K-Nearest Neighbors is developed in this study to satisfy the second objective of this dissertation about the diversity of the different statistical and artificial intelligence models and validating the result of logistic regression and gradient boosting tree models. Additionally, KNN is used for regression and classification and the application of this method is not well studied in this area.

3.3 Logistic Regression

3.3.1 Introduction

Logistic regressions are used to analyze the relationship between multiple independent variables and a categorical dependent variable. In logistic regression the probability of occurrence of an event is estimated by fitting data to a logistic curve. Logistic regression models can be classified in three groups of binary logistic regression, multinomial logistic regression and ordinal logistic regression (Park, 2013). In this dissertation binary and multinomial logistic regressions were used to develop condition prediction models. The detail of these models is explained in following sections.

3.3.2 Binary Logistic Regression

Binary logistic regression is used to develop prediction models when the output (dependent or response) variable is binary or dichotomous. A binary or dichotomous are variables which only take two values. For example, the output of the model can be true or false, success or failure and zero or one. In sewer condition prediction modeling, the dependent variable can be classified in good or poor conditions (Hosmer et al., 2013; Salman, 2010). A binary logistic regression for deterioration of pipeline has two response variables of 0 and 1. If the outcome is equal to 0, pipe is in poor condition and in contrast response variable of 1 indicates that the pipe is in good condition.

For a binary response variable Y and a single explanatory variable X, let $\mathrm{\pi(X)} = \mathrm{P(Y = 1 \mid X = x) = 1 - P(Y = 0 \mid X = x)}$, the logistic regression model has linear form for the logit of this probability as shown in Eq. 3.1 (Agresti, 2007).

$$\mathrm{logit}\,[\pi(X)] = \log\left(\frac{\pi(X)}{1-\pi(X)}\right) = \alpha + \beta x \qquad \mathrm{Eq.\,3.1}$$

And Eq. 3.2 presents the formula for the probability $\mathrm{\pi(X)}$, using the exponential function ($\exp(\alpha + \beta x) = e^{\alpha+\beta x}$).

$$\pi(X) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \qquad \text{Eq. 3.2}$$

Figure 3-1 illustrates logistic curve or logistic function which are used to estimate coefficient of the parameters in the model. In this example x varying from -4 to +4 while the y axes show the probability from 0 to 1.

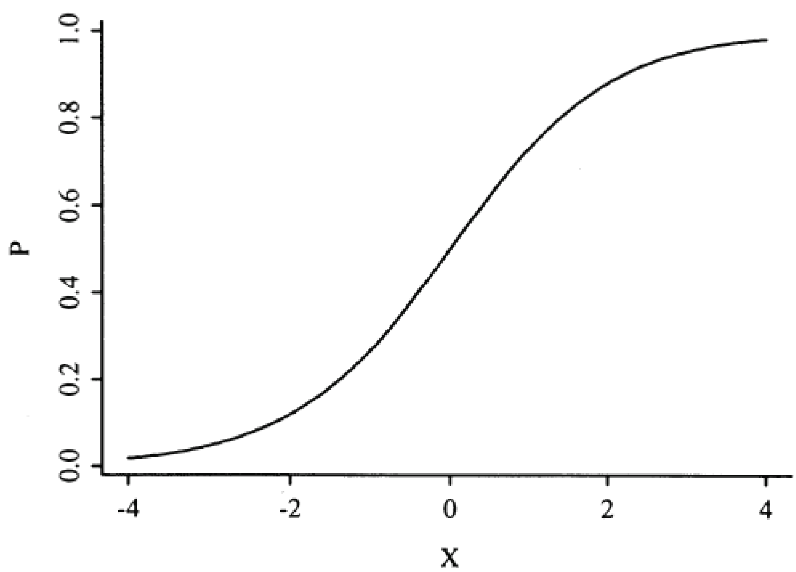


Figure 3-1 Logistic Function

(Harrell, 2016)

When there are more than one independent variables in database, multiple logistic regression is used to develop the model. The Eq. 3.3 presents the multiple logistic regression formula when the dependent variable is zero or one (Agresti, 2007).

$$\log\left[\frac{\pi}{1-\pi}\right] = \log\left[\frac{P(Y=1 \mid X_1, X_{2,} \dots, X_p)}{1 - P(Y=1 \mid X_1, X_{2,} \dots, X_p)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

$$= \alpha + \sum_{j=1}^{p} \beta_j X_j \qquad \text{Eq. 3.3}$$

where:

$X_1, X_2, \ldots, X_p$ are independent variables

α is the intercept parameter for category *i*

β is the regression coefficients

And finally, the probability of Y=1 can be measured using an exponential transformation as shown in Eq. 3.4.

$$\mathrm{P}(\mathrm{Y}=1 \mid X_1, X_{2,} \ldots, X_p) = \frac{e^{\alpha+\sum_{j=1}^{p} \beta_j \mathrm{X}_j}}{1+\ e^{\alpha+\sum_{j=1}^{p} \beta_j \mathrm{X}_j}} \qquad \text{Eq. 3.4}$$

The logistic model is easily understood by transforming the probability to a linear model, since the logistic regression is a direct probability model (Harrell, 2016).

3.3.3 Multinomial Logistic Regression

The multinomial logistic regression is used where the dependent variable is nominal with more than two levels. For example, consider a series of pipes which were assessed based on PACP method. The dependent variable has five levels indicating the condition of sewer pipes (condition 1 through 5). The objective of multinomial logistic regression in this case is to estimate the probability of having each of the five conditions and to convey the result in terms of odd ratio for choice of different conditions.

It would be possible to develop a multinomial logistic regression where the dependent variable has several levels, however, the details are more easily to understand for variables with three levels. When, a logistic regression is generated for a discrete dependent variable with more than two levels, the measurement scales should be investigated in more detail (Hosmer et al., 2013).

Equations 3.5 and 3.6 reveal how multinomial logistic regression works for a pipe system with three condition levels. Assume, levels 0, 1 and 2 indicate pipes in good, moderate and poor condition states respectively. Since, one of the categories is used as the reference value, two logit functions are required to develop the model. To develop the model, *p* covariate and a constant term denoted by the vector x (Hosmer et al., 2013).

$$\mathrm{g}_1(\mathrm{X}) = \log\left[\frac{P(Y = 1 \mid X)}{P(Y = 0 \mid X)}\right] = \beta_{10} + \beta_{11}X_1 + \beta_{12}X_2 + \cdots + \beta_{1p}\mathrm{X}_p \qquad \text{Eq. 3.5}$$

and

$$\mathrm{g}_2(\mathrm{X}) = \log\left[\frac{P(Y = 2 \mid X)}{P(Y = 0 \mid X)}\right] = \beta_{20} + \beta_{21}X_1 + \beta_{22}X_2 + \cdots + \beta_{2p}\mathrm{X}_p \qquad \text{Eq. 3.6}$$

Probability of each condition levels can be calculated by Eq.3.7 through 3.9.

$$\mathrm{P}(\mathrm{Y} = 0 \mid X) = \frac{1}{1 + e^{g_1(X)} + e^{g_2(X)}} \qquad \text{Eq. 3.7}$$

$$\mathrm{P}(\mathrm{Y} = 1 \mid X) = \frac{e^{g_1(X)}}{1 + e^{g_1(X)} + e^{g_2(X)}} \qquad \text{Eq. 3.8}$$

and

$$\mathrm{P}(\mathrm{Y} = 2 \mid X) = \frac{e^{g_2(X)}}{1 + e^{g_1(X)} + e^{g_2(X)}} \qquad \text{Eq. 3.9}$$

Multinomial logistic regression is known by other names such as polychotomous, or polytomous logistic regression in the health and life science (Hosmer et al., 2013). In the next sections more information will be given regarding significance of the models and variables.

3.3.4 Logistic Regression Assumptions

Binary logistic regression and multinomial logistic regression share the same assumptions (Salman, 2010). The assumptions of logistic regression are as follows (McDonald, 2009):

- The observations are independent and there is no relation between the outcome variables. In other word, the observations should not come from repeated measurements.
- The odds ratio and independent variables have a linear relationship.

- Logistic regression does not assume that the independent variables are normally distributed.
- There is no multicollinearity between independent variables. In other word, the correlation between independent variables should not be too high.

3.3.5 Forward and Backward Stepwise Selection

Forward and backward stepwise selection are statistical techniques to screen the independent variables. In these methods, the variables which have enough predictive power are remained in the model and idle variables are removed stepwise. For example, if a dataset has hundred independent variables it would be beneficial to keep the appropriate variables on the model and remove the rest.

Forward stepwise selection starts with the intercept and then the variables that improve the performance of the model are added sequentially. In contrast, backward stepwise selection starts with full model and then the variables that have least influence are deleted. The variables with the smallest Z-score are the candidate for removing from the model. Backward stepwise selection can only be used when total number of observations are greater than independent variables, while forward stepwise can always be used (Hastie et al., 2017).

3.3.6 Fitting the Logistic Regression Model

In logistic regression model, the intercept and the coefficient of each independent variables are estimated by Maximum Likelihood Estimation (MLE) method. In general, the method of maximum likelihood estimation assigns values for the unknown parameters that maximize the probability of obtaining the observed values. The maximum likelihood estimators of the parameters are the value that maximize likelihood function which expresses the probability of the observed data as a function of the unknow parameters (Hosmer et al., 2013). Equation 3.10 defines the detail of maximum likelihood estimation.

$$l(\beta) = \prod_{i=1}^{n} \pi(X_i)^{y_i} \left[1 - \pi(X_i)\right]^{n_i - y_i} \qquad \text{Eq. 3.10}$$

where, n_i is total number of observations, β is coefficient parameters, y_i is number of success and N is total number of observations.

3.3.7 Odds Ratio

The odds ratio (OR) is a nonnegative function, used to compare the relative odds of the occurrence of the outcomes. For a probability of success π, the odds of success are given in Equation 3.11. For example, if the probability of success is 0.8, then the odds of success equal 0.8/0.2 = 4 (Agresti, 2007).

$$odds = \frac{\pi}{1-\pi} \qquad \text{Eq. 3.11}$$

The odds ratio is a measure of association between an event occurring in one group, to the odds of it occurring in another group (Ana et al., 2009). Assume, the possible values of the logistic probabilities from a prediction model of sewer pipes including discrete dependent variable with condition states of 0 and 1. The odds ratio is the ratio of the odds for x = 1 to the odds for x = 0 as given by the Equation 3.12.

$$OR = \frac{odds_1}{odds_0} = \frac{\left[\frac{\pi(1)}{1-\pi(1)}\right]}{\left[\frac{\pi(0)}{1-\pi(0)}\right]} = e^{\beta_1} \qquad \text{Eq. 3.12}$$

The odds ratio is widely used to approximate how much more likely or unlikely is the outcome to be present in groups where x = 1 or x = 0. When, OR = 1 the outcome is equally likely in both groups of x =1 and x = 0. Odds ratio greater than one reveals that the outcome is most likely to happen when x = 1 and odds ratio less than one indicates that the event is less likely when x = 1 (Hosmer et al., 2013). For example, if the outcome (y), expresses the presence or absence of pipe failure and the independent variable (x) presents whether the groundwater is above or below the pipe, then an OR = 3 means that the odds of pipe failure is three time greater where the groundwater is above the pipes.

This simple relationship between the coefficient and odds ratio is one of the main reasons that logistic regression is widely accepted as a powerful analytical tool (Hosmer et al., 2013).

3.3.8 Significance of the Coefficients

Identifying the significant variables in the model is formulation and testing of a statistical hypothesis to determine if the independent variables are significantly related to the dependent variables. Typically, significance of the variables can be identified by comparing the observed dependent variables and predicted values after development of the model with and without independent variables. If the predicted values are more accurate by utilizing an independent variable in the model, then the variable is significant. Log-likelihood test and Wald test are the most common tests used in logistic regression to identify the significance of the variables (Hosmer et al., 2013).

3.3.8.1 Log-likelihood Test

In logistic regression, the log-likelihood function is used to compare the observed and predicted values. Equations 3.13 and 3.14 show the mathematical concept of log-likelihood function.

$$G = -2\ln\left[\frac{(likelihood\ without\ the\ variable)}{(likelihood\ with\ the\ variable)}\right] \qquad \text{Eq. 3.13}$$

$$G = 2\left\{\sum_{i=1}^{n}[y_i \ln\pi(X_i) + (1-y_i)\ln(1-\pi(X_i)] - [n_1\ln(n_1) + n_0\ln(n_0) - n\ln(n)]\right\} \qquad \text{Eq. 3.14}$$

where $n_1 = \sum y_i$ and $n_0 = \sum(1-y_i)$. This statistic is similar partial F-test in linear regression. For large samples, the statistic G follows a chi-square distribution with degree of freedom equal to the number of parameters estimated (Harrell, 2016).

3.3.8.2 Wald Test

Wald test is a method to identify significance of the individual variables in logistic regression models. The Wald test is equal to the ratio of the maximum likelihood estimate and its standard error as shown in Equation 3.15. This ratio follows a standard normal distribution (Hosmer et al., 2013).

$$W_j = \left(\frac{\beta_j}{SE(\beta_j)}\right) \qquad \text{Eq. } 3.15$$

where β_j is the coefficient of the predictor variable, and SE is the standard error of the coefficient. If the result of Wald test for an independent variable is zero, this variable is not significant, and it can be removed from model. In contrast, if Wald is not zero, the variables should be included in the model.

3.3.9 Classification Table

Classification tables are used to show the percentage of correct predictions by the logistic regression models. This table summarize the result of fitted logistic regression models. To obtain the discrete result of classification table, a cut-point (*c*) is defined (0.5 is the most common value) and it is compared to each estimated probability. If the estimated probability exceeds the cut-point, they are assigned to class one. In contrast, if the estimated probability does not exceed the cut-point, they are assigned to the other groups (Hosmer et al., 2013). The concept of classification table is matched with confusion matrix, but typically in logistic regression the term of classification table was used in different references (Harrell, 2016; Hosmer et al., 2013; Agresti, 2007).

3.4 Tree-Based Models

3.4.1 Introduction

In tree-based models, the feature space is divided into a set of rectangles and then a simple model is developed for each partition. Tree-based models are conceptually simple and powerful method for both regression and classification aim. For example, consider a regression problem with continuous dependent variable Y, and two independent variables X_1 and X_2. As shown in Figure 3-2, the space is split into several regions and then model is developed based on mean of dependent variable in each region. The regions are split into more regions until achieving the best fit for the model or applying some stopping rules (Hastie et al., 2017).

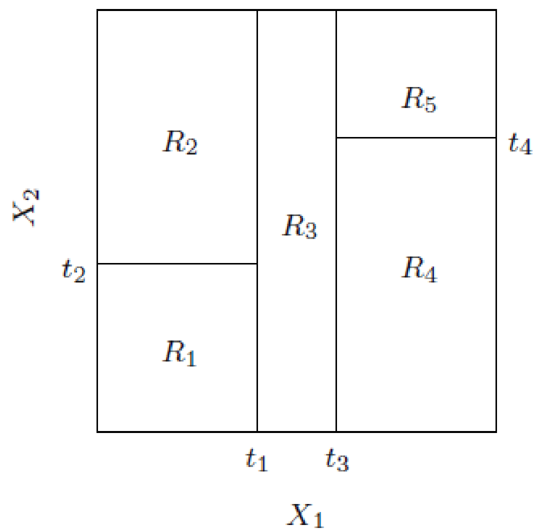


Figure 3-2 Tree-based Partitions

(Hastie et al., 2017)

In previous example, X_1 is split into t_1 and t_3 and X_2 is divided into t_2 and t_4. And the result is five regions R_1, R_2, …, R_5 shown in the figure.

3.4.2 Classification Trees

Classification trees are used as predictive models when the outcome taking discrete values 1, 2, …, *K* (Hastie et al., 2017). In other word, classification trees are used to classify an object into separate classes based on characteristics of input variables (Rokach and Maimon, 2015). Comparable with logistic regression, this method can be used to predict different condition levels of sewer pipes.

Regression trees utilize squared-error node to split the space into separate regions, while the classification trees use different criteria such as impurity-based criteria, information gain, and Gini index (Rokach and Maimon, 2015). For impurity-based criteria assume a random variable x with k discrete values. The variable is defined as pure if the probability vector x gets only one value. Information gain is an impurity-based criterion with utilizing the entropy measure to determine the split regions. This method works based on maximum likelihood estimation to make inferences about parameters of the underlying probability distribution. And Gini index is one of the most common techniques that measures the differences between the probability distributions of the dependent variables. Gini index measures how often a random event would be identified incorrectly. Therefore, a variable with lower Gini index should be preferred (Hastie et al., 2017).

The employment of decision tree is very common for classification due to its simplicity and transparency. Decision trees are self-explanatory and there is no need for data mining expert to follow a certain decision rule. Decision trees can present the graphically results which are easier to interpret than other modeling techniques, especially when the outcome result is complicated (Rokach and Maimon, 2015).

3.4.3 Gradient Boosting Tree

Gradient boosting is a machine learning technique for regression and classification, which provide a prediction model by improving the performance of a weak learner. In this method, a weak learner is run repeatedly on various training data to develop classifiers. Then, the classifiers are combined into a single strong classifier to achieve a higher accuracy (Rokach and Maimon, 2015).

In fact, gradient boosting tree is an ensemble model which employs the strengths of a collection of simpler base models to develop a prediction model (Friedman, J. 2001). Many recent machine learning approaches determined that prediction of an ensemble of models works better than only a single prediction model. The most frequent approaches to generate ensemble classifiers are bagging, boosting and random forest (Kozak, 2019).

AdaBoost is the most popular boosting algorithm. Consider a two-class dependent variable $Y \in \{-1, 1\}$. The error rate of training sample can be determined by Eq.3.16, where X is independent variable and G(X) is a classifier produces a prediction. A weak classifier has slightly better error rate than random guessing (Hastie et al., 2017).

$$err = \frac{1}{N}\sum_{i=1}^{N} I\big(y_i \neq G(x_i)\big) \qquad \text{Eq. 3.16}$$

Based on the objective of boosting, a series of weak classification algorithm are generated to provide a sequence of weak classifiers. Then the final prediction is developed from combination of weighted classifiers as shown in Figure 3-3 and estimated from Eq.3.17.

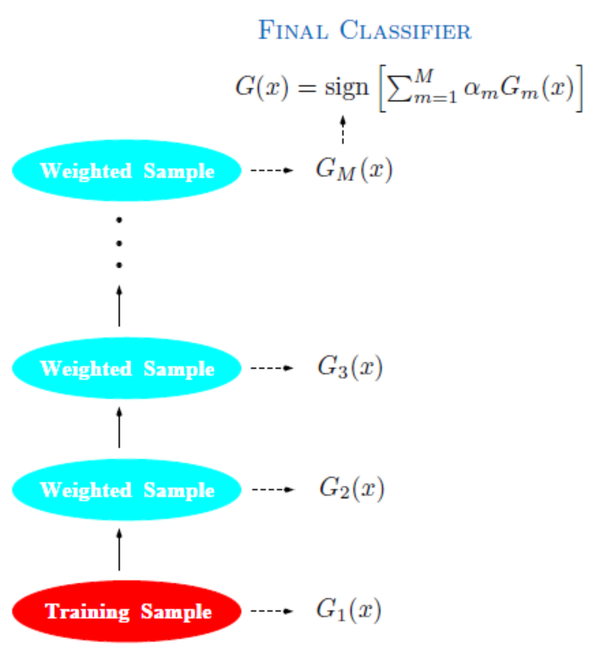


Figure 3-3 Schematic of AdaBoost

(Hastie et al., 2017)

$$G(x) = sign\left(\sum_{m=1}^{m} \alpha_m G_m(x)\right) \qquad \text{Eq. 3.17}$$

where $\alpha_1, \alpha_2, \dots, \alpha_M$ are computed by boosting algorithm and weight the contribution of classifiers.

The weights are assigned to each training observation sequentially and the classification algorithm is replied to the weighted observations. At the final step, the weights of misclassified observations are increased, while the weights are decreased for those which were predicted correctly.

3.4.4 Fitting Gradient Boosting

Boosting models employs a set of basic functions to fit an additive expansion. Equation 3.18 presents the form of basic function expansions (Hastie et al., 2017).

$$f(x) = \sum_{m=1}^{M} \beta_m \, b(x; \gamma_m) \qquad \text{Eq. 3.18}$$

where β_m are the expansion coefficients and $b(x; \gamma_m)$ are simple functions characterized by a set of parameters γ. In tree models, γ determines the split variables and points at the internal nodes and the predictions at the terminal nodes.

Typically, minimizing a loss function such as the squared-error or a likelihood-based loss function is used to fit the gradient boosting trees. Loss function is a machine learning technique to evaluate the prediction performance of the model. When the loss function is a high value, the model is not appropriate for prediction. While, the lower value of loss function determines the capability of the model to achieve better accuracy. Therefore, minimizing the loss function is a technique to increase the performance of the models. Equation 3.19 shows the detail of minimizing loss function in boosting trees.

$$\min_{\{\beta_m, \gamma_m\}_1^M} \sum_{m=1}^{N} L\left(y_i, \sum_{m=1}^{M} \beta_m \, b(x; \gamma_m)\right) \qquad \text{Eq. 3.19}$$

3.4.5 Importance of Independent Variables

Decision trees models are capable to rank the importance of the independent variables in both regression and classification aims. Typically, two measure of significance are used to determine the magnitude of the variables. The first technique is Mean Decrease Impurity (MDI) and the second one is Mean Decrease Accuracy (MDA) (Biau and Scornet, 2016).

Mean decrease impurity method measures the weighted decrease of impurity from splitting on the variable, averaged over all trees. In simple word, MDI counts the times that an independent variable is used to split a node as given in Eq.3.20. This method is used more to develop forest decision trees and gradient boosting (Biau and Scornet, 2016).

$$\widehat{MDI}\left(X^{(j)}\right) = \frac{1}{M}\sum_{l=1}^{M} \sum_{\substack{t \in \mathcal{T}_\ell \\ j^*_{n,t}=j}} p_{n,t} L_{class,n}(j^*_{n,t}\,,\ z^*_{n,t}) \qquad \text{Eq. 3.20}$$

where $p_{n,t}$ is the fraction of observations falling in the node *t*, $\mathcal{T}_\ell$ is the collection of trees in the forest and $(\,j^*_{n,t}, z^*_{n,t})$ the split that maximizes the empirical criterion in node *t*.

Mean decrease accuracy is one of the most interesting measures in tree based models, because it is based on averaging the difference in out-of-bag error estimation before and after the permutation over all trees. Equation 3.21 present the mathematical concept of MDA (Biau and Scornet, 2016).

$$\widehat{MDA}\left(X^{(j)}\right) = \frac{1}{M}\sum_{l=1}^{M} \left[R_n\left[m_n(.\,;\theta_l), D^j_{l,n}\right] - R_n\left[m_n(.\,;\theta_l), D_{l,n}\right]\right] \qquad \text{Eq. 3.21}$$

where $X^{(j)}$ is variable, $D_{l,n}$ is out-of-bag dataset, $D^j_{l,n}$ is the same dataset when the variable has been randomly selected.

3.4.6 Evaluation of Gradient Boosting Tree

Typically, the supervised learning techniques are trained by a set of data and the objective is to perform a model which can make prediction. The performance of the prediction models always must be evaluated to understand the quality of the model and to identify the important parameters in the model. There are several techniques to evaluate performance of machine learning models. In this dissertation, confusion matrix, Receiver Operating Characteristic (ROC) curve and Area Under Curve (AUC) are used to evaluate machine learning models.

3.4.6.1 Alternatives to the Accuracy Measure

In this technique, sensitivity and specificity measures are used to evaluate the accuracy of the model. Sensitivity expresses how well the classifier can predict the positive samples and specificity determines how well the negative samples are recognized by classifiers. Equations 3.22 and 3.23 show the detail of measurements (Biau and Scornet, 2016).

$$Sensitivity = \frac{true\ positive}{positive} \qquad \text{Eq. 3.22}$$

$$Specificity = \frac{true\ negative}{negative} \qquad \text{Eq. 3.23}$$

where true positive and true negative are number of true positive and true negative samples respectively, positive is number of positive and negative is number of negative samples. And then, the accuracy of the model can be specified as a function of sensitivity and specificity as given in Eq.3.24.

$$Accuracy = Sensitivity.\frac{positive}{positive + negative} + specificity.\frac{negative}{positive + negative} \qquad \text{Eq. 3.24}$$

3.4.6.2 Confusion Matrix

The confusion matrix is used to identify the number of elements that have been correctly or incorrectly predicted for each class. In confusion matrix, for every test samples the actual class is compared to the class that was assigned by the trained classifier. True positive or true negative (TP/TN) determine the examples that are classified correctly in the model. In contrast, false positive or false negative (FP/FN) identify the positive or negative examples that are classified incorrectly as shown in Table 3-1 (Biau and Scornet, 2016).

Table 3-1 Confusion Matrix

	Predicted positive	Predicted negative
Positive Examples *(P)*	True positive *(TP)*	False negative *(FN)*
Negative Examples *(N)*	False positive *(FP)*	True negative *(TN)*

Based on the values on Table 3-3 below measurements can be calculated in confusion matrix method (Biau and Scornet, 2016):

- Accuracy: (TN+TP) / (TP+FN+FP+TN)
- Misclassification rate: (FP+FN) / (TP+FN+FP+TN)
- Precision: (TP) / (FP+TP)
- True positive rate (recall or sensitivity): (TP) / (FN+TP)
- False positive rate: (FP) / (TN+FP)
- True negative rate (specificity): (TN) / (TN+FP)
- False negative rate: (FN) / (FN+TP)

3.4.6.3 ROC Curve

Receiver Operating Characteristic (ROC) curve illustrates the exchange between true positive to false positive rates. In ROC curve, the X-axis illustrates a false positive rate (specificity) and the Y-axis presents a true positive rate (sensitivity). When all positive examples are predicted correctly in the model, the best point on the ROC curve is (0,100) (Hastie et al., 2017). Therefore, when the ROC curve is closer to upper left corner, the overall accuracy of the model is higher. Figure 3-4 illustrates a typical ROC curve.

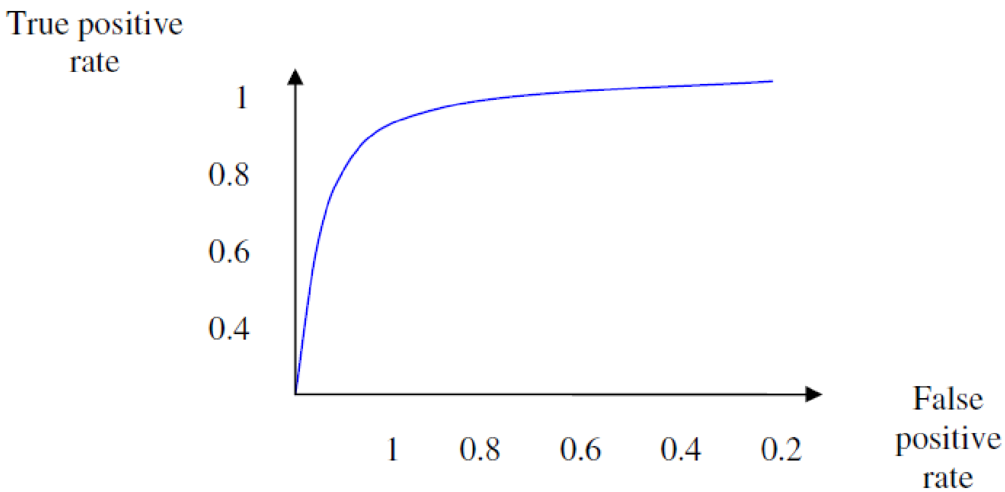


Figure 3-4 A Typical ROC Curve

(Biau and Scornet, 2016)

3.4.6.4 Area under Curve (AUC)

Area under ROC Curve (AUC) is a useful method to evaluate the performance of classification models since it is independent from prior probabilities and decision criterion. When the total area under curve is higher, the prediction performance of the model is better. The imbalance of the training set does not affect the area under ROC curve, therefore the comparison of AUC of two or more classification model is more suitable than comparing their misclassification rates (Biau and Scornet, 2016). Figure 3-5 illustrates an example of AUC.

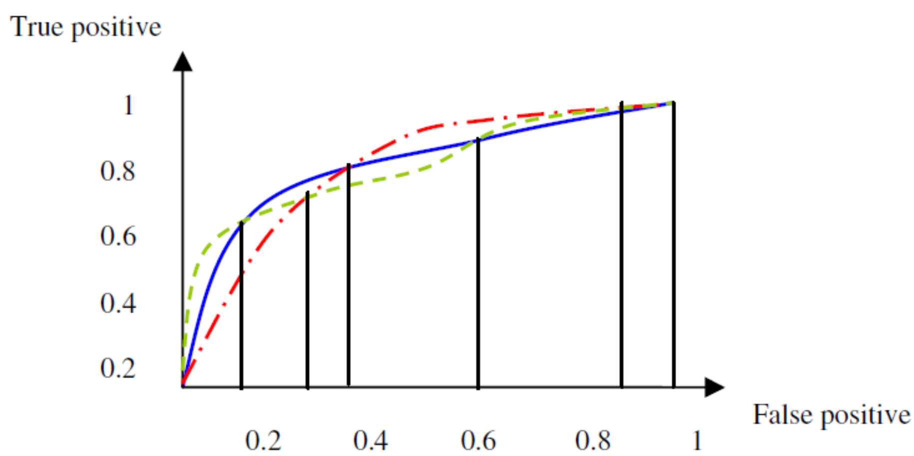


Figure 3-5 Area Under Curve (AUC)

(Biau and Scornet, 2016)

3.5 K-Nearest Neighbors

3.5.1 Introduction

This section provides information about developing classification models by K-nearest neighbors method. K-nearest neighbors are applicable to develop both regression and classification models. Nearest neighbors method works based on identifying the labels of K-nearest patterns in data space. Nearest neighbor techniques have better performance when the datasets are large with low dimensions (Kramer, 2016). This model can be used for prediction in both supervised and unsupervised learning approaches.

3.5.2 KNN Classifier

K-nearest neighbors (KNN), also known as nearest neighbor classification, works based on recognizing the nearest patterns to a target pattern x', to deliver label information of different classes in the dataset. For an unknown pattern x_j the class labels are assigned based on the majority of the K-nearest patterns in data space. Equation 3.25 defines a similarity measure in data space based on Minkowski metric (Kramer, 2016). In Minkowski metric the distance between two vectors is the norm of their different.

$$\left\|\mathrm{x}' - \mathrm{x_j}\right\|^p = \left(\sum_{i=1}^{q}\left|(x_i)' - (x_i)_j\right|^p\right)^{1/p} \qquad \text{Eq. 3.25}$$

Additionally, the adequate distance function can be measured by Hamming distance which measures the minimum number of errors that could have transformed data points. For example, in the case of binary classification with set of dependent variables $y = (1, -1)$, KNN is defined in Equation 3.26.

$$f_{KNN}(\mathrm{x}') = \begin{cases} 1 & if \sum_{i \in N_k(\mathrm{x}')} y_i \geq 0 \\ -1 & if \sum_{i \in N_k(\mathrm{x}')} y_i < 0 \end{cases} \qquad \text{Eq. 3.26}$$

where K is size of neighborhood with a set of $N_k(\mathrm{x}')$ of K-nearest patterns. The size of neighborhood describes the locality of KNN. When size of neighborhoods are small, little and scattered neighborhoods appear in regions and the model tends to overfit. In contrast, a model with higher neighborhood size ignore the patterns which are in minority. Figure 3-6 illustrates a classification model with K = 1 and K = 20 on a two-dimensional data.

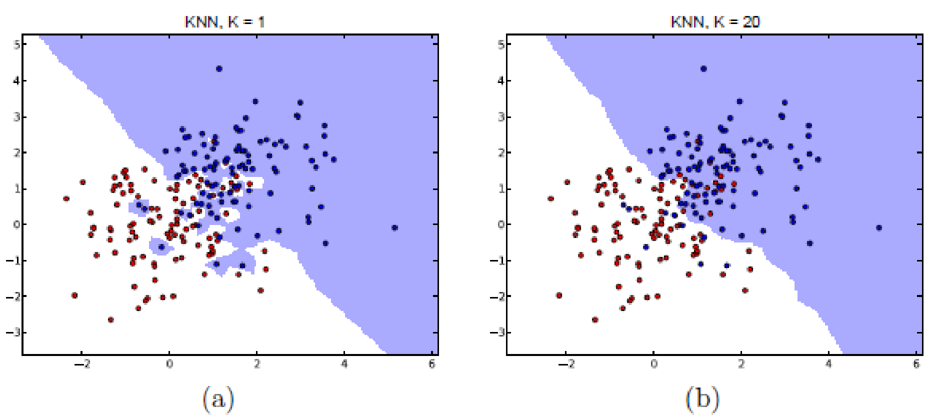


Figure 3-6 Comparison of KNN Classification ((a) K=1 and (b) K=20))

(Kramer, 2016)

For K = 1, several neighborhoods raised around the blue outliers located at the center of the red data points. For K = 20, the classifier ignored small patterns and the KNN search for the K-nearest patterns in the whole space. In larger neighborhood size, the risk of overfitting is lower, and the model yield a good approximation. Selecting an appropriate size of neighborhood is an important part of developing K-nearest Neighbors models. Various techniques such as, cross-validation can be used to select the best model and parameters in KNN models (Kramer, 2016).

3.5.3 Cross-Validation

Cross-validation is a strategy to avoid overfitting during training and testing the artificial intelligence models. In this method, the N observations $\{(\mathrm{x_i}, \mathrm{y_i})\}_{i=1}^{N}$ split up into training, validation and test set. The training set is used to learn the algorithm in the model. The validation set is utilized to evaluate the model and the test set are used to evaluate the final independent test set (Biau and Scornet, 2016).

K-fold cross-validation is one the most common and advanced strategies to avoid overfitting. In this method, the learning process is repeated *k* times with different training and validation sets. To generate K-fold cross-validation, the dataset is split up into *k* separate groups and in each step $k - 1$ sets are employed for training and the remaining validation set is used to evaluate the model. For example, to split the data into 80% training and 20% testing, 5-fold cross-validation should be used during model development.

In this method, all the observation is used to train and test of the model. All the developed models in this dissertation were trained and tested by employment of k-fold cross-validation strategy.

3.5.4 Evaluation of KNN Model

In this dissertation, the validation and performance evaluation of KNN model were performed by several techniques such as, confusion matrix, ROC curve and Area Under Curve (AUC). These methods were presented in section 3.4.6.2 through 3.4.6.4.

3.6 Chapter Summary

In this chapter the detail of logistic regression, gradient boosting tree and KNN models was comprehensively reviewed. The discussions in this chapter reinforced the suitability of statistical and artificial intelligence models to work as a classifier for predicting condition of sanitary sewer pipes. Furthermore, model selection process and various techniques for training and evaluation of the models were widely investigated. The source of sanitary sewer pipe database and different steps of data preparation will be presented in next chapter.

Chapter 4 Data Preparation and Analysis

4.1 Background Information of Sanitary Sewer Dataset

The framework of this study is based on collected data from the City of Tampa, Florida. The City of Tampa's Wastewater Department receives and treats wastewater collected from the Tampa area and surrounding suburbs. On average, more than 50 million gallons of raw sewage is treated per day to an advanced level that meets or exceeds federal regulations. In addition to treating the City's wastewater, the maintenance of gravity and force main sewer lines and over 30,000 manholes are performed regularly. Overall, sewer inventory networks included a total number of 70,172 manhole to manhole gravity and force main pipe segments approximately 1,800 miles in length. Table 4-1 shows a summary of the sewer networks proportions in the City of Tampa.

Table 4-1 Sewer Network Proportion in Tampa City

Feature	Description
Miles of sanitary system	1,800 miles
Number of gravity pipes	31,364
Number of force mains	38,808
Range of pipe diameter	2 – 96 in.
Majority pipe size	8 in.
Burial depth range	0.50 – 29.64 ft
Pipe Material	Asbestos-cement (AC) Cast iron (CAS) Corrugated metal pipe (CMP) Ductile iron pipe (DIP) Fiberglass pipe (FRP) Prestressed concrete cylinder pipe (PCCP) Plastic pipe (PE) Polypropylene pipe (PP) Polyvinyl chloride (PVC) Reinforced concrete pipes (RCP) Vitrified clay pipe (VCP)
Majority of Older Pipes	VCP, ~60% constructed prior to 1970, ~20% constructed prior to 1950
Majority of Newer Pipes	PVC
Start of CCTV inspection	~2005

The inspection of sewer pipes began in 2005 by the City of Tampa and Pipeline Assessment and Certification Program (PACP) guidelines were used to assess the condition of pipes on a scale from 1 to 5, with 1 indicating a pipe with no or very few defects and 5 representing failing conditions. In this dissertation, sanitary sewer pipes that were assessed according to PACP guideline were used to develop the prediction models due to avoiding any inconsistencies and imbalance during data preparation and building the models.

The geographic information system (ArcGIS) is the primary source of information to manage and maintain wastewater systems for the City of Tampa. A list of data layers, that are compatible with GIS software were collected to develop the deterioration models as shown below:

- Sewer inventory data for sanitary pipes
- Closed Circuit Television (CCTV) inspection data
- Soil data

Sewer inventory dataset contained a total number of 30,739 individual manhole to manhole pipe segments. A unique number was assigned to each pipe segment with the intention of making it easier to identify and track individual pipes. Total length of pipe segments in the inventory was approximately 1,250 miles. Several pipe attributes such as, installation year, material, diameter, length, depth, shape, slope, down elevation, up elevation and location of pipes were in the inventory. Furthermore, the pipes that have any repair or replacement were all specified by date and type of repair in the dataset. The newest pipe in the dataset had an installation date of 2018 and the oldest pipe was installed in 1947. This inventory involved essential information regarding condition of each pipe segment, such as, PACP condition level, total score, and quick rating structural and operational defects. Figure 4-1 illustrates location of sanitary sewer pipes in Tampa city.

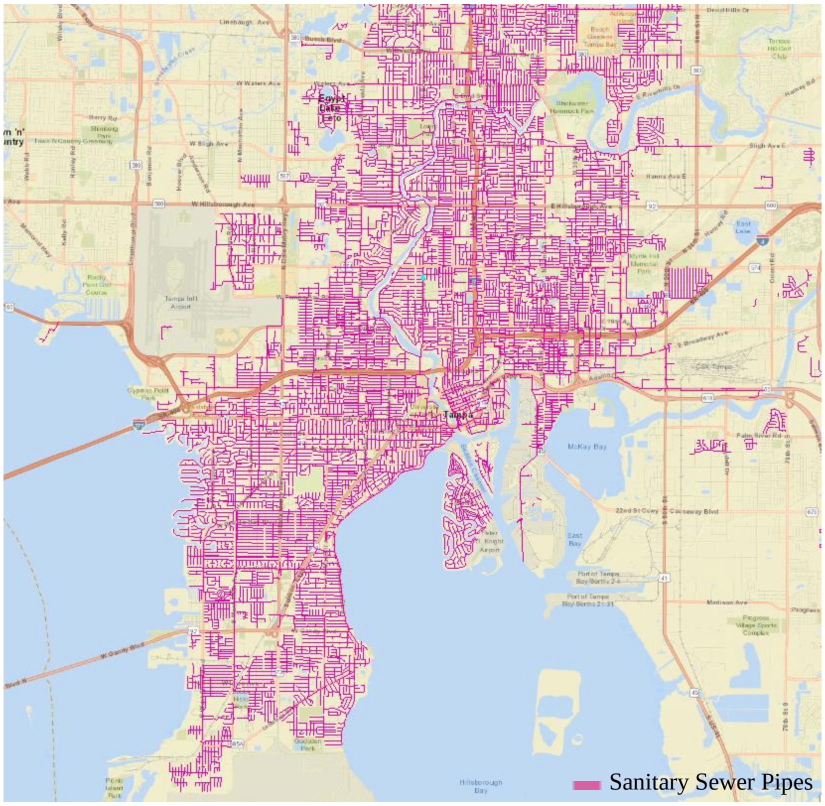


Figure 4-1 Location of Sanitary Sewer Pipes in Tampa City

CCTV inspection data involved detail of main inspection and scoring for all individual pipe segments based on PACP condition rating method. For each pipe segment some information such as, pipe rating, quick rating and pipe rating index were available for both structural and operational conditions. Additionally, the overall condition of pipes was existed in this database. The detail of calculating condition of sewer pipes based on PACP guideline was comprehensively presented in chapter two. As shown in Figure 4-2, type and total number of pipe defects were accessible in observations section. CCTV inspection database was used in this dissertation to fix any incorrect or missing information of sewer inventory dataset. For example, if condition of a pipe was missed in the inventory, the correct information was obtained from CCTV inspection data.

Scores

Calculated at: **4/3/2018 9:43:43 AM**

Grade	Structural:				O&M:				Overall:	
	Segment Grade	Pipe Rating	Quick Rating	Pipe Rating Index	Segment Grade	Pipe Rating	Quick Rating	Pipe Rating Index	Pipe Rating	Pipe Rating Index
1	0				0					
2	2				0					
3	9	15	4133	3.0	0	0	0000	0.0	15	3.0
4	4				0					
5	0				0					

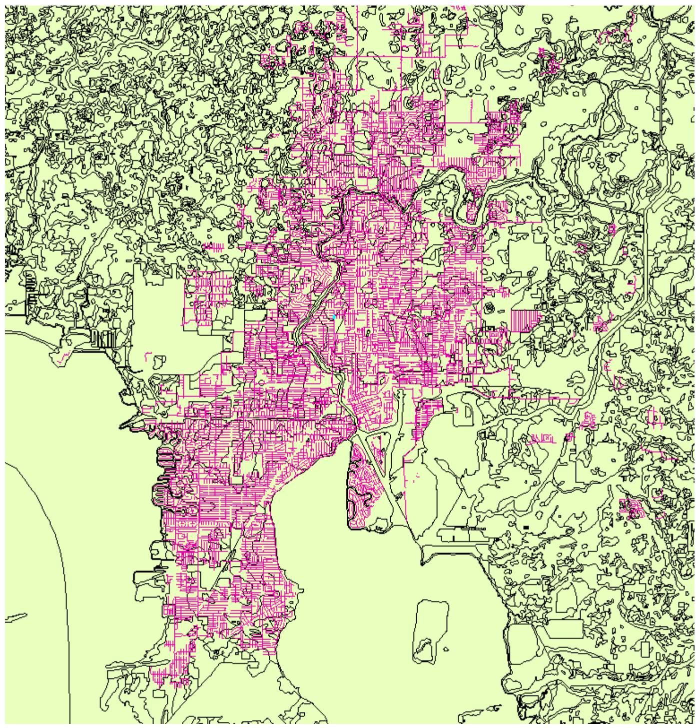
Observations

Distance	Dir.	Length	From/To	Code	Modifier	Rating
0.0 ft.	U		/	AMH		
0.0 ft.	U		/	MWL		
18.1 ft.	U		10 /	TB		
21.1 ft.	U		6 /	FL		3
21.7 ft.	U		12 /	FL		3
22.4 ft.	U		/	JOM		3
26.6 ft.	U		3 /	TB		
39.2 ft.	U		12 /	TB		
69.6 ft.	U		9 /	TFC		
72.3 ft.	U		3 /	TB		
74.7 ft.	U		1 /	TFC		
113.7 ft.	U		9 /	TB		
117.2 ft.	U		3 /	TB		
117.4 ft.	U		1 / 5	FM		4
122.2 ft.	U		9 /	TFC		
137.1 ft.	U		3 /	TB		
151.3 ft.	U		1 /	TFC		
193.6 ft.	U		12 /	CL		2
220.1 ft.	U		1 /	TFC		
233.6 ft.	U		12 /	TB		
242.2 ft.	U		/	AMH		

Figure 4-2 CCTV Inspection Database

Soil database was another important part of this study which was collected from web soil survey (2018). Some important environmental variables such as, soil type, soil sulfate, soil pH, soil hydraulic group and soil corrosivity were collected based on depth of each individual pipe segments. Several appropriate information was available in this dataset and the variables were selected based on the possibility of their impact on condition states of sanitary sewer pipes.

Soil data was a polygon type GIS and spatial join feature was used to combine pipe and soil datasets. Spatial join is a GIS operation to combine data from one attribute table to another from a spatial perspective. In other word, spatial join is used to add data from one feature class to another class. Figure 4-3 illustrates the combination of sanitary sewer pipes location and soil datasets.



(Sanitary Sewer Pipes; Soil Areas)

Figure 4-3 Combination of Sewer and Soil Dataset

To generate this operation, latitude and longitude of each pipe segment were utilized to provide combination of soil area and sanitary sewer pipe datasets.

4.2 Dataset Preparation

Combination of sanitary sewer pipes and soil datasets provided a huge amount of geographical information. Before developing statistical or artificial intelligence models, data needs to be prepared. Data preparation is combination of strategies to work with dataset for feeding pure data as input to model algorithms due to achieving higher accuracy. Data preparation is not a completely automated process and several techniques should be applied to prepare the dataset (Pyle, 2007). Typically, combination of rules and techniques are used to clean the dataset based on type of data and output of the model. Prior to start statistical analysis of sewer dataset, data were filtered and several evaluations have been done to find inappropriate and missing information. A unique “Facility ID” was assigned to each pipe segment with the intention of making it easier to identify and track individual pipes. Additionally, this facility IDs were used to identify any duplicate record in the pipe dataset.

As a first step, missing information was identified and analyzed based on the variables included in the sewer dataset. For example, approximately 4,000 pipe segments were available in the dataset without any information regarding installation year. Pipes with missing information on pipe age, depth, material, slope, length and condition scales were excluded from the dataset.

Secondly, Inspection results of pipes which previously underwent a lining, repair or replacement were excluded from the study. Approximately, 1,972 pipe segments were observed with historical repair and rehabilitation information. Typically, pipe segments with repair or lining history have higher age, while their condition states are excellent. Therefore, considering this groups of pipes in the dataset highly affect the result of the condition prediction models.

Thirdly, pipe segments with negative age, depth and length values were removed from the dataset. Some inspection results were outdated, and the pipe attributes included several negative and constant values. After communication with engineers in City of Tampa, any infrequent datapoint was excluded from the dataset. Fourthly, pipe material such as cast iron, concrete, ductile iron, reinforced concrete and plastic pipes which had low population in the dataset were removed. The total number of all these pipes were approximately 3,000.

As a final step, different techniques, such as, boxplot and Mahalanobis Distance were used to remove the outliers from the dataset. Observed datasets often contain outliers which have numerically distant from the rest of the data. Outliers are typically larger or smaller than observed values in the dataset. Boxplot is a well-known simple graphical tool to display the variation of continuous data. The median, lower quartile, upper quartile, lower extreme, and upper extreme are the thresholds identified by boxplot (Seo, 2006). Mahalanobis Distance is a measure of the distance between a point and its distribution and it is widely used for regression and classification problems (Hastie et al., 2017). It was observed that after removing outliers the correlation between dependent and independent variables was improved. The final dataset contains 19,766 individual pipe segments with different physical and environmental variables as shown in table 4-2.

Table 4-2 Variables Included in Sewer Pipe Dataset

Category	Variables	Description
Physical	Age	Time difference between the installation date of the pipe segment and date of inspection in years
	Material	Type of sewer pipes material
	Diameter	Diameter of the sewer pipe segment in inches
	Depth	Depth of overburden above the sewer pipe segment in feet
	Slope	Vertical displacement of the pipe section per horizontal displacement in percentage
	Length	Length of the sewer pipe segment between two manholes in feet
Environmental	Soil Type	Type of soil surrounding the pipe
	Soil Sulfate	The weight percentage of hydrated calcium sulfate in the soil
	Soil pH	A numerical expression of the relative acidity or alkalinity of a soil sample
	Water Table	The average depth of water table during the year in inches
	Soil Hydraulic Group	Soil hydraulic groups indicate soil runoff potential and the rate of water transmission through the soil layers
	Soil Corrosivity	Corrosivity of soil based on soil texture, pH, and amounts of magnesium and sodium sulfate or sodium chloride in the saturated soil paste
Operational	Pipe Flow	The amount of sewage transmission in gallon per minutes

Some features, such as, pipe age, size, depth, water table, soil pH and soil sulfate are numeric variables and some features, such as, pipe material, soil type, soil hydraulic groups and soil corrosivity are categorical variables to develop sewer deterioration model. Table 4-3 illustrates more information about variables used in this study.

Table 4-3 Type of variables in Sewer Pipe Dataset

<table>
<tr><th>Variables</th><th>Variable Type</th></tr>
<tr><td>Age</td><td rowspan="9">Continuous quantitative</td></tr>
<tr><td>Diameter</td></tr>
<tr><td>Depth</td></tr>
<tr><td>Slope</td></tr>
<tr><td>Length</td></tr>
<tr><td>Soil Sulfate</td></tr>
<tr><td>Soil pH</td></tr>
<tr><td>Water Table</td></tr>
<tr><td>Pipe Flow</td></tr>
<tr><td>Material</td><td>Nominal categorical
• PVC
• VCP</td></tr>
<tr><td>Soil Type</td><td>Nominal categorical
• Clayey soil
• Fine sand
• Silty gravel and sand
• Silty soil</td></tr>
<tr><td>Soil Hydraulic Group</td><td>Ordinal categorical
• Group A
• Group B
• Group C
• Group D</td></tr>
<tr><td>Soil Corrosivity</td><td>Ordinal categorical
• Low
• Moderate
• High</td></tr>
</table>

4.3 Description of Variables in Final Dataset

4.3.1 Pipe Age

Pipe age is time difference between the installation date of the pipe segment and date of inspection in year. The age of pipes in sanitary sewer dataset ranges from 1 to 69 years. As shown in Figure 4-4, most of the pipes (13.07%) are between 50 to 55 years old and only 0.57% of them are below 5 years.

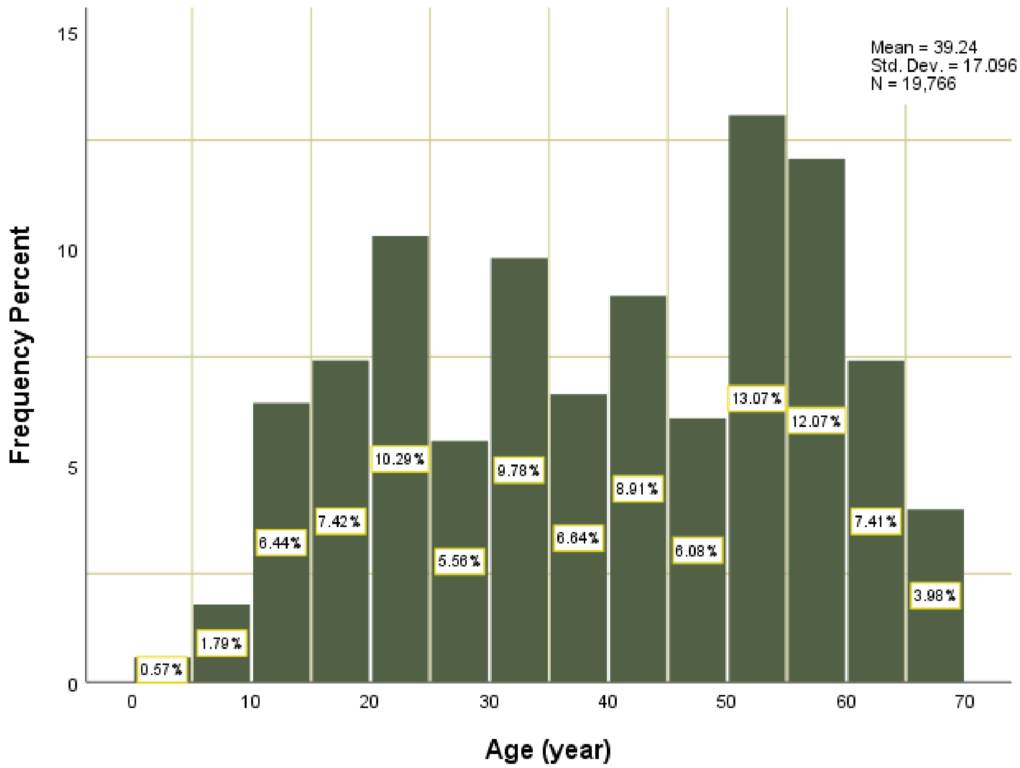


Figure 4-4 Frequency of Pipe Age

4.3.2 Pipe Material

The sanitary sewer pipe dataset involved different type of pipe material such as asbestos-cement (AC), cast iron (CAS), ductile iron pipe (DIP), prestressed concrete cylinder pipe (PCCP), polyvinyl chloride (PVC) and vitrified clay pipe (VCP). In this dissertation only polyvinyl chloride and vitrified clay pipes were used to develop the prediction models due to their frequency in sewer dataset. Rest of pipe material was excluded to avoid any misclassification or error during model development. As shown in Figure 4-5, vitrified clay pipes are dominant material type with 65.21% frequency rather than 34.79% polyvinyl chloride pipes.

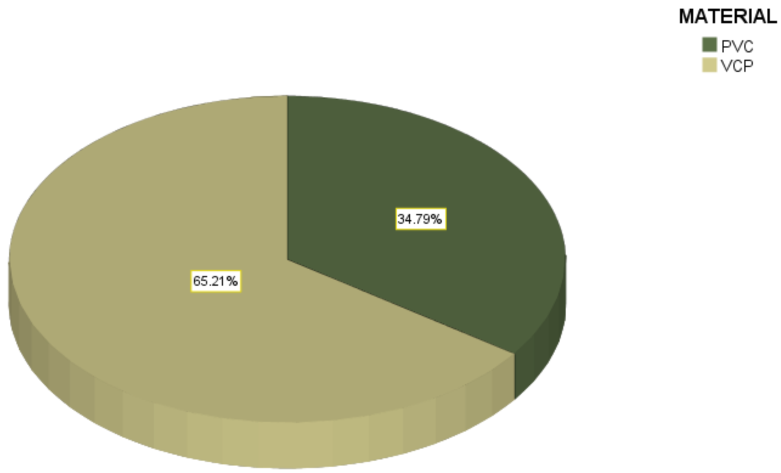


Figure 4-5 Frequency of Pipe Material

4.3.3 Pipe Diameter

Pipe diameter indicates size of sanitary sewer pipes in inch. As shown in Figure 4-6 majority of the pipes had a diameter less than 15 inches. Pipes with diameter 5 to 10 inches form 86% of the pipes in the dataset.

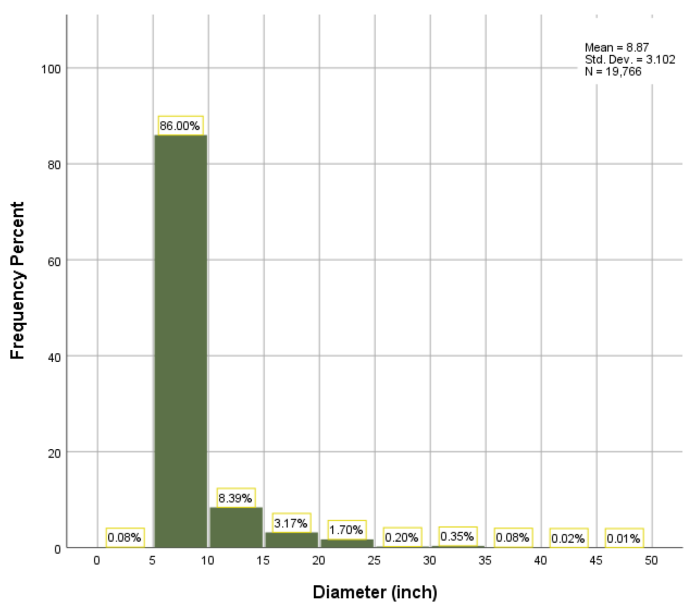


Figure 4-6 Frequency of Pipe Diameter

4.3.4 Pipe Flow

Pipe flow is an operational factor that indicates the amount of sewage transmission in gallon per minutes. Pipe slope, size, viscosity and population in urban areas govern the behavior of sanitary sewer pipes flow. Dataset used in dissertation had different values of pipe flow for each pipe segment.

4.3.5 Pipe Depth

Depth of sanitary sewer pipe is the depth of backfill over the top of the pipe in feet. According to the Figure 4-7 most of the pipes were buried within the depth of less than 10 feet and 36.73% of the pipes were covered by 4 to 6 feet backfill material. Just few pipe segments buried under a depth of less than 2 feet and more than 20 feet. The average age of pipes buried below 2 feet is 41 years and it indicates that more restrictions have been placed to install newer pipes.

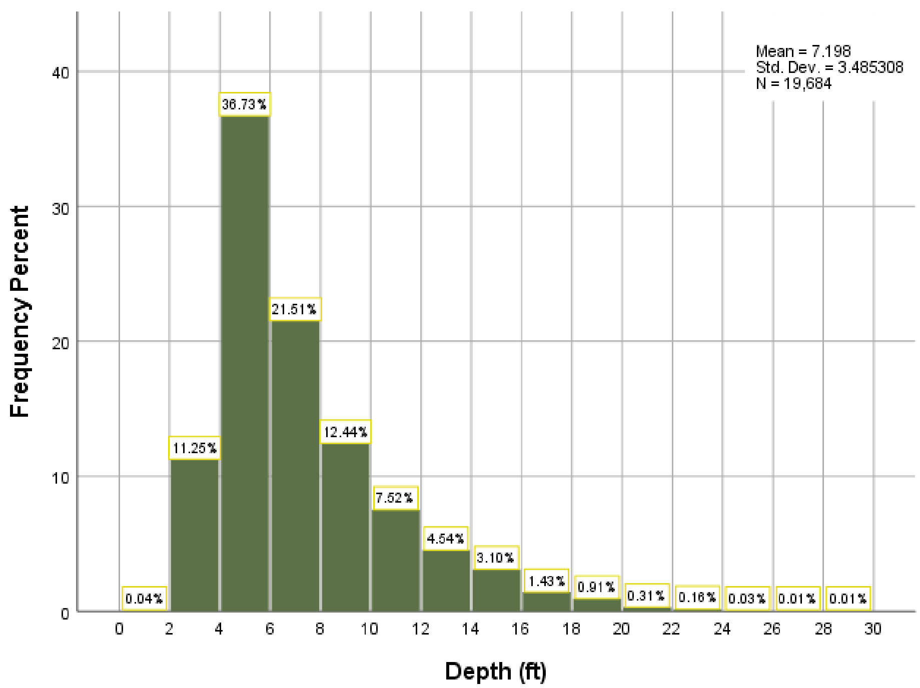


Figure 4-7 Frequency of Pipe Depth

4.3.6 Pipe Slope

Pipe slope is vertical displacement of the pipe section per horizontal displacement in percentage. As Tampa city is placed in a flat geographical location, most of the sewer pipes had very low slope. As illustrated in Figure 4-8 the minimum slope was -1.76% and the maximum was 21%.

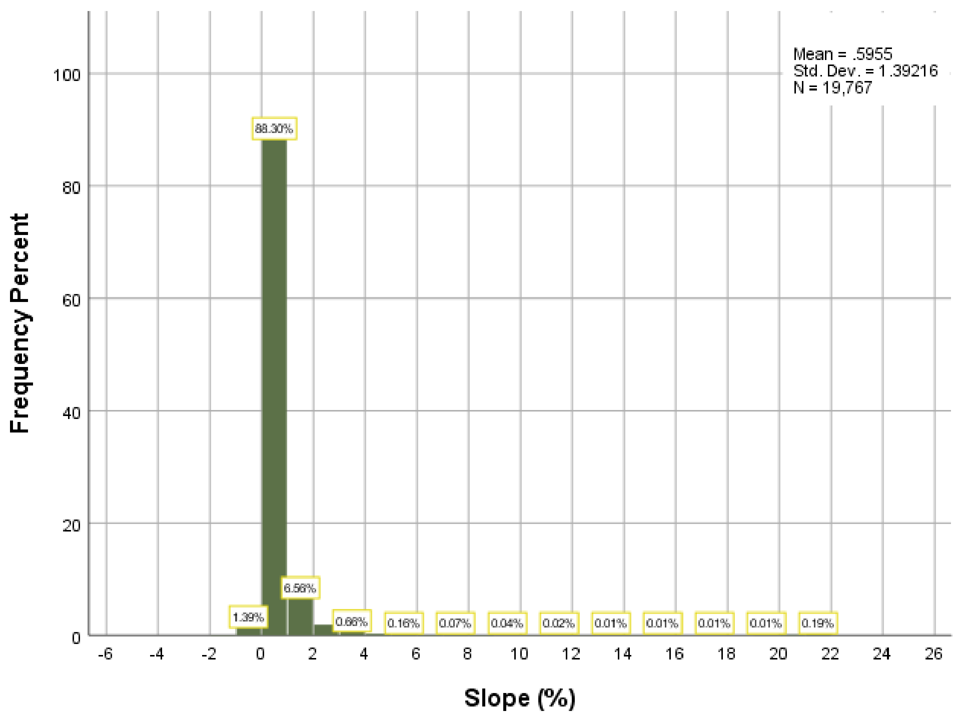


Figure 4-8 Frequency of Pipe Slope

4.3.7 Pipe Length

Pipe length is manhole to manhole length of sewer pipe segments in feet. most of sewer pipe inventories storage manhole to manhole length of sewer pipes due to difficulty of data collection. The length of pipes in sanitary sewer dataset ranges from 3 to 680 feet. As shown in Figure 4-9 the highest frequency percentage belongs to the pipes with length of 250 to 300 feet. Only few percentages of the sewer pipes had the length of more than 400 ft.

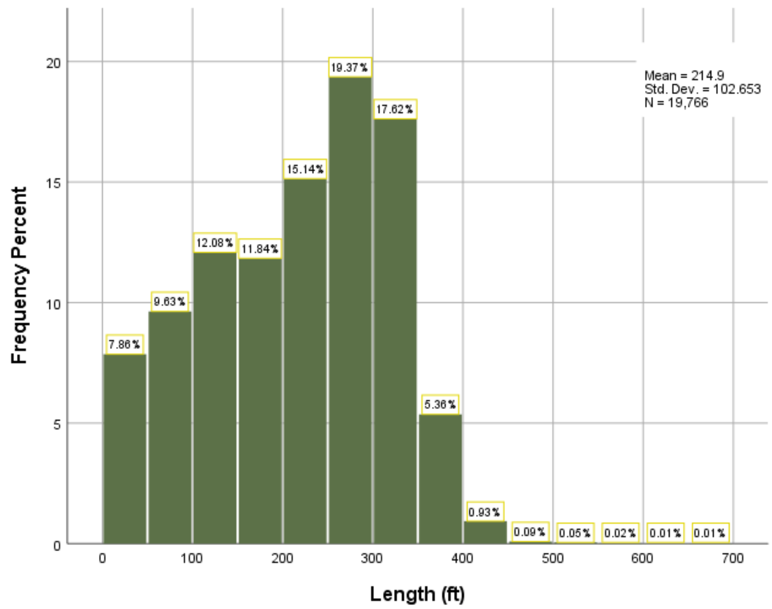


Figure 4-9 Frequency of Pipe Length

4.3.8 Soil Type

Soil type indicates the type of backfill material surrounding the sewer pipes. Type of soil is one of the important factors can affect the ground loos and stability of the sewer pipes. As shown in Figure 4-10 silty gravel and sand is the most common soil type surrounding the sewer pipes with 71.72%.

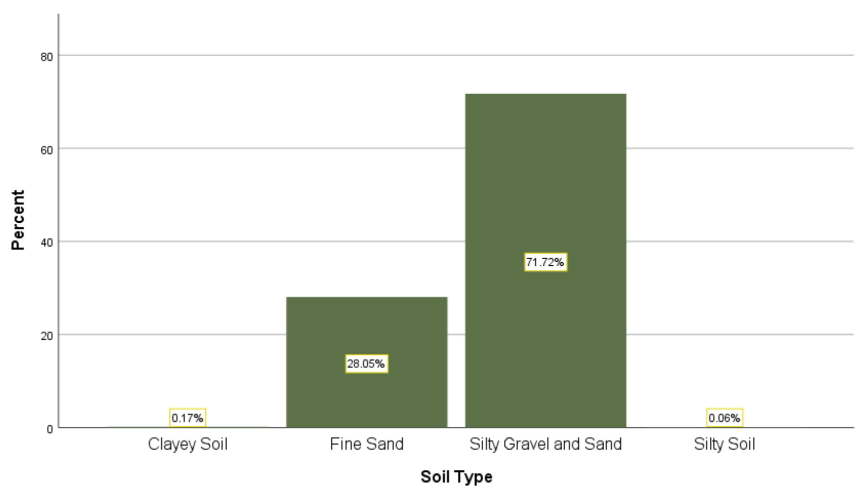


Figure 4-10 Distribution of Soil Type

4.3.9 Soil Sulfate

Soil sulfate in sewer dataset is the weight percentage of hydrated calcium sulfate in the soil. The amount of gypsum or hydrous calcium sulfate ($CaSO_4 \cdot 2H_2O$) in the soil can affect the condition of pipes gradually over time. Sulfur components inside the soil can react with oxygen and release sulfuric acid which is harmful for environment and infrastructures. The combination of sulfates and chlorides is argued to be the leading cause of corrosion of steel reinforcements inside the concrete pipes (Bhattarai, 2013). High level of soil sulfate is the primary cause of corrosion in buried stainless steels (Sjogren et al., 2011). As illustrated in Figure 4-11, approximately 60% of the pipes were buried in soil areas within 0.02 to 0.05% sulfate content.

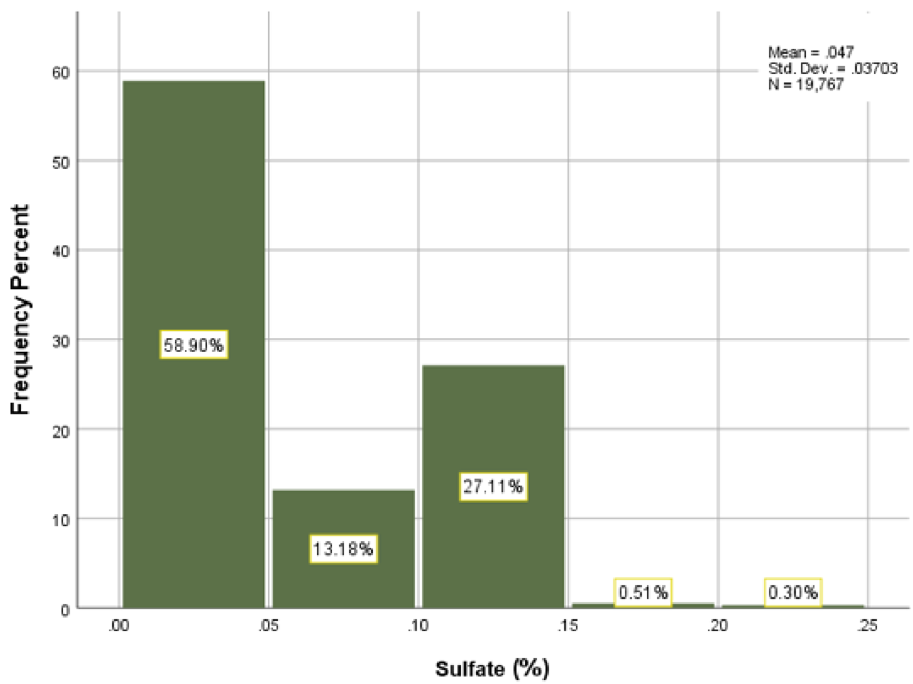


Figure 4-11 Frequency of Soil Sulfate

4.3.10 Soil pH

The soil pH is considered as the most important factor affecting underground corrosion. Typically, most of the studies in the field of underground corrosion indicated that the pH of the soil increases the corrosion rate of buried pipes (Wasim et al., 2018). In sanitary sewer dataset used in this study, pH is a numerical expression of the relative acidity or alkalinity of a soil sample. Buried metallic structures are vulnerable to corrosion at any pH value (Oguzie et al., 2004). The soil sample collected from Tampa area showing the pH values in the range of 4 to 8.2. Range of pH can be described as alkaline (pH>7), natural (pH=7) and acidic (pH<7). Figure 4-12 illustrates the distribution of pH in sanitary sewer dataset. The histogram shows that 73% of soil areas have pH between 5 to 6 which indicates the risk of acidity and high rate corrosion.

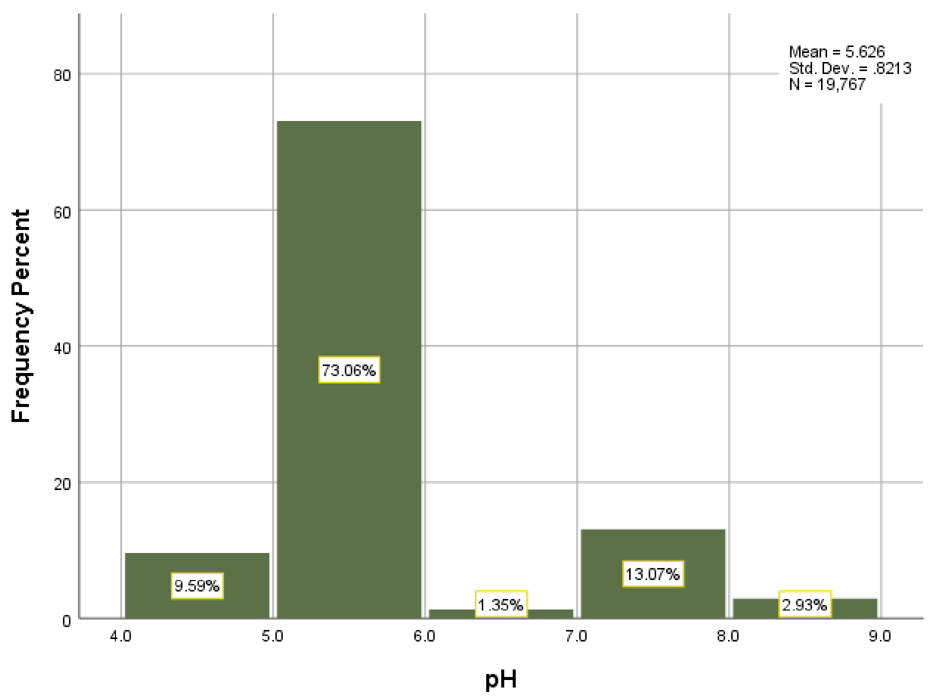


Figure 4-12 Frequency of Soil pH

4.3.11 Water Table

Water table or groundwater level indicates the average depth of water during the year in inches. Tampa city is located to the west coast of Florida on Tampa Bay, near the Gulf of Mexico and the highest point in the city is only 48 feet above sea level. Therefore, the groundwater level is always high during all seasons. The availability of groundwater at or above sewer pipelines may cause water flowing through the pipe, increasing the structural defects, formation of void and loss of sewer support. As shown in Figure 4-13 the average water table between 20 to 40 inches has the highest frequency in Tampa city with approximately 47%.

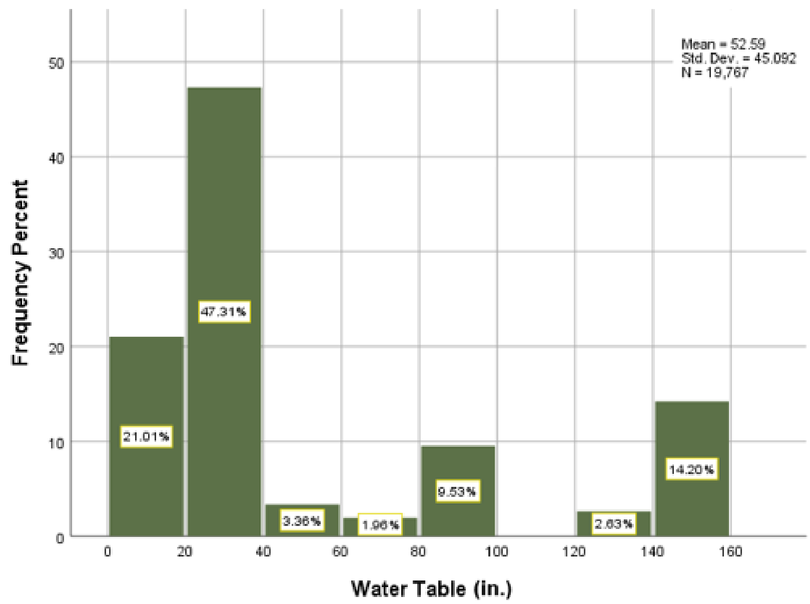


Figure 4-13 Frequency of Water Table

4.3.12 Soil Hydraulic Group

Soil hydraulic group indicates soil runoff potential and the rate of water transmission through the soil layers. According to National Engineering Handbook (2007), hydrologic soil groups can be classified in four general groups of A, B, C and D. This classification is based on several parameters such as water table, transmission rate of water, texture, structure and degree of swelling when soil is saturated. Soil hydraulic group was used as an input variable in this dissertation based on the level of groundwater in Tampa city (presented in previous section). The four hydrologic soil groups are described as below:

Group A: Soils in this group have low runoff potential and the water is freely transmitted through the soil layers. Typically, group A soils have less than 10 percent clay and more than 90 percent sand or gravel. Loamy sand, sandy loam, loam or silt loam textures may be placed in this group.

Group B: Soils in this group have moderately low runoff potential and the water transmission through the soil layers is unimpeded. Typically, group B soils have between 10 to 20 percent clay and 50 to 90 percent sand. Some soils having loam, silt loam, silt, or sandy clay loam textures may be placed in this group.

Group C: Soils in this group have moderately high runoff potential and the water movement through the soil layers is somewhat restricted. Typically, group C soils have between 20 to 40 percent clay and less than 50 percent sand. Some soils having clay, silty clay, or sandy clay textures may be placed in this group.

Group D: Soils in this group have high runoff potential and the water movement through the soil layers is restricted or very restricted. Typically, group D soils have greater than 40 percent clay and less than 50 percent sand and have clayey textures.

Figure 4-14 illustrates the distribution of soil hydraulic groups in sanitary sewer pipe dataset. Hydraulic group A is the most common type of soil hydraulic in dataset with approximately 79%.

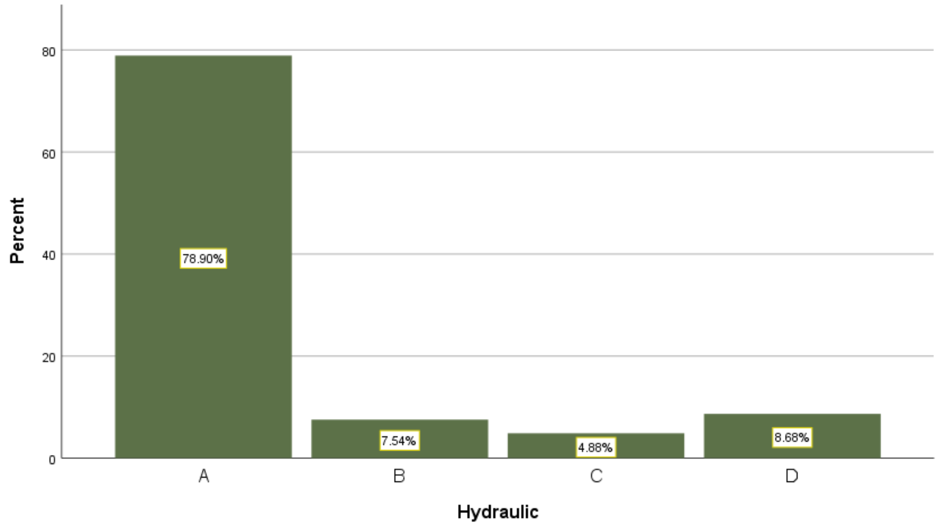


Figure 4-14 Frequency of Soil Hydraulic

4.3.13 Soil Corrosivity

This factor indicates the corrosivity level of soil based on soil texture, pH, and amounts of magnesium and sodium sulfate or sodium chloride in the saturated soil. The rate of corrosion is highly influenced by the characteristic of the pipe material and surrounding soil around the pipe. In sewer dataset, the level of corrosivity is classified into three groups of low, moderate and high as shown in Figure 4-15.

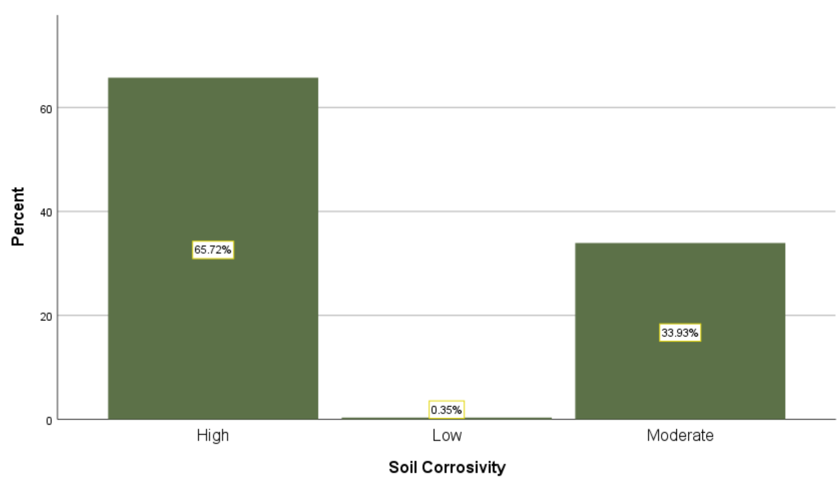


Figure 4-15 Frequency of Soil Corrosivity

4.3.14 Pipe Condition Rating

The condition states of sewer pipes are the output or dependent variable used in this study. As explained before, City of Tampa used Pipeline Assessment and Certification Program (PACP) guidelines to assess the condition of pipes on a scale from 1 to 5, with 1 indicating a pipe with no or very few defects and 5 representing failing conditions. Therefore, 13 different physical and environmental independent variables were used in this study to predict condition rating of sewer pipes. Table 4-4 and Figure 4-16 illustrates the descriptive analysis and distribution of pipe conditions in dataset.

Table 4-4 Descriptive Statistics of Sewer Pipes Condition

Pipe Condition	Frequency	Percent
1	10,338	52.3
2	2,957	15
3	978	4.9
4	1,589	8
5	3,904	19.8
Total	19,766	100

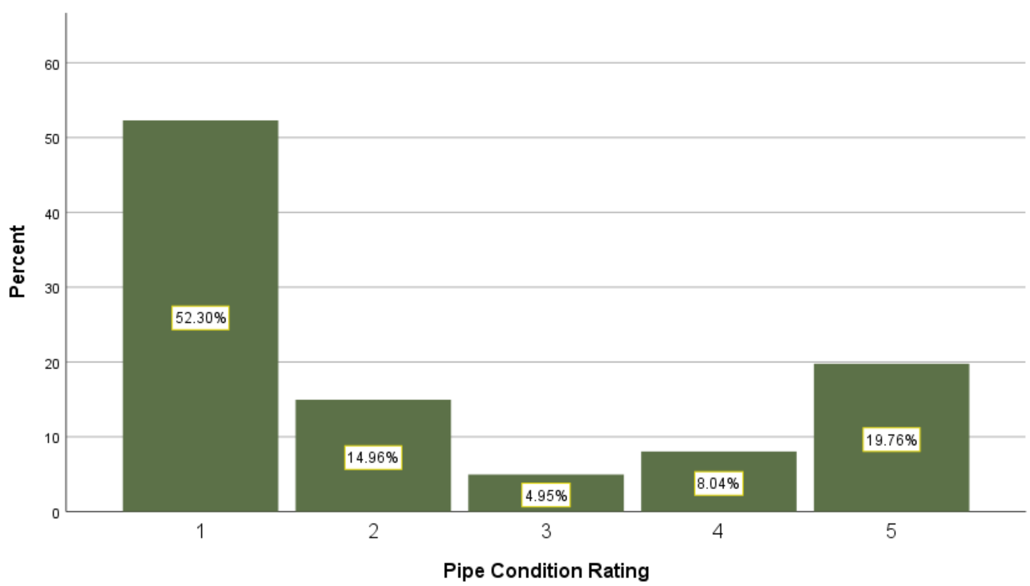


Figure 4-16 Distribution of Pipe Conditions

The highest percentage of condition rating observed in dataset corresponds to condition rating 1 (52.3%), which is followed by condition rating of 5 (19.8%). Condition ratings 2, 3 and 4 involved 27.9% of pipes in the dataset. The lowest percentage of condition state corresponds to condition rating 3 with only 4.9 percent. Figure 4-17 is the area graph which illustrates the percentage of each condition rating for age groups (bin size = 5 years).

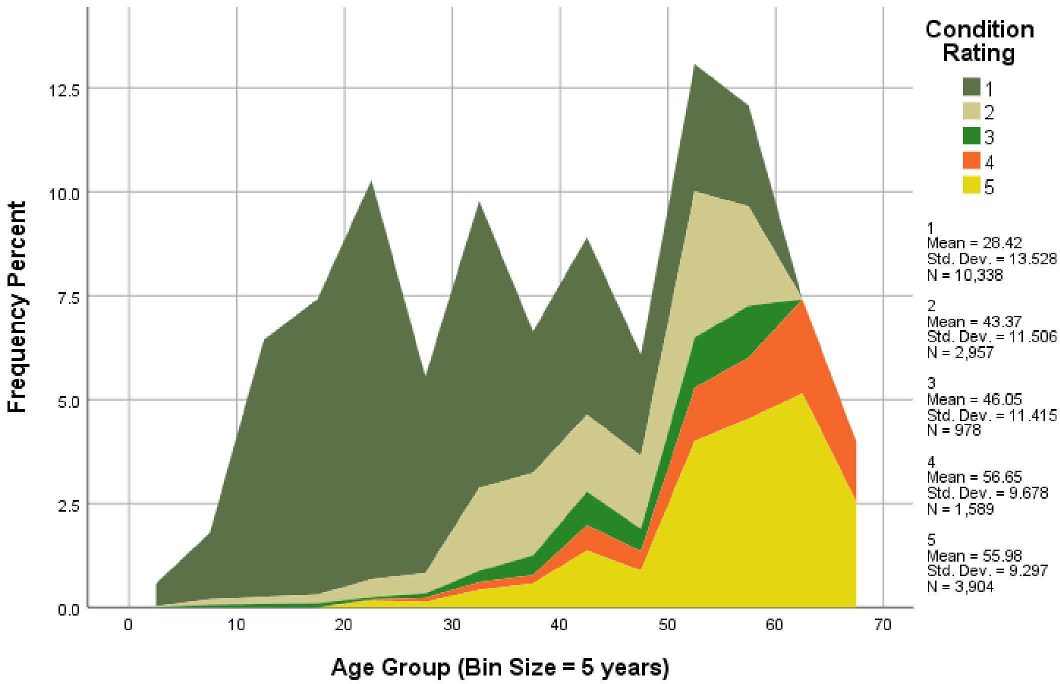


Figure 4-17 Percentage of Pipes in each Condition Rating

It is obvious that condition rating 3 and 4 nave the lowest frequency percentage in the dataset. Additionally, some sanitary sewer pipes reached condition rating 5 before age 20 and in contrast some pipes are in condition rating 1 after passing 50 years old. In this dissertation PVC and VCP pipes are used in same dataset to evaluate the effect of pipe material on condition of sanitary sewer pipes. Thus, the area diagram shows all the pipe material in sewer dataset.

4.4 Descriptive Statistics

Descriptive statistics describes a summary of quantitative analysis for the numerical variables in the dataset. The objective of development descriptive statistics in this study is to display a simple summary about the data sample and characteristics of variables in the sewer dataset. Table 4-5 presents the descriptive statistics of numerical variables in this study.

Table 4-5 Descriptive Statistics of Numerical Variables

Variables	Minimum	Maximum	Mean	Standard Deviation
Age (year)	1	69	39.24	17.1
Diameter (in.)	2	48	8.87	3.1
Flow (gallons/min)	0	28,100	656	1499
Depth (feet)	0.59	28.63	7.2	3.5
Slope (%)	-1.76	21	0.6	1.4
Length (feet)	3	680	214.92	102.65
Sulfate (%)	0.02	0.24	0.047	0.04
pH	4.0	8.2	5.63	0.82
Water Table (in.)	8	145	52.6	45.1

4.5 Correlation Analysis

Correlation analysis is a statistical method to determine the degree of relationship between two different variables. This relationship can vary from strong to weak and sometimes there is no relationship between two variables. The strong relationship means the value of one variable can be predicted based on the value of the other variable. Contrary, the variables cannot be predicted well when their relationship is weak. The correlation coefficient between variables can be positive or negative and it can only range from -1.00 to +1.00 (Lewis, 2016).

Correlation coefficient is presented by the value “r”. When the correlation is close to one (-1 or +1), there is a strong relationship between two variables. For example, assume diameter and depth of sewer pipes as two variables. Overall, the pipes with larger diameter are installed in greater depth. Therefore, there can be a positive correlation between depth and size of the pipes. A coefficient correlation close to zero means, there is no relationship between variables in the model. For example, there is no relationship between length of the pipe and type of soil in a database and their correlation is close to zero. Incorporating highly correlated independent variables in a model may cause multicollinearity problem which affects the outcomes of the model (Salman, 2010). In general, development of model with highly correlated independent variables are not recommended. The three common types of correlation analysis are (Lewis, 2016):

- Pearson: A measure of the strength of a relationship between two continuous variables.
- Spearman: A measure of the similarity between two ordinal rankings of a single set of data.
- Point-Biserial: A measure of the strength of a relationship between one continuous variable and one dichotomous variable (a two-level-only variable like gender).

Pearson correlation is the most commonly used correlation analysis. Pearson correlation assumes that the distribution between two variables are normal and only linear relationship between two variables can be described by this method. Development of condition prediction models in this study is not based on linear relationship between variables. Additionally, as described in section 4.3, most of the variables were not normally distributed in the model. Therefore, spearman's rank correlation was to examine the correlation between the variables and to avoid any multicollinearity problem. Spearman's rank correlation can be used to describe the association between nonlinearly related variables (Meyers et al. 2017). This method does not assume any assumption regarding the distribution of the variables in the model. Equation 4-1 is used to calculate the Spearman rank correlation coefficient.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \qquad \text{Eq. 4.1}$$

where ρ is Spearman rank correlation coefficient, d_i is the difference between the ranks of corresponding values X_i and Y_i and n is number of values in each data set. Table 4-6 presents the Spearman rank correlation analysis of sanitary sewer dataset.

Table 4-6 Spearman Rank Correlation Analysis

Variables	Age	Diameter	Flow	Depth	Slope	Length	Sulfate	pH	Water Table
Age	1.000	-.016*	-.011*	-.294**	.023**	.347**	-.298**	.215**	.085**
Diameter		1.000	.529**	.404**	-.306**	.066**	-.048**	.040**	-0.007*
Flow			1.000	.167**	.397**	-0.004*	.071**	-.056**	-.030**
Depth				1.000	-.212**	-.113**	.163**	-.097**	-.096**
Slope					1.000	-.027**	.102**	-.090**	-0.004*
Length						1.000	-.088**	.033**	.021**
Sulfate							1.000	-.375**	-.262**
pH								1.000	-.131**
Water Table									1.000

*Correlation is significant at the 0.05 level
**Correlation is significant at the 0.01 level

According to the result of Spearman rank analysis, all the variables were significantly correlated at the level of 0.01 and 0.05. The highest correlation coefficient is between diameter and flow (+0.529) which indicates that the larger diameter provides more flow. Additionally, there is no strong correlation between independent variables which means none of them needs to be removed from the model to avoid overfitting and multicollinearity.

4.6 Chapter Summary

In this chapter the source of sanitary sewer dataset was comprehensively reviewed. Additionally, the detail of variables included in the model and data preparation techniques were discussed. The raw database was transformed into a standardized format ready for development of the models. The available parameters for the model development were identified and their relevance examined through statistical analysis. The detail of developing logistic regression, gradient boosting tree and KNN models will be presented in next chapter.

Chapter 5 Development of Prediction Models

5.1 Introduction

The previous chapter described the sanitary sewer dataset, data preparation processes and statistical analysis of acquired data. This chapter deals with the detailed account of developing multinomial and binary logistic regressions, gradient boosting tree and k-nearest neighbors models. The model development in this chapter includes the detail of training and testing of the models. Moreover, the influence of independent variables on deterioration of sanitary sewer pipes is presented comprehensively.

Prior development of the models, five-fold cross validation method was used to randomly divide the dataset into two groups for train and validation purposes. This technique was used to reduce the risk of uncertainty and overfitting during generating the models. 80% of the data from sanitary sewer inventory was used to develop training dataset and the remaining 20% was utilized for validation of the models. Total number of records was approximately 15,800 pipe segments for training and 3,966 pipes for validation (based on the k-fold cross validation rules, the numbers were varying in different models). The cross validation was conducted manually for multinomial and binary logistic regressions, but python libraries were used during the development of gradient boosting trees and KNN models.

IBM SPSS Statistics packages (SPSS 25) was the primary software to develop the statistical models (multinomial and logistic regressions) and python was selected to perform gradient boosting trees and KNN models. Numerical variables, namely age, diameter, flow, depth, slope, length, sulfate, pH and water table; and categorical variables, namely material, soil type, soil hydraulic group, and soil corrosivity were entered as independent variables to develop the condition prediction models. Table 5-1 demonstrates a portion of sanitary sewer dataset used to train the models.

Table 5-1 Sample Portion of Sanitary Sewer Dataset

Facility ID	Age	Material	Diameter	Flow	Depth	Slope	Length	Soil	Sulfate	pH	Water Table	Hydraulic	Corrosivity	Condition
276914	63	VCP	8	366	3.29	0.45	679.99	Fine Sand	0.02	5.3	145	A	High	4
271516	50	VCP	24	2523	8.08	0.10	657.95	Silty Gravel and Sand	0.05	5.5	84	A	High	2
290774	19	PVC	10	524	5.88	0.28	349.40	Clayey Soil	0.10	5.1	31	A	High	1
2895105	29	PVC	8	347	5.6	0.41	349.39	Fine Sand	0.02	5.6	31	B	High	1
295715	69	VCP	8	343	4.2	0.40	349.30	Fine Sand	0.02	5.3	145	A	High	4
294314	43	VCP	8	509	4.24	0.88	349.24	Fine Sand	0.02	5.3	145	A	High	2
273546	38	VCP	18	2107	4.65	0.20	349.20	Clayey Soil	0.02	5.3	145	A	High	1
274321	50	VCP	30	5091	8.32	0.08	471.46	Fine Sand	0.10	8.2	59	A	High	1
275920	34	VCP	8	296	11.42	0.30	349.00	Fine Sand	0.02	4.8	31	B	High	1
1984206	13	PVC	8	351	4.75	0.42	90.97	Clayey Soil	0.02	5.3	145	A	High	1
290076	29	VCP	8	344	4.92	0.40	349.00	Fine Sand	0.10	4.8	31	B	High	5
288675	46	VCP	8	314	5.58	0.34	349.00	Clayey Soil	0.05	5.5	15	A	Moderate	2
293769	69	VCP	8	344	10.22	0.40	349.00	Fine Sand	0.02	4.8	31	B	High	5
272205	47	VCP	8	349	9.25	0.41	343.17	Silty Gravel and Sand	0.05	5.5	84	A	High	5
292932	17	PVC	8	342	8.68	0.40	343.11	Fine Sand	0.05	5.5	15	A	Moderate	1
286752	41	VCP	8	360	6.58	0.44	343.07	Silty Gravel and Sand	0.02	5.5	31	A	Moderate	4
296805	41	VCP	8	464	5.83	0.73	343.07	Silty Gravel and Sand	0.02	5.5	31	A	Moderate	2
277405	49	VCP	8	194	4.57	0.13	343.00	Fine Sand	0.02	5.3	145	A	High	3
278194	54	VCP	8	638	4.33	1.38	343.00	Silty Gravel and Sand	0.10	5.3	8	A	High	1
284399	64	VCP	8	309	6.1	0.32	343.00	Fine Sand	0.02	4.8	31	B	High	5
297087	24	PVC	8	567	4.99	1.09	343.00	Fine Sand	0.02	5.3	145	A	High	1

5.2 Multinomial Logistic Regression

5.2.1 Description of the Model

As described in chapter three, the multinomial logistic regression is used where the dependent variable is nominal with more than two levels. Since there are five possible pipe condition levels, multinomial logistic regression was used as a first model to predict all the probable conditions of sanitary sewer pipes. Therefore, four different multinomial logistic regression equations were developed according to the result of the model. Equation 5-1 presents the general form of the multinomial logistic regression when all the independent variables are significant in the model.

$$
\begin{aligned}
ln\left(\frac{P(C=i)}{P(C=5)}\right) = {} \\
& \alpha_i + \beta_{i1} \times Age + \beta_{i2} \times Diameter + \beta_{i3} \times Flow + \beta_{i4} \times Depth + \beta_{i5} \times Slope \\
& + \beta_{i6} \times Length + \beta_{i7} \times Sulfate + \beta_{i8} \times pH + \beta_{i9} \times Water\ Table \\
& + \beta_{i10} \times D_{PVC} + \beta_{i11} \times D_{VCP} + \beta_{i12} \times D_{Soil=FS} + \beta_{i13} \times D_{Soil=SG} \\
& + \beta_{i14} \times D_{Soil=CS} + \beta_{i15} \times D_{Soil=SS} + \beta_{i16} \times D_{Hydraulic=A} \\
& + \beta_{i17} \times D_{Hydraulic=B} + \beta_{i18} \times D_{Hydraulic=C} + \beta_{i19} \times D_{Hydraulic=D} \\
& + \beta_{i20} \times D_{Corrosivity=Low} + \beta_{i21} \times D_{Corrosivity=Medium} \\
& + \beta_{i22} \times D_{Corrosivity=High} \qquad \text{Eq. 5.1}
\end{aligned}
$$

where i = 1, 2, 3 and 4 determines the condition level of sewer pipes, α_i is intercept, β_{i1}, $\beta_{i2}, \ldots,$ β_{i22} are regression coefficients, and D_i is dummy variable to assign different values to categorical independent variables. Dummy variable is a numerical variable to take the value 0 or 1 to specify the absence or presence of a categorical variable. For example, assume the condition of a PVC pipe is predicted through above equation. In this condition, dummy variable assigns value 1 to PVC pipe and 0 to VCP pipe. Table 5-2 demonstrates the categories of dummy variables used in this study to develop multinomial and logistic regressions.

Table 5-2 Description of Dummy Variables

Independent Variable	Dummy Variable	Category
Pipe material	D_{PVC}	PVC pipes
	D_{VCP}	VCP pipes
Soil type	$D_{Soil=FS}$	Fine sand
	$D_{Soil=SG}$	Silty gravel and sand
	$D_{Soil=CS}$	Clayey soil
	$D_{Soil=SS}$	Silty soil
Hydraulic group	$D_{Hydraulic=A}$	Hydraulic group A
	$D_{Hydraulic=B}$	Hydraulic group B
	$D_{Hydraulic=C}$	Hydraulic group C
	$D_{Hydraulic=D}$	Hydraulic group D
Soil Corrosivity	$D_{Corrosivity=Low}$	Low corrosivity
	$D_{Corrosivity=Medium}$	Medium corrosivity
	$D_{Corrosivity=High}$	High corrosivity

5.2.2 Parameters Estimation

As described before, 80% of data was used to train the multinomial logistic regression by SPSS software. In logistic regression, if the dependent variable includes N categories, one of these categories is selected as the reference category. The remaining $N-1$ categories are used to generate logistic regression equations. For development of multinomial logistic regression in this dissertation, condition level 5 was selected as reference category. Pipe age, diameter, flow, depth, slope, length, sulfate, pH and water table were entered as covariate, and pipe material, soil type, hydraulic group and soil corrosivity were the factors to generate multinomial logistic regression.

Maximum Likelihood Estimation (MLE) was used to estimate the parameters in the model. Significance of the variables was identified by Wald test and P-test with the confidence interval of 95%. Parameter estimates for different condition of sanitary sewer pipes are provided in Tables 5-3 through 5-6.

Table 5-3 Parameter Estimates for Condition Level 1

Variable	Coefficient (β)	Standard Error	Wald	P Value	Expected Value
Intercept	-0.838	0.984	0.725	0.000	
Age	0.026	0.002	107.140	0.000	1.026
Diameter	-0.015	0.008	3.139	0.006	0.985
Flow	0.000	0.000	31.760	0.000	1.000
Depth	0.008	0.008	0.927	0.336	1.008
Slope	-0.020	0.017	1.458	0.227	0.980
Length	0.000	0.000	2.590	0.108	1.000
Sulfate	-1.921	0.997	3.712	0.054	0.147
pH	0.086	0.038	5.204	0.023	1.090
Water Table	0.000	0.001	0.120	0.729	1.000
Material = PVC	-2.211	0.091	586.270	0.000	0.110
Material = VCP (Reference)	0	.	.	.	.
Soil Type = Clayey Soil	-0.965	1.181	0.667	0.414	0.381
Soil Type = Fine Sand	-1.019	0.936	1.186	0.276	0.361
Soil Type = Silty Gravel and Sand	-1.166	0.933	1.559	0.212	0.312
Soil Type=Silty Soil (Reference)	0	.	.	.	.
Hydraulic =A	0.188	0.112	2.825	0.093	1.207
Hydraulic =B	-0.136	0.171	0.636	0.425	0.873
Hydraulic =C	0.413	0.135	9.334	0.122	1.511
Hydraulic =D (Reference)	0	.	.	.	.
Corrosivity = High	-0.283	0.080	12.666	0.060	0.753
Corrosivity = Low	0.404	0.325	1.544	0.214	1.498
Corrosivity = Moderate (Reference)	0	.	.	.	.

Table 5-4 Parameter Estimates for Condition Level 2

Variable	Coefficient (β)	Standard Error	Wald	P Value	Expected Value
Intercept	-15.58	0.515	917.230	0.000	
Age	0.054	0.004	194.311	0.000	1.055
Diameter	-0.110	0.019	32.616	0.000	0.896
Flow	0.000	0.000	5.099	0.024	1.000
Depth	-0.24	0.014	3.064	0.080	0.976
Slope	0.002	0.026	0.008	0.928	1.002
Length	0.004	0.000	89.140	0.000	1.004
Sulfate	-3.269	1.570	4.334	0.037	0.038
pH	0.106	0.058	3.395	0.065	1.112
Water Table	0.001	0.001	0.374	0.541	1.001
Material = PVC	-1.432	0.152	88.735	0.000	0.239
Material = VCP (Reference)	0	.	.	.	.
Soil Type = Clayey Soil	1.627	1.161	1.578	0.265	0.301
Soil Type = Fine Sand	1.590	0.103	1.512	0.248	0.321
Soil Type = Silty Gravel and Sand	1.477	0.000	1.568	0.212	0.316
Soil Type=Silty Soil (Reference)	0	.	.	.	.
Hydraulic =A	-0.057	0.166	0.120	0.729	0.944
Hydraulic =B	0.250	0.256	0.956	0.328	1.284
Hydraulic =C	-0.194	0.219	0.786	0.375	0.823
Hydraulic =D (Reference)	0b	.	.	.	.
Corrosivity = High	-0.094	0.124	0.581	0.446	0.910
Corrosivity = Low	1.007	0.457	4.854	0.028	2.736
Corrosivity = Moderate (Reference)	0	.	.	.	.

Table 5-5 Parameter Estimates for Condition Level 3

Variable	Coefficient (β)	Standard Error	Wald	P Value	Expected Value
Intercept	-7.308	1.266	33.306	0.000	
Age	0.167	0.004	1600.086	0.000	1.182
Diameter	-0.081	0.016	25.483	0.000	0.922
Flow	0.000	0.000	.884	0.347	1.000
Depth	-0.028	0.014	4.077	0.043	0.972
Slope	-0.145	0.047	9.306	0.002	0.865
Length	0.004	0.000	142.327	0.000	1.004
Sulfate	-2.114	1.419	2.221	0.136	0.121
pH	0.109	0.061	3.242	0.072	1.115
Water Table	0.000	0.001	.006	0.939	1.000
Material = PVC	-0.520	0.228	5.218	0.022	0.595
Material = VCP (Reference)	0	.	.	.	.
Soil Type = Clayey Soil	-2.015	1.168	3.003	0.090	0.185
Soil Type = Fine Sand	-2.005	1.148	3.051	0.081	0.135
Soil Type = Silty Gravel and Sand	-2.266	1.144	3.920	0.048	0.104
Soil Type=Silty Soil (Reference)	0	.	.	.	.
Hydraulic =A	-0.400	0.151	7.052	0.008	0.670
Hydraulic =B	0.210	0.249	0.707	0.400	1.233
Hydraulic =C	-0.549	0.181	9.144	0.002	0.578
Hydraulic =D (Reference)	0	.	.	.	.
Corrosivity = High	-0.250	0.114	4.802	0.028	0.779
Corrosivity = Low	0.610	0.543	1.265	0.261	1.841
Corrosivity = Moderate (Reference)	0	.	.	.	.

Table 5-6 Parameter Estimates for Condition Level 4

Variable	Coefficient (β)	Standard Error	Wald	P Value	Expected Value
Intercept	-6.428	1.431	20.182	0.000	
Age	0.158	0.003	2327.545	0.000	1.171
Diameter	-0.158	0.015	110.906	0.000	0.853
Flow	0.000	0.000	1.185	0.276	1.000
Depth	-0.002	0.011	0.035	0.851	0.998
Slope	-0.012	0.024	0.254	0.614	0.988
Length	0.009	0.000	727.361	0.000	1.009
Sulfate	-0.453	1.112	0.166	0.683	0.635
pH	-0.093	0.053	3.104	0.078	0.911
Water Table	0.000	0.001	0.258	0.611	1.000
Material = PVC	-0.435	0.146	8.829	0.003	0.647
Material = VCP (Reference)	0	.	.	.	.
Soil Type = Clayey Soil	-1.156	0.833	1.759	0.312	0.412
Soil Type = Fine Sand	-0.673	1.360	0.245	0.621	0.510
Soil Type = Silty Gravel and Sand	-0.897	1.358	0.436	0.509	0.408
Soil Type=Silty Soil (Reference)	0	.	.	.	.
Hydraulic =A	-0.703	0.130	29.344	0.000	0.495
Hydraulic =B	-0.046	0.203	0.051	0.821	0.955
Hydraulic =C	-0.881	0.158	31.245	0.000	0.414
Hydraulic =D (Reference)	0	.	.	.	.
Corrosivity = High	-0.400	0.091	19.244	0.000	0.670
Corrosivity = Low	-0.477	0.588	0.660	0.417	0.620
Corrosivity = Moderate (Reference)	0	.	.	.	.

The result of parameter estimates shows the significant level of variables are varying in different condition levels of sanitary sewer pipes. For example, pipe length is significant in condition levels 3, 4 and 5, but in condition level 2 the p-value is 0.108 which indicates it is an insignificant variable.

5.2.3 Significance of the Model

Significance of multinomial logistic regression model was evaluated based on a likelihood ratio test. This test indicates the likelihood ratio of the model with all independent variables (final model) to the model which all the parameter coefficients are 0 (null). The chi-square indicates the difference between – 2 log-likelihoods of the null and saturated models. As shown in Table 5-7, the significance level of final model is less than 0.05 and the model with all independent variables outperforms the null model.

Table 5-7 Significance Test of Multinomial Logistic Regression

Model	– 2 Log-likelihood	Chi Square	Degree of Freedom	Significance
Null	51,191.933			
Full	35,764.902	15,427.032	76	0.000

5.2.4 Validation of the Model

The result of multinomial logistic regression provided four equations for pipes in condition levels of 1, 2, 3, and 4. These equations are used to predict the probability of the pipes to be in a certain condition level. Logistic regression equations were developed based on the parameters estimated in section 5.2.2. In the first step, the coefficient of variables (*β*) in each condition level were used to build the logistic regression equations. Multinomial logistic regression equations are presented from Eq. 5.2 through 5.5.

$$g_1(x) = ln\left(\frac{P(C=1)}{P(C=5)}\right)$$

$$= -0.838 + 0.026 \times Age - 0.015 \times Diameter + 0.008 \times Depth$$

$$- 0.020 \times Slope - 1.921 \times Sulfate + 0.086 \times pH - 2.211 \times D_{PVC}$$

$$- 1.019 \times D_{Soil=FS} - 1.166 \times D_{Soil=SG} - 0.965 \times D_{Soil=CS}$$

$$+ 0.188 \times D_{Hydraulic=A} - 0.136 \times D_{Hydraulic=B} + 0.413 \times D_{Hydraulic=C}$$

$$+ 0.404 \times D_{Corrosivity=Low} - 0.283 \times D_{Corrosivity=High} \qquad Eq. 5.2$$

$$g_2(x) = ln\left(\frac{P(C=2)}{P(C=5)}\right)$$

$$= -15.58 + 0.054 \times Age - 0.110 \times Diameter - 0.024 \times Depth$$

$$+ 0.002 \times Slope + 0.004 \times Length - 3.269 \times Sulfate + 0.106 \times pH$$

$$+ 0.001 \times Water\ Table - 1.432 \times D_{PVC} + 1.590 \times D_{Soil=FS}$$

$$+ 1.477 \times D_{Soil=SG} + 1.627 \times D_{Soil=CS} - 0.057 \times D_{Hydraulic=A}$$

$$+ 0.250 \times D_{Hydraulic=B} - 0.194 \times D_{Hydraulic=C} + 1.007 \times D_{Corrosivity=Low}$$

$$- 0.094 \times D_{Corrosivity=High} \qquad Eq. 5.3$$

$$g_3(x) = ln\left(\frac{P(C=3)}{P(C=5)}\right)$$

$$= -7.380 + 0.167 \times Age - 0.081 \times Diameter - 0.028 \times Depth$$

$$- 0.145 \times Slope + 0.004 \times Length - 2.114 \times Sulfate + 0.109 \times pH$$

$$- 0.520 \times D_{PVC} - 2.005 \times D_{Soil=FS} + 2.266 \times D_{Soil=SG} - 2.015 \times D_{Soil=CS}$$

$$- 0.400 \times D_{Hydraulic=A} + 0.210 \times D_{Hydraulic=B} - 0.549 \times D_{Hydraulic=C}$$

$$+ 0.610 \times D_{Corrosivity=Low} - 0.250 \times D_{Corrosivity=High} \qquad Eq. 5.4$$

and

$$g_4(x) = ln\left(\frac{P(C=4)}{P(C=5)}\right)$$

$$= -6.428 + 0.158 \times Age - 0.158 \times Diameter - 0.002 \times Depth$$

$$- 0.012 \times Slope + 0.009 \times Length - 0.453 \times Sulfate - 0.093 \times pH$$

$$- 0.435 \times D_{PVC} - 0.673 \times D_{Soil=FS} - 0.897 \times D_{Soil=SG} - 1.156 \times D_{Soil=CS}$$

$$- 0.703 \times D_{Hydraulic=A} - 0.046 \times D_{Hydraulic=B} - 0.881 \times D_{Hydraulic=C}$$

$$- 0.477 \times D_{Corrosivity=Low} - 0.400 \times D_{Corrosivity=High} \qquad Eq. 5.5$$

Once the odds ratios were calculated, the probability of pipes being in each condition levels can be estimated. Equations 5.6 to 5.10 demonstrate the details of calculation probabilities associated with each condition level.

$$P(C=1) = \frac{e^{g_1(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \qquad \text{Eq. 5.6}$$

$$P(C=2) = \frac{e^{g_2(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \qquad \text{Eq. 5.7}$$

$$P(C=3) = \frac{e^{g_3(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \qquad \text{Eq. 5.8}$$

$$P(C=4) = \frac{e^{g_4(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \qquad \text{Eq. 5.9}$$

and

$$P(C=5) = \frac{1}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \qquad \text{Eq. 5.10}$$

Remained 20% of data was used to test the model by presented equations. The probability of all condition levels was calculated for each pipe segment and the higher probability was considered as predicted value for condition state of sanitary sewer pipe. Classification table was selected to evaluate the result of multinomial logistic regression as shown in Table 5-8.

Table 5-8 Multinomial Logistic Regression Classification Table

Observed	Predicted					Percent Correct Predicted
	1	2	3	4	5	
1	1,873	17	0	0	134	92.5%
2	391	25	0	0	157	4.4%
3	89	2	1	1	76	0.6%
4	57	2	0	21	240	6.6%
5	130	8	0	10	610	80.5%
Overall Percentage						65.8%

According to the result of classification table, in overall 65.8% of the pipe conditions were predicted correctly by multinomial logistic regression. 92.5% of the pipes in condition states 1 and 80.5% in condition state 5 were estimated correctly which indicates a high accuracy. In contrast, overall percentage of correct prediction for condition rating 2, 3 and 4 is only 10% which is not acceptable.

5.2.5 Results of Multinomial Logistic Regression

Multinomial logistic regression was used in this dissertation to predict all five condition levels of sanitary sewer pipes. According to the result of model, the condition of 65.8% of sanitary sewer pipes was predicted correctly, however pipes in condition levels 2, 3 and 4 were not estimated properly by multinomial logistic regression. The result of classification table indicated that most of the pipes which had condition states of 2, 3 and 4 were predicted in condition 1 or 5. Low number of appropriate pipe data in condition levels of 2, 3 and 4 caused low prediction rate for these three categories.

As shown in table 5-9, several researchers indicated that understanding the condition rating 4 and 5 (pipes in poor conditions) is more critical for utility companies and municipalities for prioritizing the pipes. Therefore, they just used two condition levels of 0 (pipe in good condition) and 1 (pipe in poor condition) to predict the future condition of sewer pipes.

Table 5-9 Prediction Models Developed by Binary Dependent Variables

Authors	Year	Condition Level 0	Condition Level 1
Davies et al.	2001	1, 2, 3, 4	5
Ariaratnam et al.	2001	1, 2, 3	4, 5
Koo and Ariaratnam	2006	1, 2, 3	4, 5
Ana et al.	2009	1, 2, 3	4, 5
Salman and Salem	2012	1, 2, 3	4, 5
Sousa et al.	2014	1, 2, 3	4, 5
Harvey and McBean	2014	1, 2, 3	4, 5
Kabir et al.	2018	1, 2, 3	4, 5
Laakso et al.	2018	0, 1, 2	3, 4

Since the prediction rate of multinomial logistic regression was not satisfactory in this dissertation, the condition of sanitary sewer pipes was classified into two groups of 0 as good and 1 as poor condition. Table 5-10 and Figure 5-1 illustrate the new category of pipe condition levels and frequency of them in sanitary sewer dataset.

Table 5-10 Category of Pipe Condition Levels

Datasets	Pipe Condition Levels	
Original Dataset	1, 2, 3	4, 5
Binary Dataset	0 (good)	1 (poor)

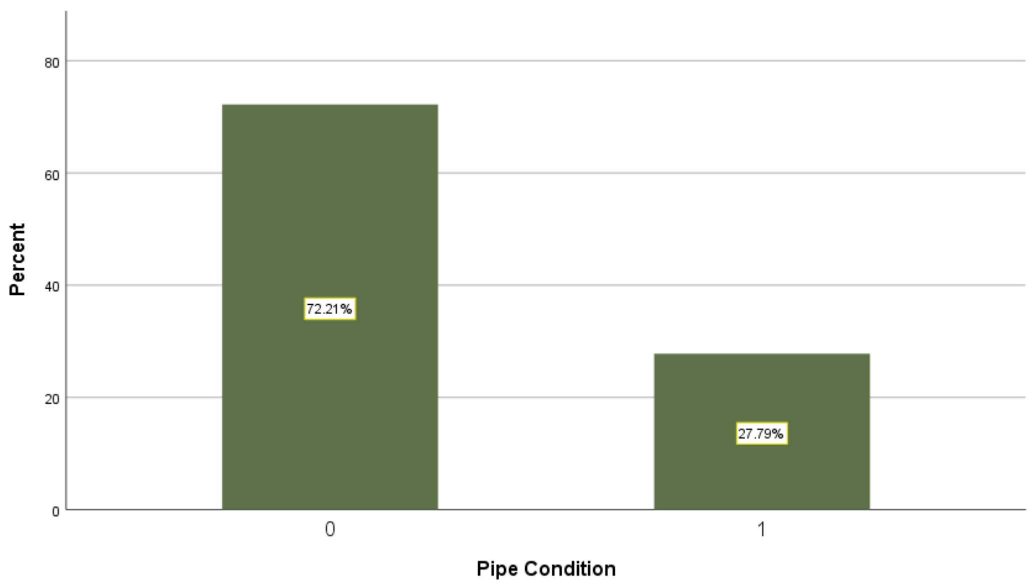


Figure 5-1 Percentage of Pipe Conditions

As shown in figure 5-1, the recoded dataset includes 72.21% and 27.79% pipes with condition levels 0 and 1 respectively. Sanitary sewer pipes in conditions 1, 2 and 3 were transformed into condition 0 and pipes in condition levels 4 and 5 were converted to condition 1. Next section will cover the detail of binary logistic regression developed in this study.

5.3 Binary Logistic Regression

5.3.1 Description of the Model

Binary logistic regression is used to develop prediction models when the output (dependent or response) variable is binary or dichotomous. A binary or dichotomous are variables which only take two values. For example, the output of the model can be true or false, success or failure and zero or one. In sewer condition prediction modeling, the dependent variable can be classified in good or poor conditions (Hosmer et al., 2013; Salman, 2010).

As described in multinomial logistic regression results, the dependent variable was transformed into a binary condition rating. Condition ratings 1, 2 and 3 were assigned to the pipes which are in good and stable condition (condition rating 0). And, the remaining pipes with condition ratings 4 and 5 were considered as pipes which are in poor condition and need immediate attention (condition rating 1). Therefore, binary logistic regression was developed to predict whether the pip is in good or poor condition states.

Based on the characteristics of dependent variable which has only two values, one regression equation is generated to estimate the condition of each pipe segments as shown in Equation 5.11.

$$
\begin{aligned}
ln\left(\frac{P(C=1)}{1-P(C=1)}\right) = {} & \\
& \alpha + \beta_1 \times Age + \beta_2 \times Diameter + \beta_3 \times Flow + \beta_4 \times Depth + \beta_5 \times Slope \\
& + \beta_6 \times Length + \beta_7 \times Sulfate + \beta_8 \times pH + \beta_9 \times Water\ Table \\
& + \beta_{10} \times D_{PVC} + \beta_{11} \times D_{VCP} + \beta_{12} \times D_{Soil=FS} + \beta_{13} \times D_{Soil=SG} + \beta_{14} \times D_{Soil=CS} \\
& + \beta_{15} \times D_{Soil=SS} + \beta_{16} \times D_{Hydraulic=A} + \beta_{17} \times D_{Hydraulic=B} \\
& + \beta_{18} \times D_{Hydraulic=C} + \beta_{19} \times D_{Hydraulic=D} + \beta_{20} \times D_{Corrosivity=Low} \\
& + \beta_{21} \times D_{Corrosivity=Medium} + \beta_{22} \times D_{Corrosivity=High} \qquad \text{Eq. 5.11}
\end{aligned}
$$

where α is intercept, β_1, β_2,…, β_{22} are regression coefficients, and D_i is dummy variable to assign different values to categorical independent variables.

5.3.2 Parameters Estimation

Similar to development of multinomial logistic regression, 80% of data was used to train the binary logistic regression by SPSS software. In logistic regression, if the dependent variable includes *N* categories, one of these categories is selected as the reference category. For development of binary logistic regression in this dissertation, condition level 0 was selected as reference category. Pipe age, diameter, flow, depth, slope, length, sulfate, pH, material, soil type, hydraulic group, soil corrosivity and water table were entered as covariate to generate binary logistic regression.

Maximum Likelihood Estimation (MLE) was used to estimate the parameters in the model. Significance of the variables was identified by Wald test and P-test with the confidence interval of 95%. A backward stepwise variable selection was used to identify the variables that have more predictive power to forecast condition of sanitary sewer pipes. Forward and backward stepwise selection are statistical techniques to screen the independent variables. In these methods, the variables which have enough predictive power are remained in the model and idle variables are removed stepwise. For example, if a dataset has hundred independent variables it would be beneficial to keep the appropriate variables on the model and remove the rest.

backward stepwise selection started with full model and considering all 13 independent variables and then the variables that have least influence were excluded from the model. The variables with the highest p-score were the candidate for removing from the model. Parameter estimates for different condition of sanitary sewer pipes are provided in Table 5-11.

Table 5-11 Parameter Estimates in Binary Logistic Regression for Condition Level 1

Variable	Coefficient (β)	Standard Error	Wald	P Value	Expected Value
Intercept	-6.114	0.963	40.290	.000	
Age	0.143	0.003	2638.310	.000	1.154
Diameter	-0.109	0.012	88.831	0.000	0.897
Flow	0.000	0.000	0.219	0.640	1.000
Depth	-0.009	0.009	1.005	0.316	0.991
Slope	-0.030	0.022	1.910	0.167	0.970
Length	0.007	0.000	641.608	0.000	1.007
Sulfate	0.114	0.947	0.014	0.904	1.120
pH	-0.066	0.043	2.388	0.122	0.936
Water Table	-0.001	0.002	18.64	0.006	1.001
Material = PVC	-.317	0.307	1.065	0.002	0.728
Material = VCP (Reference)	0	.	.	.	.
Soil Type = Clayey Soil	-10.873	190.997	0.003	0.955	0.524
Soil Type = Fine Sand	-0.884	0.890	0.986	0.321	0.413
Soil Type = Silty Gravel and Sand	-1.052	0.888	1.402	0.236	0.349
Soil Type=Silty Soil (Reference)	0	.	.	.	.
Hydraulic =A	-0.645	0.106	2.234	0.064	0.525
Hydraulic =B	0.037	0.171	0.047	0.828	1.038
Hydraulic =C	-0.904	0.125	1.398	0.075	0.405
Hydraulic =D (Reference)	0	.	.	.	.
Corrosivity = High	-0.215	0.076	8.049	0.085	0.807
Corrosivity = Low	-0.282	0.410	0.474	0.491	0.754
Corrosivity = Moderate (Reference)	0	.	.	.	.

As shown in Table 5-11, pipe age, material, length, diameter and water table are the significant variables in binary logistic regression. Table 5-12 illustrates the result of backward stepwise method after elimination of 8 insignificant variables from the saturated model.

Table 5-12 Parameter Estimates in Binary Logistic Regression (Backward Stepwise)

Variable	Coefficient (β)	Standard Error	Wald	P Value
Intercept	-8.060	0.171	2227.929	0.000
Age	0.142	0.003	2865.361	0.000
Diameter	-0.114	0.009	175.763	0.000
Length	0.006	0.000	653.456	0.000
Water Table	-0.002	0.000	14.070	0.000
Material = PVC	-0.189	0.126	8.418	0.015
Material = VCP (Reference)	0	.	.	.

5.3.3 Significance of the Model

Similar to multinomial logistic regression, significance of binary regression was evaluated based on a likelihood ratio test. This test indicates the likelihood ratio of the model with all independent variables (final model) to the model which all the parameter coefficients are 0 (null). The chi-square indicates the difference between – 2 log-likelihoods of the null and saturated models. As shown in Table 5-13, the significance level of final model is less than 0.05 and the model with all independent variables outperforms the null model.

Table 5-13 Significance Test of Multinomial Logistic Regression

Model	– 2 Log-likelihood	Chi Square	Degree of Freedom	Significance
Null	23,360.459			
Full	12,742.063	10,618.396	19	0.000

5.4 Gradient Boosting Tree

5.4.1 Description of the Model

The third objective of this dissertation was to compare the performance of different statistical and artificial intelligence models for predicting the condition levels of sanitary sewer pipes. Gradient boosting tree is the first AI model developed in this study. This model is a machine learning technique for regression and classification, which provides a prediction model by improving the performance of a weak learner. In this method, a weak learner is run repeatedly on various training data to develop classifiers. Then, the classifiers are combined into a single strong classifier to achieve a higher accuracy (Rokach and Maimon, 2015).

Similar to previous models, the objective of machine learning approach is to find a relationship between dependent and independent variables. In this study, the relationship between thirteen independent variables and condition level of sanitary sewer pipes (dependent variable) was investigated to predict whether pipe is in good or poor condition. Gradient boosting tree model was developed by Python 3.7.3 with application of different computing libraries, such as Numpy, Pandas and Sklearn. XGBoost algorithm was the primary source to develop the gradient boosting tree in this dissertation.

XGBoost is a machine learning technique for tree boosting that designed by Chen and Guestrin in 2014. According to Zhang et al. (2018), XGBoost was one of the most popular machine learning methods in 2015 for developing prediction models. This technique implements machine learning algorithms under the gradient boosting frameworks by combining weak base learning models into a stronger learner.

5.4.2 Development of the Model

5.4.2.1 Five-Fold Cross Validation

Cross-validation is a strategy to avoid overfitting and uncertainty during training and testing the models. The basic idea is that the dataset is partitioned into K equal size folders. For example, if there are 200 datapoints and 10 folds, there will be 20 datapoints in each folder. In order to develop gradient boosting tree in this dissertation, five-fold cross validation method was used to randomly select 80% of data for train and 20% for testing the model. Sklearn library was employed to split sanitary sewer dataset into 5 folders (each folder consists 20% of data).

In 5-Fold cross validation, five separate learning experiments were run. In each iteration one folder was selected as testing set and the remaining four folders were combined to build training set. This procedure was repeated in five different iterations and then the average value was showed as result of the model. The key element of cross validation method is that all the datapoints are used for testing and training the model. Figure 5-2 illustrates the detail of 5-Fold cross validation method in different iteration of the model development.

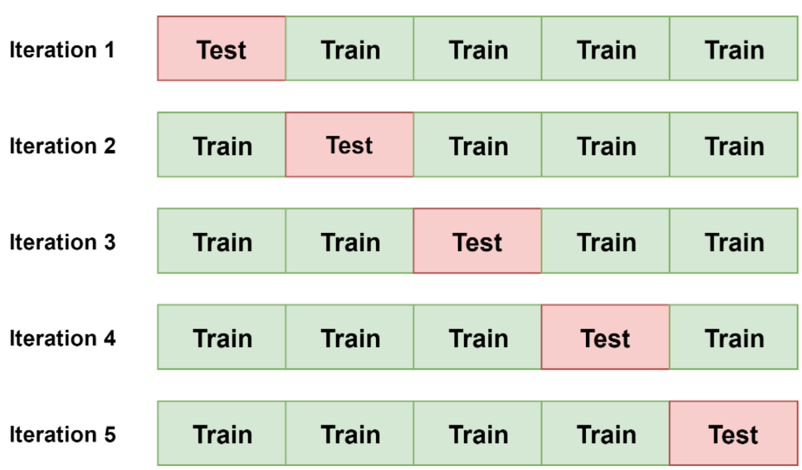


Figure 5-2 Five-Fold Cross Validation

5.4.2.2 Training the Model

As explained before, XGBoost algorithm was used to develop gradient boosting tree model in this study. Three main parameters of XGBoost algorithm are; 1) General parameters, 2) booster parameters, and 3) learning tasks. General parameters specify the overall functionality of the algorithm. Booster parameters involve tree booster, linear booster or dart, and determine how the algorithm boosts the performance of the model. And, learning task parameters define the optimization objectives at each iteration. The description of XGBoost parameters in Python are presented in Table 5-14.

Table 5-14 Description of XGBoost Parameters in Scikit-learn API

Main Parameters	Parameters	Description
General Parameters	booster	Specify type of model to be run at each iteration: gbtree, gblinear or dart
	silent	Whether to print messages while running boosting
	nthread	Used for parallel processing and set number of parallel threads
Booster Parameters	max_depth	Maximum depth of tree for base learners
	learning_rate	Boosting learning rate
	n_estimators	Number of trees to fit
	n_jobs	Number of parallel threads
	gamma	Minimum loss reduction required to make a split
	min_child_weight	Define the minimum sum of weights of all observations required in a child
	max_delta_step	Maximum delta step which allows each tree's weight estimation
	subsample	Subsample ratio of the training instance
	colsample_bytree	Subsample ratio of columns when constructing each tree
	colsample_bylevel	Subsample ratio of columns for each level
	reg_alpha	L1 regularization term on weight (analogous to Lasso regression)
	reg_lambda	L2 regularization term on weights (analogous to Ridge regression)
	scale_pos_weight	Balancing of positive and negative weights
	base_score	The initial prediction score of all instances, global bias
	missing	Value in the data which needs to be present as a missing value
Learning Task Parameters	objective	Defines the loss function to be minimized: binary: logistic, multi:softmax, multi:softprob
	seed	Random number seed

XGBoost classification was implemented using the Scikit-learn API libraries. Firstly, tree booster was selected to specify type of model to be run at each iteration. Tree booster was preferred to linear booster and dart as it is a tree-based model and can detect non-linear relationships between variables. The maximum depth of tree was set 4 to reduce the complexity of the model and overfitting. This feature presents depth of each tree and is used to control overfitting. The value of maximum depth is usually between 3 and 10, and the higher depth can determine very specific relationship between variables and increase the risk of overfitting. The performance of the model was evaluated with higher depth values such as 5 and 6, but the results showed very complex outcome.

The learning rate shrinks the weights in each step to develop stronger model and typically it is between 0.01 to 0.2. The average value of 0.1 was selected to implement the model in this dissertation. The higher values close to 0.2 and lower values close to 0.01 showed weaker training results. The number of estimators determines the number of required trees to fit the model. This value is usually between some hundreds to thousands and changes based on parameters used in the model. The number of estimators was set 500 for developing the model. As dependent variable has two classes (good and poor), binary logistic was selected for objective part. Other values were set as default and gradient boosting tree was implemented.

5.5 K-Nearest Neighbors

5.5.1 Description of the Model

K-Nearest Neighbors is the second AI model developed in this study to predict condition of sanitary sewer pipes. K-nearest neighbors are applicable to develop both regression and classification models. Nearest neighbors method works based on identifying the labels of K-nearest patterns in data space. Nearest neighbor techniques have better performance when the datasets are large with low dimensions (Kramer, 2016).

Similar logistic regression and gradient boosting tree models, thirteen independent variables were used to predict whether pipe is in good or poor condition.

KNN model was developed by Python 3.7.3 with application of different computing libraries, such as Numpy, Pandas, Sklearn and seaborn. Sklearn library provides both supervised and unsupervised learning methods. K neighbors classifier and radius neighbors classifier are two main classifier algorithms that Sklearn provides. K neighbors classifier implements learning based on the nearest neighbors of each query point, while radius neighbors classifier implements learning based on the number of neighbors within a fixed radius of each training point.

K neighbors classifier (sklearn.neighbors.KNeighborsClassifier) algorithm was the primary source to develop the KNN model in this dissertation. This algorithm is the most commonly used technique to develop KNN model. Fast computation of KNN is one of the most important advantages of this machine learning approach (Pedregosa et al., 2011).

5.5.2 Development of the Model

5.5.2.1 Five-Fold Cross Validation

Similar to the gradient boosting tree model, five-fold cross validation method was used to randomly select 80% of data for train and 20% for testing the model. Sklearn library was employed to split sanitary sewer dataset into 5 folders and each folder including 20% of data. The detail of cross validation was presented in previous section.

5.5.2.2 Training the Model

As explained before, K neighbors classifier algorithm was used to develop KNN model for predicting whether sanitary sewer pipes are in good or poor condition levels. This model has much less complexity than gradient boosting tree and less parameters are required to be set during training the model. Table 5-15 presents the K neighbors classifier parameters in Scikit-learn.

Table 5-15 Description of KNN Parameters in Scikit-learn

Parameters	Description
n_neighbors	Specify number of neighbors: default = 5
weights	weight function used in prediction: uniform, distance
algorithm	Algorithm used to compute the nearest neighbors: auto, ball_tree, kd_tree, brute
leaf_size	This parameter is related to BallTree or KDTree
p	Power parameter for the Minkowski metric: 1 for Manhattan distance, 2 for Minkowski distance
metric	The distance metric to use for the tree
metric_params	Additional keyword arguments for the metric function
n_jobs	The number of parallel jobs to run for neighbors search

In order to generate KNN model, some parameters such as, number of jobs, metrics, p, and leaf size were set as a default. Auto algorithm was selected to compute the nearest neighbors as this function attempts to find the most appropriate algorithm. The weight parameter was set uniform to keep the consistency of the model.

Identifying the number of neighbors is the most important activity during implementation of KNN model. In general, risk of overfitting is high when lower values are selected for number of neighbors. This parameter can be set manually or be determined through automatic methods. Therefore, two techniques were used to identify right number of neighbors and decrease the risk of overfitting. Firstly, the KNN model was run using different neighbor values from 1 to 15 and then the accuracy of the test and train dataset were compared to identify the most appropriate number of neighbors. Figure 5-3 illustrates the varying number of neighbors in KNN model.

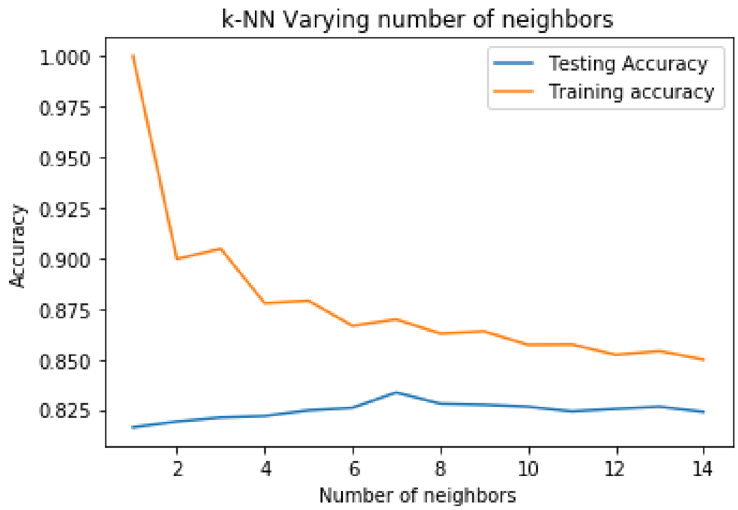


Figure 5-3 Varying Number of Neighbors in KNN Model

The KNN plot showed that the highest accuracy achieved when the number of neighbors is 7. The second method identified the number of neighbors based on misclassification error in different number of neighbors as shown in Figure 5-4. The optimal number of neighbors is 7 with lowest error in KNN model.

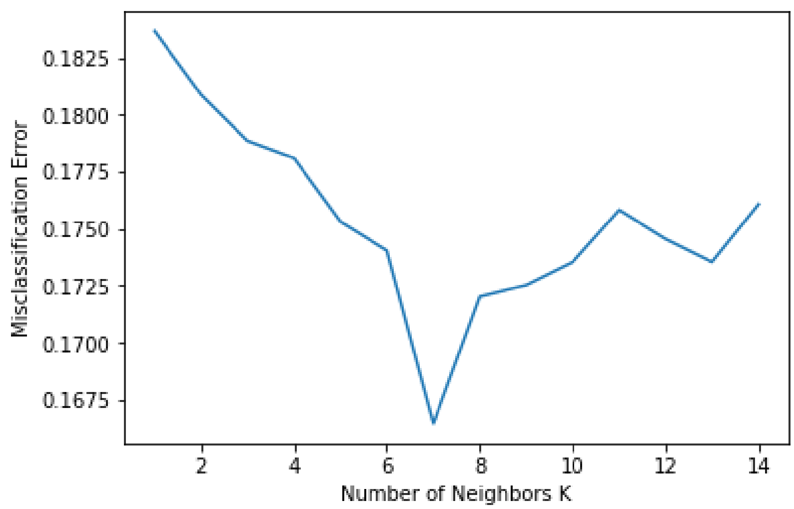


Figure 5-4 Misclassification Error in Different Neighbor Numbers

Both techniques identified that number 7 is the best value for selecting the quantity of neighbors in KNN model. Therefore, the model was implemented with 7 neighbors and the results of validation is presented in next chapter.

5.6 Chapter Summary

This chapter presented the detailed overview of developing sanitary sewer pipes condition prediction models, through multinomial logistic regression, binary logistic regression, gradient boosting tree and KNN models. Numerous structures were tested and the best architecture among them was chosen for further explanation and development. Different steps of training and generating the statistical and artificial intelligence models were presented. The validation of the models and the significance of the variables will be presented in next chapter.

Chapter 6 Results and Discussions

6.1 Introduction

The previous chapter information about the development of the models was provided. This chapter presents the results and validation of binary logistic regression, gradient boosting tree and KNN models. Additionally, the effect of influence variables will be comprehensively investigated in this chapter.

6.2 Binary Logistic Regression

6.2.1 Validation of the Model

The result of binary logistic regression provided one equation to predict the condition levels of sanitary sewer pipes. The independent variables and parameter estimate in section 5.3.2 are used to develop the odds ratio. These equations are used to predict the probability of the pipes being in poor condition level. As significant variables were identified, the binary logistic regression equation was generated using only significant variables in the model. Equation 6.1 present the result of binary logistic regression.

$$g(x) = ln\left(\frac{P(C=1)}{1-P(C=1)}\right) =$$

$$-8.06 + 0.142 \times Age - 0.114 \times Diameter + 0.006 \times Length$$
$$- 0.002 \times Water\ Table - 0.189 \times D_{PVC} \qquad \text{Eq. 6.1}$$

Once the odds ratio was calculated, the probability of pipes being in poor or good condition can be estimated by using Equation 6.2 and 6.3.

$$P(C=1) = \frac{1}{1+e^{-g(x)}} \qquad \text{Eq. 6.2}$$

and

$$P(C=0) = 1 - P(C=1) \qquad \text{Eq. 6.3}$$

Remained 20% of data was used to test the model by presented equations. The probability of pipe being in poor condition was calculated for each pipe segment and the higher probability was considered as predicted value for condition state of sanitary sewer pipe. Classification table was selected to evaluate the result of multinomial logistic regression as shown in Table 6-1.

Table 6-1 Binary Logistic Regression Classification Table

Observed	Predicted		Percent Correct Predicted
	0	1	
0	2,542	315	89.0%
1	300	824	73.3%
Overall Percentage			84.6%

According to the result of classification table, in overall 84.6% of the pipe conditions were predicted correctly by binary logistic regression. 89% of the pipes in condition level 0 and 73.3% in condition level 1 were estimated correctly which indicates a high accuracy. The detail of confusion matrix was presented in chapter 3. When pipes in good condition are considered the positive class of interest, TP (true positive) and TN (true negative) represent correctly classified pipes. TP demonstrates number of pipes which are actually in good condition and correctly predicted to be in good condition, and TN determine number of pipes which are actually in poor condition and correctly predicted to be in poor condition state.

Incorrect classifications are presented by FP (false positive) and FN (false negative) values. FP reveals the pipes which are predicted in good condition, when in fact they are pipes in poor condition, and FN determined the pipes predicted to be in poor condition, while they are actually good pipes. Equations 6.4 through 6.7 determine how to calculate true and false positive and negative rates based on the results of confusion matrix.

$$True\ Positive\ Rate = TPR = Sensitivity = \frac{TP}{TP + FN} \qquad \text{Eq. 6.4}$$

$$True\ Negative\ Rate = TNR = Specificity = \frac{TN}{FP + TN} \qquad \text{Eq. 6.5}$$

$$False\ Positive\ Rate = FPR = 1 - Specificity \qquad \text{Eq. 6.6}$$

$$False\ Negative\ Rate = FNR = 1 - Sensitivity \qquad \text{Eq. 6.7}$$

Table 6-2 presents the result of calculating true positive, true negative, false positive and false negative rates. The prediction performance of binary logistic regression was also evaluated by Receiver Operating Characteristic (ROC) curve. ROC curve is a useful visual tool which is a plot of true positive rate (TPR) and false positive rate (FPR). The area under the ROC curve illustrates the model performance, where perfect models have an area close to 1 and random models have an area close to 0.5. An area under the ROC curve greater than 0.7 demonstrates the model is acceptable (Hosmer et al., 2013). Figure 6-1 illustrates the ROC curve for binary logistic regression.

Table 6-2 Binary Logistic Regression Model Performance

Rates	Values
True positive rate (TPR)	89%
True negative rate (TNR)	73.3%
False positive rate (FPR)	26.7%
False negative rate (FNR)	11%

The area under ROC curve is 0.903 which shows binary logistic regression has acceptable result. Therefore, logistic regression equation can be used to predict the condition of pipes which have not been inspected yet.

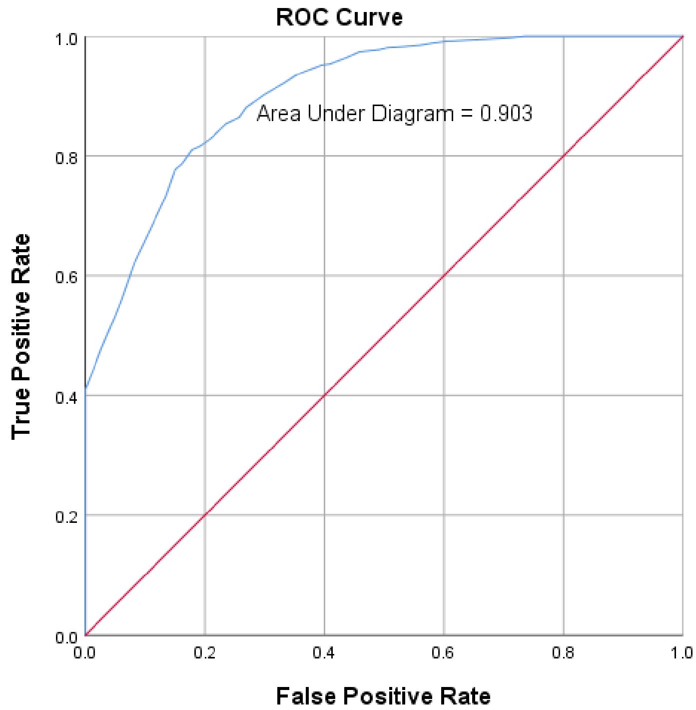


Figure 6-1 Binary Logistic Regression ROC Curve

6.2.2 Deterioration Curve

The outcomes of binary logistic regression can be used to develop a visual presentation of the probability of pipes being in poor or good conditions. A deterioration curve was developed in this dissertation to show how the condition of sewer pipes degrade over time and the age of the asset while considering the effect of all significant variables. The deterioration curve was developed by using the mean values of the numerical dependent variables and changing the age by one-year increments in the binary logistic regression equation. Figure 6-2 illustrates the deterioration curve of sanitary sewer pipes in the network for two different pipe materials: PVC and VCP.

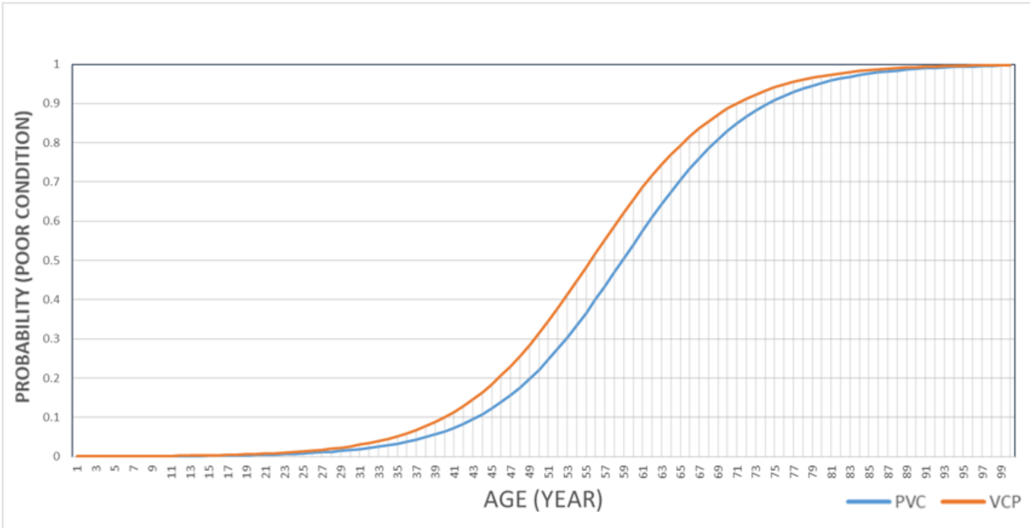


Figure 6-2 Sanitary Sewer Pipes Deterioration Curve

According to the result of deterioration curve, PVC and VCP pipes have almost similar behavior over time. PVC pipes seems to degrade slower resulting in a delayed poor condition score. In general, VCP pipes would have a shorter life than PVC pipes since they have more brittle qualities than PVC. The primary benefit of a deterioration curve is the possibility of predicting future pipe conditions within the network. Furthermore, the short-term and long-term behavior of sewer pipes can be monitored over time. Deterioration curves can be generated for each individual pipe and they can be a valuable tool for prioritizing the pipes and providing a logical inspection schedule (Malek Mohammadi et al., 2019).

6.2.3 Influence Variables

As presented in previous sections, of the 13 independent variables considered only five variables were retained in the final model. The significant variables were pipe age, diameter, length, material and water table as listed in Table 6-3.

Table 6-3 Significant Variables in Binary Logistic Regression Model

Variable	Coefficient (β)	Standard Error	Wald	P Value	Expected Value
Age	0.142	0.003	2865.361	0.000	1.053
Diameter	-0.114	0.009	175.763	0.000	0.892
Length	0.006	0.000	653.456	0.000	1.006
Water Table	-0.002	0.000	14.070	0.000	0.998
Material = PVC	-0.189	0.126	8.418	0.015	0.828
Material = VCP (Reference)	0	.	.	.	.

In this part the influence of each significant variable is presented in detail to understand in what way and how much they affect condition of sewer pipes gradually.

6.2.3.1 Significant Variables

Pipe Age. The binary logistic regression results identified that pipe age affects condition of sanitary sewer pipes strongly with Wald = 2865.361 (Sig. = 0.000). As could be expected, when the sanitary sewer pipes aged, the probability of pipes being in poor condition increased. According to the binary logistic regression equation, the coefficient of pipe age is positive, and a unit increase in age results in an increase in the probability that the pipe is in poor condition level.

As shown in Table 6-3, the odds ratio (expected value) of pipe age is 1.053 which reveals that for a unit increase in age, the odds of sewer being in poor condition are multiplied by 1.053. This result shows a 5.3% raise (relative to the odds of sanitary sewer with age 1 year less) when all other conditions are constant. This means, when a sanitary sewer pipe ages by 20 years, the odds of being in poor condition increase to $(1.053)^{20}=2.81$ which is 28.1% over the 20-year period. Each pipe material has a specific useful life and the physical properties of pipe material are changed during aging process. Thus, it is obvious that with aging the sewer pipes the deterioration rate increases and other factors such as external load, corrosion and groundwater level can provide higher risk of collapse or failure. The above results support the finding of several studies presented in section 2.7.2.1.

Pipe Diameter. Pipe diameter was also found to affect deterioration of sanitary sewer pipes largely with Wald = 175.763 (Sig. = 0.000). According to the binary logistic regression equation, the coefficient of pipe diameter is negative, and a unit increase in size results in a decrease in probability of pipe being in poor condition. Therefore, larger pipes are more resistant to pipe deterioration.

The odds ratio of pipe diameter is 0.892 which reveals that for a unit increase in diameter, the odds of sewer being in poor condition are multiplied by 0.892. This result shows a 10.8% reduction (relative to the odds of sanitary sewer with 1 in. smaller diameter) when all other conditions are constant. Thus, increasing the diameter of a sanitary sewer pipe from 8 in. to 18 in., would reduce the odds of being in poor condition to $(0.892)^{10}=0.319$ which is 68.1% over the 10 in. increase. Figure 6-3 and 6-4 illustrates the deterioration of PVC and VCP pipes with different diameters of 8, 16, 25, 32 and 40 inches. The effect of pipe diameter on condition of sanitary sewer pipes is more evident in Figure 6-5 which shows the variance of pipe diameter from 2 to 80 inches for a PVC and VCP pipe with constant age of 50 years.

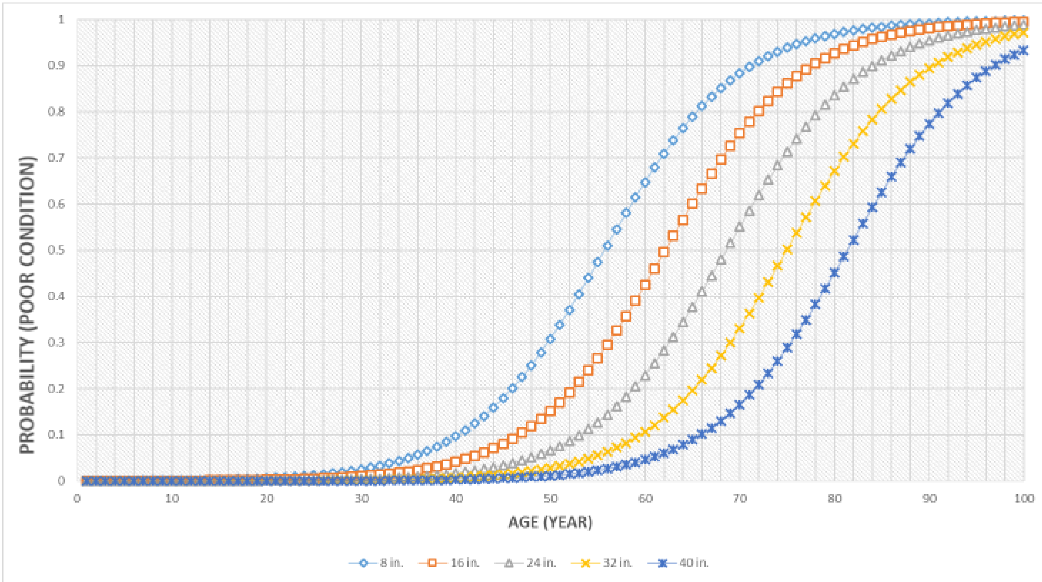


Figure 6-3 Deterioration Curve for PVC Pipes with Different Diameter Ranges

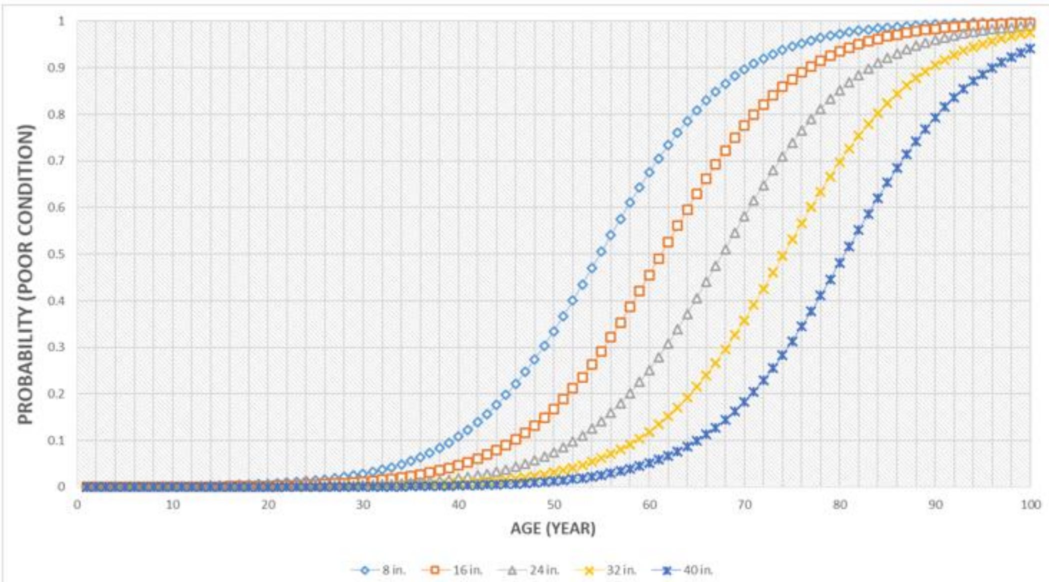


Figure 6-4 Deterioration Curve for VCP Pipes with Different Diameter Ranges

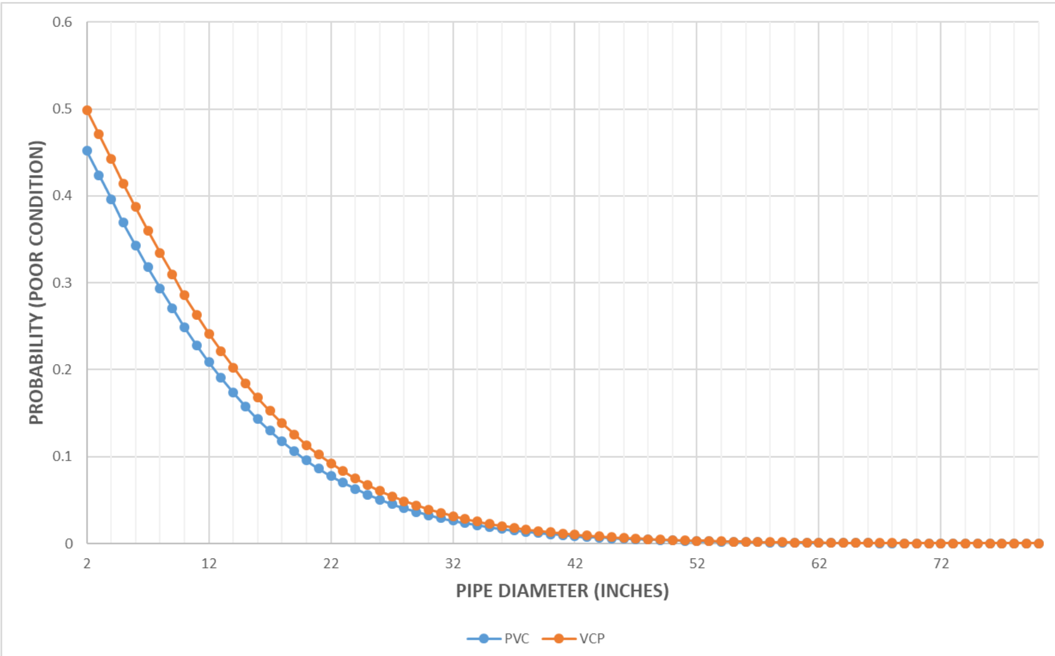


Figure 6-5 Effect of Pipe Diameter on Condition of a 50-year Old Pipe

Previous studies stated contradictory results regarding the effect of pipe diameter on condition of sewer pipes as presented in section 2.7.2.3. The result of this dissertation indicated that when pipe diameter increases the likelihood of a pipe being in a poor condition decrease and larger sewers are at a lower risk than small ones. A probable explanation could be that the pipe designers underestimate the required depth of cover and loading traffics for the smaller pipes. Additionally, larger pipes are often buried deeper and more appropriate design and construction crew are used to install them. With the occurrence of obstacles in the conduit, segments with larger diameter still enable to convey wastewater and small diameters are more likely to deteriorate due to lose of hydraulic flow. The above results support the finding of several studies presented in section 2.7.2.3.

Pipe Length. Sewer manhole to manhole length was also found to be a significant variable with Wald = 653.456 (Sig. = 0.000). The results of binary logistic regression revealed that as sewer reach increased in length, the probability of pipe being in poor condition increased. The coefficient of pipe length is positive in binary logistic regression equation, therefore longer pipe is deteriorated faster than shorter one.

The odds ratio of pipe length is 1.006 which reveals that for a unit increase in length, the odds of sewer being in poor condition are multiplied by 1.006. This result shows a 0.6% increase (relative to the odds of sanitary sewer with length 1 ft less) when all other conditions are constant. Thus, increasing the length of a sanitary sewer pipe from 50 ft to 100 ft, would reduce the odds of being in poor condition to $(1.006)^{50} = 1.348$ which is 34.8% over the 50 ft increase. The effects of pipe length on condition of PVC and VCP pipes is shown in Figures 6-6 and 6-7.

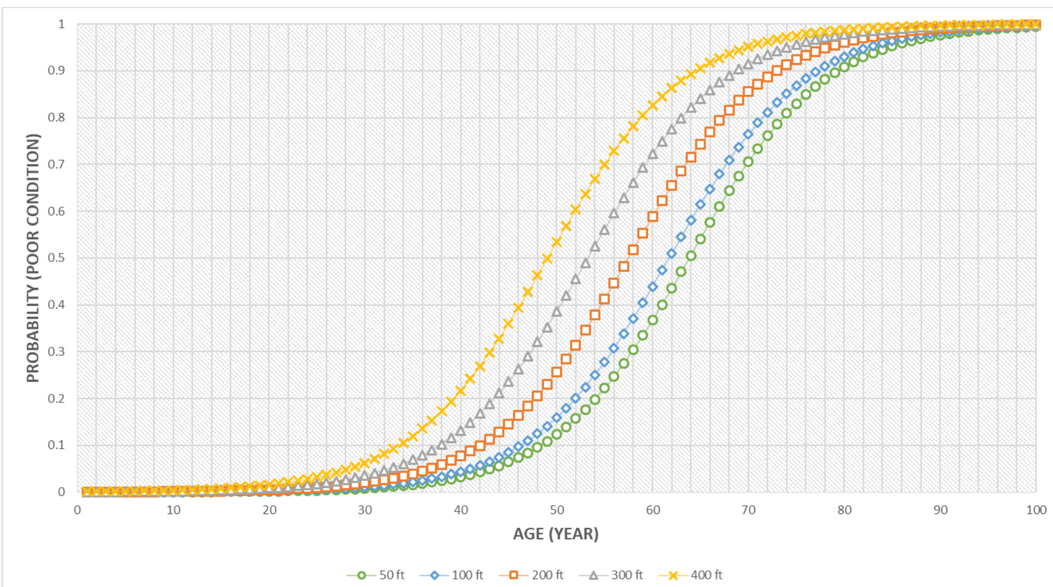


Figure 6-6 Deterioration Curve for PVC Pipes with Different Length

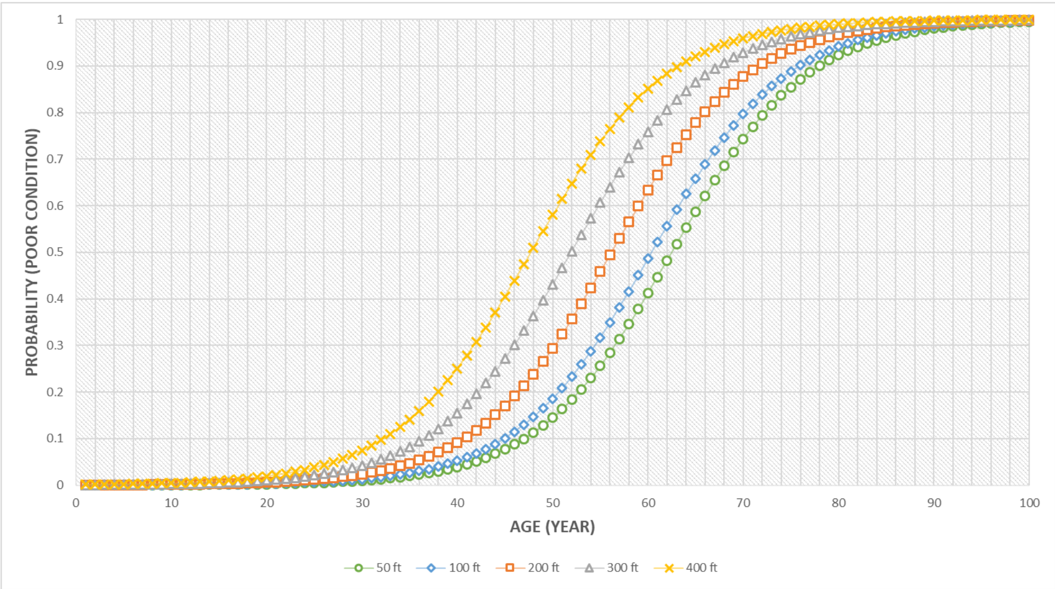


Figure 6-7 Deterioration Curve for VCP Pipes with Different Length

The effect of pipe length on condition of sanitary sewer pipes is more evident in Figure 6-8 which shows the pipe length from 10 to 500 feet for a PVC and VCP pipe with constant age of 50 years.

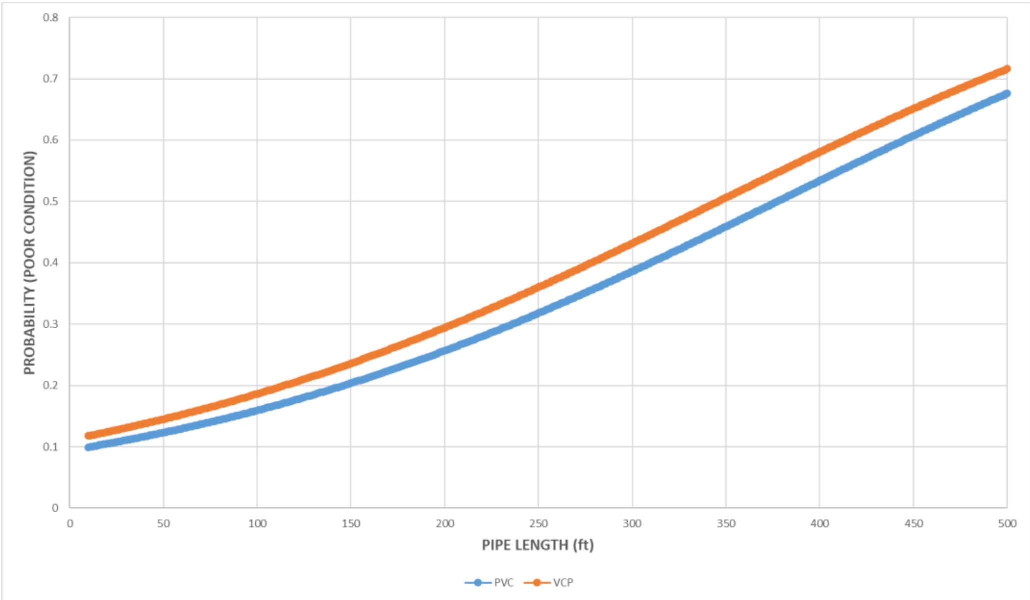


Figure 6-8 Effect of Pipe Length on Condition of a 50-year Old Pipe

The results of previous literatures showed a dual behavior regarding the effect of pipe length on deterioration of sewer pipes. Practically in all sewer pipe inventories, length of pipes is stored as manhole to manhole length of pipe segments, since CCTV is the most common tool for inspecting the sewers. Typically, longer sewer pipes have higher deterioration rate because the probability of occurring defects is more in longer pipes. When individual longer sewer pipes (pipe section) are used in sewer networks, the number of joints per unit length of sewer is reduced, therefore, the risk of infiltration, exfiltration and other important defects are decreased. However, most of available sewer inspection inventory are based on the manhole to manhole length of sewer pipes. Lack of appropriate data regarding number of joints or length of pipe section is one of the reasons that the effect of pipe length on deterioration of sewer pipes is contradictory in different studies.

Pipe joints are the main source of infiltration and longer pipes have more points and areas of possible failure specially in joints. Joint defect is one of the common defects in sewer systems and increases the probability of failure. Additionally, longer pipes are more vulnerable to have blockage and sediment deposition which facilitate the deterioration of sewer pipes.

Water Table. Water table is the next significant variable found in binary logistic regression model with Wald = 14.070 (Sig. = 0.000). The results indicated that sanitary sewer pipes are deteriorated faster when the water table is higher around the pipe. The unit of water table is inches in this study and a lower value shows a higher water table. The coefficient of this variable is negative in binary logistic regression equation, therefore larger numbers (lower water table) decrease the risk of sewer pipes being in poor condition.

The odds ratio of water table is 0.998 which demonstrates that for a unit increase in depth of groundwater, the odds of sewer being in poor condition are multiplied by 0.998. This result shows a 0.2% increase (relative to the odds of sanitary sewer surrounding around 1 in. less water table) when all other conditions are constant. Therefore, increasing the depth of water table from 10 in. to 50 in., would reduce the odds of being in poor condition to $(0.998)^{40} = 0.923$ which is 7.7% over the 40 in. increase. The effects of water table on condition of PVC and VCP pipes is shown in Figures 6-9 and 6-10.

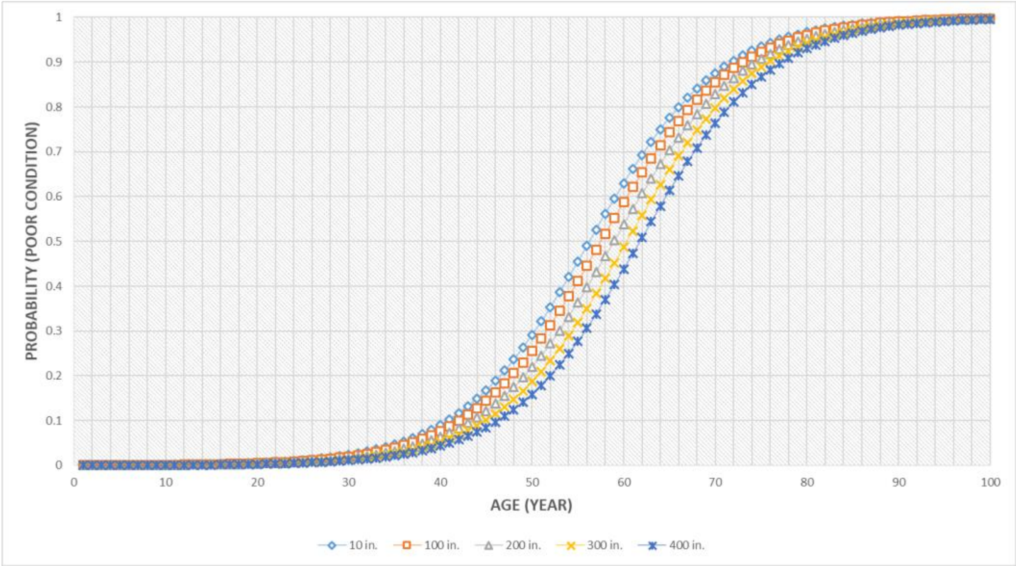


Figure 6-9 Deterioration Curve for PVC Pipes with Different Water Table Depth

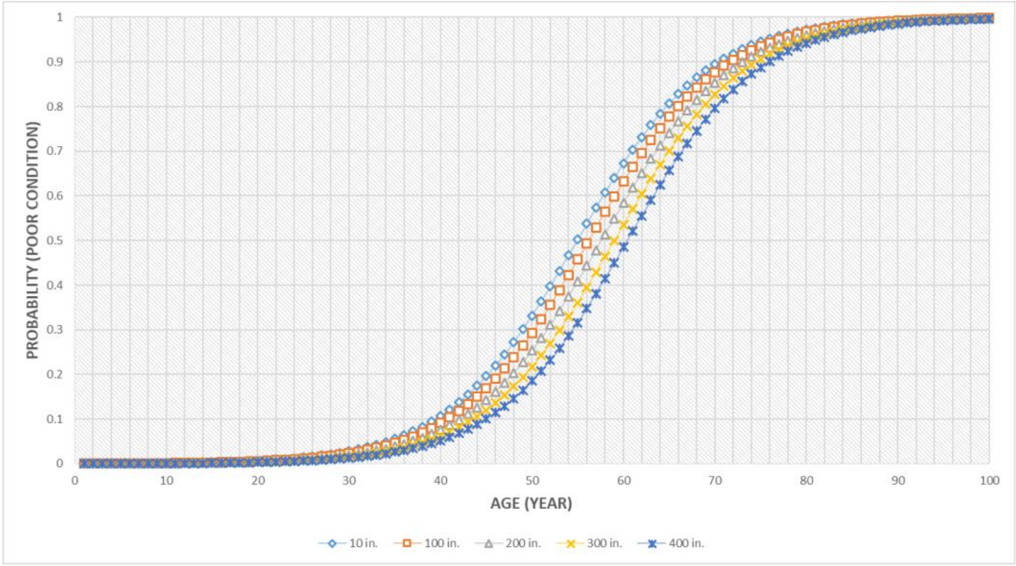


Figure 6-10 Deterioration Curve for VCP Pipes with Different Water Table Depth

The effect of water table on condition of sanitary sewer pipes is more evident in Figure 6-11 which illustrates the water table height from 0 to 500 inches for a PVC and VCP pipe with constant age of 50 years.

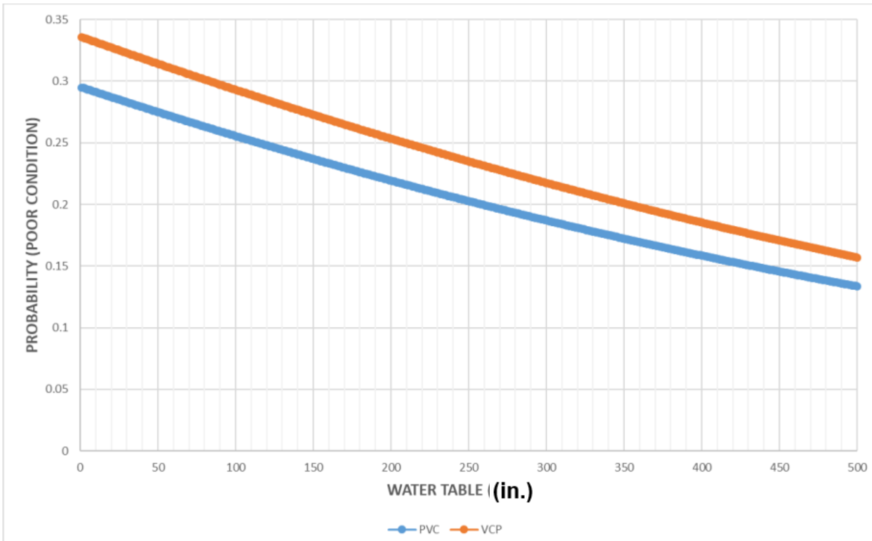


Figure 6-11 Effect of Water Table on Condition of a 50-year Old Pipe

(Zero means water table is at the ground surface)

The availability of water table at or above sewer pipelines may cause water flowing through the pipe due to possible joints and cracks, increasing the structural defects, formation of void and loss of sewer support. Raising the water table level may cause a reduction in the soil cohesive strength and growing the void around the pipe. Consequently, supporting soil can be washed (loosed) easily and the pipe is more likely to collapse in this condition. The above results support the finding of few studies investigated the effect of water table on deterioration of sewer pipes presented in section 2.7.3.2.

Pipe Material. Sanitary sewer pipe material was also found significant variable affecting condition of pipes with Wald = 8.418 (Sig. = 0.015). The result of binary logistic regression showed a moderate difference in deterioration of PVC and VCP pipes. As illustrated in deterioration curves VCP pipes are most likely to deteriorate than PVC pipes. The odds of PVC pipes being in poor condition are only -0.189 in comparison to the odds of VCP pipes.

Sewer pipes constructed with different material have different reaction to the environmental factors, such as soil type and water table. For example, concrete pipes are highly resistant to abrasion and clay pipes act very well against acids. Plastic pipes, such as PVC or HDPE, resist to acidic and alkaline wastes, however they can suffer excessive deformations under loading (Singh and Adachi, 2013). Pipe material can be used as an independent variable during development of condition prediction models and it is possible to identify whether this variable is significant or insignificant through the results of the model. The above results support the finding of studies presented in section 2.7.2.2.

6.2.3.2 Insignificant Variables

Sulfate. The results of binary logistic regression showed that soil sulfate is an insignificant variable in the model with Wald = 0.014 (Sig.=0.904). Based on the backward stepwise analysis, soil sulfate was the first variable to be dropped from the model (second step). One probable reason for this low significant value might be the frequency of sulfate ranges in Tampa area. Approximately 60% of the pipes were buried in soil areas within 0.02 to 0.05% sulfate content. Additionally, the amount of sulfate in the soil can cause corrosion which is an important reason of failure in steel and reinforced concrete pipes, while in this study only VCP and PVC pipes were used to develop prediction models which are strongly resistant to corrosion.

Flow. Sewage flow was another variable found to be insignificant in binary logistic regression model. This variable was removed from the model on third step with Wald = 0.219 (Sig. = 0.640). The City of Tampa has regular plan for cleaning and maintenance of the sewer pipes and usually they do not have overflow and operational problems. Regular maintenance could be one probable reason that flow is an insignificant variable in this dissertation. Only few studies have investigated the effect of sewage flow on condition of sewer pipes during development of condition prediction models as presented in section 2.7.4.2. Most of the sewers in Tampa dataset have 8 in. diameter and similarity of their flow rate could be one probable reason of insignificancy of this variable.

Pipe Depth. Sanitary sewer depth was also found to be an insignificant variable in binary logistic regression with Wald = 1.005 (Sig. = 0.316). Several factors such as soil type, water table, pipe material, pipe diameter and regulations must be considered to identify the appropriate depth of sewer pipes. The results of investigating the effect of depth on deterioration of sewer pipes is contradictory in different prediction models.

Several factors such as soil type, water table, pipe material, pipe diameter and regulations must be considered to identify the appropriate depth of sewer pipes. The results of investigating the effect of depth on deterioration of sewer pipes is contradictory in different prediction models and the above results support the finding of studies described in section 2.7.2.6. insignificancy of pipe depth is not to say that sewer depth does not affect deterioration of pipes when considered on its own, but in data analysis based on the features of pipe datasets, there may not be a direct relationship between pipe depth and condition level of sewer pipes.

Soil pH. The next insignificant variable in this study is soil pH with Wald = 2.388 (Sig.=0.122). The effect of soil pH was investigated more in water pipe systems. For example, Rajani and Maker (2000) and Doyle et al. (2003) used soil pH as a variable to predict the remaining useful life of water pipeline. The outcome showed that the pH was not a significant factor to generate the model. Based on their results, the pH alone is not a good indicator to predict the condition of pipes and there is no positive relationship between pH and deterioration of pipes in prediction models. The effect of soil pH was presented in section 2.7.3.6.

Pipe Slope. Pipe slope was also found to have no significant effect on the condition of sanitary sewer pipes in Tampa city with Wald = 1.910 (Sig. = 0.167). Based on the geographical information of Tampa, this city is located in a flat area and 88% of sanitary sewer pipes have the slope ranged 0 to 1% (standard deviation = 1.39). It could be one probable reason that pipe slope is not a significant factor in Tampa network. This result is supported by the finding of several studies presented in section 2.7.2.5.

Soil Corrosivity. The results of binary logistic regression showed that soil corrosivity is also insignificant variable in the model. Soil corrosivity is a soil characteristic that increases the probability of external corrosion on pipe surface. The rate of corrosion is highly influenced by the characteristic of the pipe material and surrounding soil around the pipe. Only few studies have investigated the effect of soil corrosivity on deterioration of sewer pipelines as described in section 2.7.3.4. Population of PVC and VCP pipes in this dissertation could be the probable reason that soil corrosivity is an insignificant variable in this model since they are highly resistant to the corrosion.

Soil Type. Soil type was also found to be an insignificant variable in binary logistic regression model. One probable reason might be the frequency of silty gravel and sand in Tampa dataset. As explained in chapter 4, approximately 72% of sanitary sewer pipes were covered by silty gravel and sand soil in Tampa city.

Soil Hydraulic. Soil hydraulic was the last insignificant variable in binary logistic regression model. Soil hydraulic group indicates soil runoff potential and the rate of water transmission through the soil layers. It seems that soil hydraulic is more related to the soil surface and it does not affect the layers closer to pipe. Additionally, during pipe installation the backfill soil is compacted, and soil hydraulic properties might be changed after this process. The effect of soil hydraulic on deterioration of sewer pipes was not investigated in previous studies.

6.3 Gradient Boosting Tree

6.3.1 Validation of the Model

The performance of gradient boosting tree model was evaluated using confusion matrix and ROC curve. The confusion matrix was used to identify the number of pipes that have been correctly or incorrectly predicted in good or poor condition levels. In confusion matrix, for every test samples the actual class is compared to the class that was assigned by the trained classifier. 20% of the data was used to evaluate the performance of the model. Table 6-4 presents the result of confusion matrix for gradient boosting tree.

Table 6-4 Gradient Boosting Tree Confusion Matrix

Observed	Predicted		Percent Correct Predicted
	0	1	
0	2,688	194	93.3%
1	305	766	71.5%
Overall Percentage			87.4%

According to the result of confusion matrix, in overall 87.4% of the pipe conditions were predicted correctly by gradient boosting tree. 93.3% of the pipes in condition states 0 and 71.5% in condition state 1 were estimated correctly which indicates a high accuracy. Table 6-5 presents the result of calculating true positive, true negative, false positive and false negative rates. Table 6-5 Gradient Boosting Tree Model Performance

Rates	Values
True positive rate (TPR)	93.3%
True negative rate (TNR)	71.5%
False positive rate (FPR)	28.5%
False negative rate (FNR)	6.7%

The prediction performance of gradient boosting tree was also evaluated by Receiver Operating Characteristic (ROC) curve. ROC curve is a useful visual tool which is a plot of true positive rate (TPR) and false positive rate (FPR). The area under the ROC curve illustrates the model performance, where perfect models have an area close to 1 and random models have an area close to 0.5. An area under the ROC curve greater than 0.7 demonstrates the model is acceptable (Hosmer et al., 2013). Figure 6-12 illustrates the ROC curve for gradient boosting tree model.

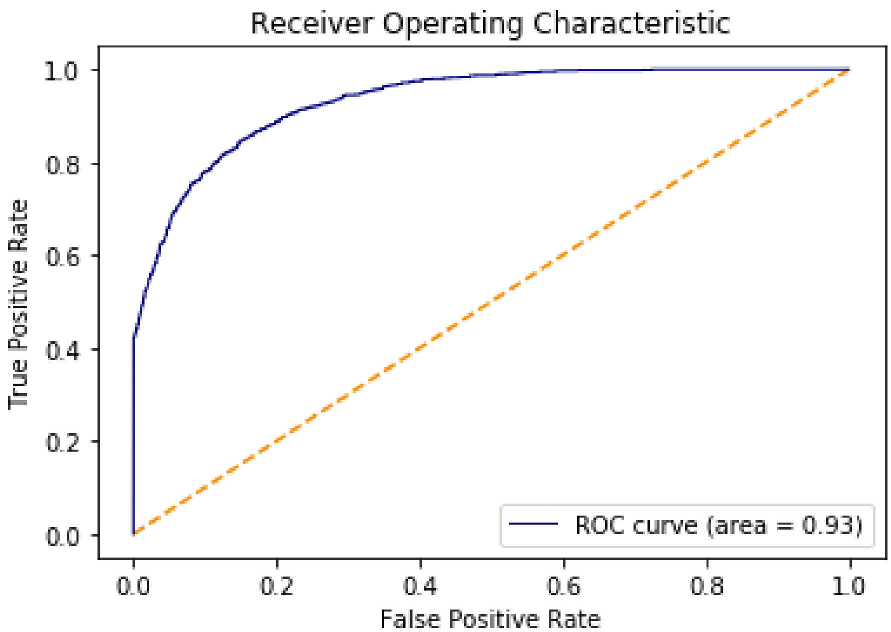


Figure 6-12 Gradient Boosting Tree ROC Curve

The area under ROC curve is 0.93 which shows gradient boosting tree has acceptable result. Therefore, gradient boosting tree model can be used to predict the condition of pipes which have not been inspected yet.

6.3.2 Feature Importance

One benefits of gradient boosting tree models is that they are capable to rank the importance of the independent variables in both regression and classification aims. In general, feature importance provides a score that indicates how useful a variable is in implementation of the model. The importance of variables in sanitary sewer dataset was evaluated based on weight method. This is a metric that presents the number of times a dependent variable is split in the trees of the model. Importance is calculated for each developed tree by the amount that independent variable split points improve the prediction performance of the gradient boosting tree model. For example, if pipe age is split 5, 1 and 4 times in each of tree 1, tree 2 and tree 3 respectively, then the weight for pipe age will be 5+1+4=10. Figure 6-13 illustrates the feature importance in gradient boosting tree model.

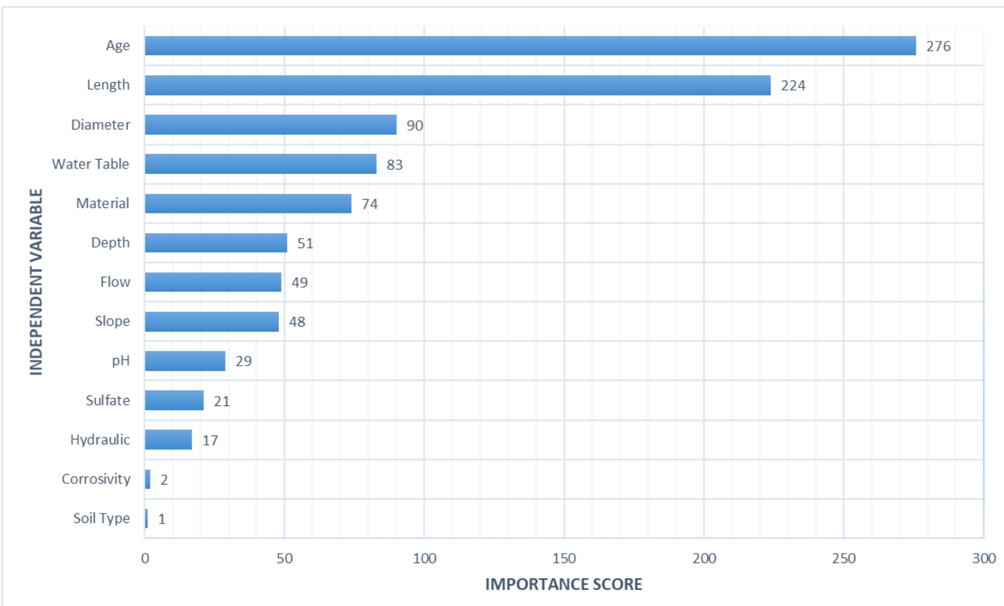


Figure 6-13 Feature Importance in Gradient Boosting Tree Model

According to the results of feature importance, pipe age, length, diameter, water table and material are the most important variables for predicting condition of sanitary sewer pipes in Tampa dataset. Other variables such as, soil type, soil corrosivity, soil hydraulic group, sulfate, and pH have lower prediction power in gradient boosting tree model.

6.3.3 Gradient Boosting Tree Plot

XGBoost algorithm provides a function to plot decision tree based on the importance of independent variables in dataset. This plot shows different layers of decision tree and split decisions of independent variables in the model. The branches and leaves of decision tree provides insight into the role of independent variables on predicting condition of sanitary sewer pipes. As explained before, several trees are created in gradient boosting tree models to find the relationship between variables and predict the target. Figure 6-14 illustrates the role of various independent variables in gradient boosting tree model. This figure shows the first tree that was created in the model.

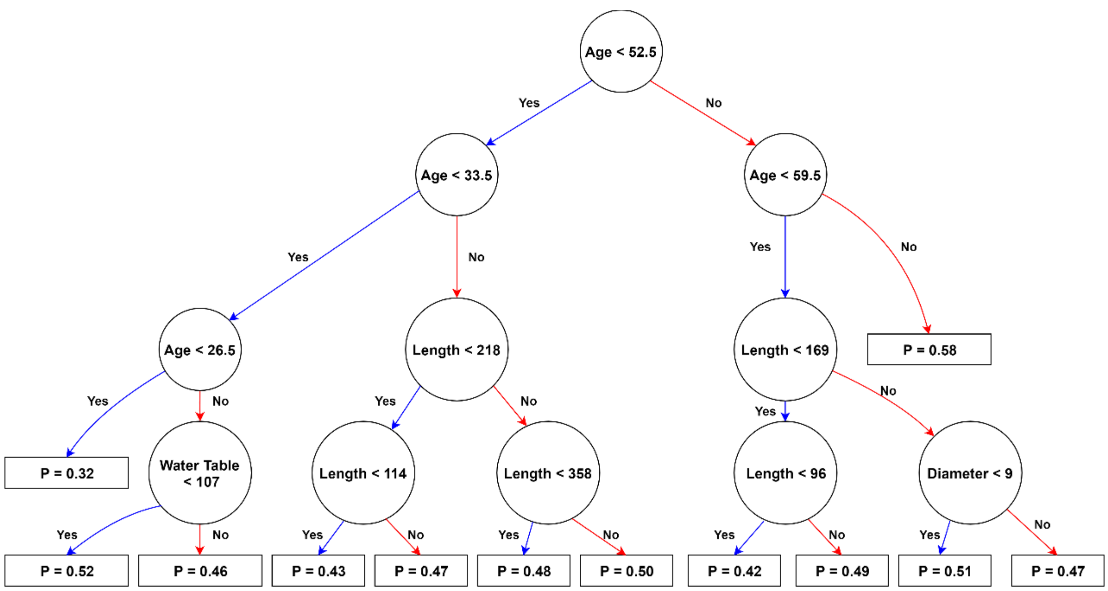


Figure 6-14 Gradient Boosting Tree Plot

The first split of the tree shows the effect of age on condition of pipes. Sanitary sewer pipes are divided into two groups of pipes with age less or more than 52.5 years. The second layer of the tree consists again the age of pipe as an influence variable. In the left node the split point is 33.5, while the right node filters the pipes more or less than 59.5

years old. In the third layer, pipe length is appeared as the second influence variable in the model. In this layer, there is also leaf node in the right side of the plot.

In general, for a binary classification tree with two classes of 0 and 1, the leaf value presents the raw score for class one. This score can be converted to probability using the logistic function (sigmoid function). Therefore, the P value indicates the probability of being in condition 1 or poor condition in plotted tree. The first leaf node shows that sanitary sewer pipes more than 59.5 years have a 58% chance of being in poor condition.

In addition to pipe age and length, forth layer involves water table and pipe diameter. The full tree is easier to interpret than single layers. For example, in the left side, pipes less than 26.5 years old have only 32% chance of being in poor condition. If they are older than 26.5 years, the water table determines the probability. When water table is higher (less than 107 in.), the probability of being in poor condition is 52%, while the lower water table decreases the likelihood to 46%.

In the right side, the condition of pipes depends on diameter when they are less than 59.5 years old with length of more than 169 ft. In this situation, larger pipes (diameter greater than 9 in.) are less probable to be in poor condition with 47% probability than smaller size pipes with 51%. The gradient boosting tree plot can be generated for all the trees created in the model.

The results of gradient boosting tree supported the outcomes of binary logistic regression model. In general, the older pipes had more chance of being in poor condition in both logistic and tree models. Additionally, the probability of being in poor condition is higher in longer pipes like logistic regression results. Water table was also an influence variable in gradient boosting tree model and sanitary sewer pipes are deteriorated faster when the water table is higher around the pipe. Moreover, influence of pipe diameter on pipe condition demonstrated that smaller diameter pipes had more probability of being in poor condition rather than the larger pipes.

6.4 K-Nearest Neighbors

6.4.1 Validation of the Model

The performance of KNN model was evaluated using confusion matrix and ROC curve. In confusion matrix, for every test samples the actual class is compared to the class that was assigned by the trained classifier. 20% of the data was used to evaluate the performance of the model. Table 6-6 presents the result of confusion matrix for KNN.

Table 6-6 Gradient Boosting Tree Confusion Matrix

Observed	Predicted		Percent Correct Predicted
	0	1	
0	2,661	195	93.2%
1	462	635	57.9%
Overall Percentage			83.4%

According to the result of confusion matrix, in overall 83.4% of the pipe conditions were predicted correctly by KNN model. 93.2% of the pipes in condition states 0 and 57.9% in condition state 1 were estimated correctly which indicates a high accuracy. Table 6-7 presents the result of calculating true positive, true negative, false positive and false negative rates for KNN model.

Table 6-7 K-Nearest Neighbors Model Performance

Rates	Values
True positive rate (TPR)	93.2%
True negative rate (TNR)	57.9%
False positive rate (FPR)	42.1%
False negative rate (FNR)	6.8%

The prediction performance of KNN model was also evaluated by Receiver Operating Characteristic (ROC) curve. Figure 6-15 illustrates the ROC curve for K-Nearest Neighbors model.

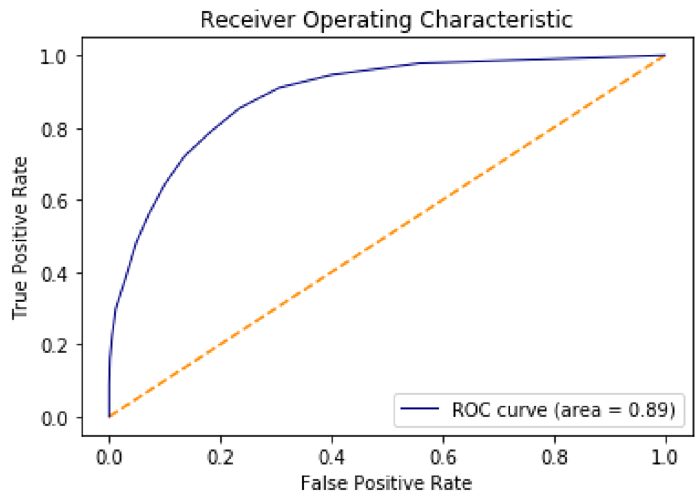


Figure 6-15 K-Nearest Neighbors ROC Curve

The area under ROC curve is 0.89 which shows KNN has acceptable result. Therefore, K-Nearest Neighbors model can be used to predict the condition of pipes which have not been inspected yet.

6.4.2 Feature Importance

Linear models and tree-based models have specific libraries in Python to identify importance of variables, however KNN model does not support this feature. Therefore, Sequential Feature Algorithms (SFAs) were used to automatically select a subset of variables that are most relevant to predict condition of sanitary sewer pipes. The objective of feature selection is to improve the performance of the models by removing the inappropriate variables. In general, sequential feature algorithms are classified into four groups of Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Forward Floating Selection (SFFS) and Sequential Backward Floating Selection (SBFS).

Sequential forward selection was used in this dissertation to identify the important variables in KNN model. This algorithm starts from the empty set and sequentially adds the variables to investigate the performance of the model. Equation 6.8 presents the detail of SFS method.

$Input: Y = \{y_1, y_2, \dots, y_d\}$

$Output: X_k = \{x_i | \, j = 1, 2, \dots, k; \; x_i \in Y\}, where\ k = (0, 1, 2, \dots, d)$

$$x^+ = \arg max \, J(x_k + x), \; where\ x \in Y - X_k \qquad Eq. 6.8$$

$X_k + 1 = X_k + x^+$

$K = K + 1$

where x^+ is additional feature that maximizes the criterion function, d is dimension of the input variable, and k is the number of selected variables. Figure 6-16 illustrates the result of sequential forward selection in KNN model.

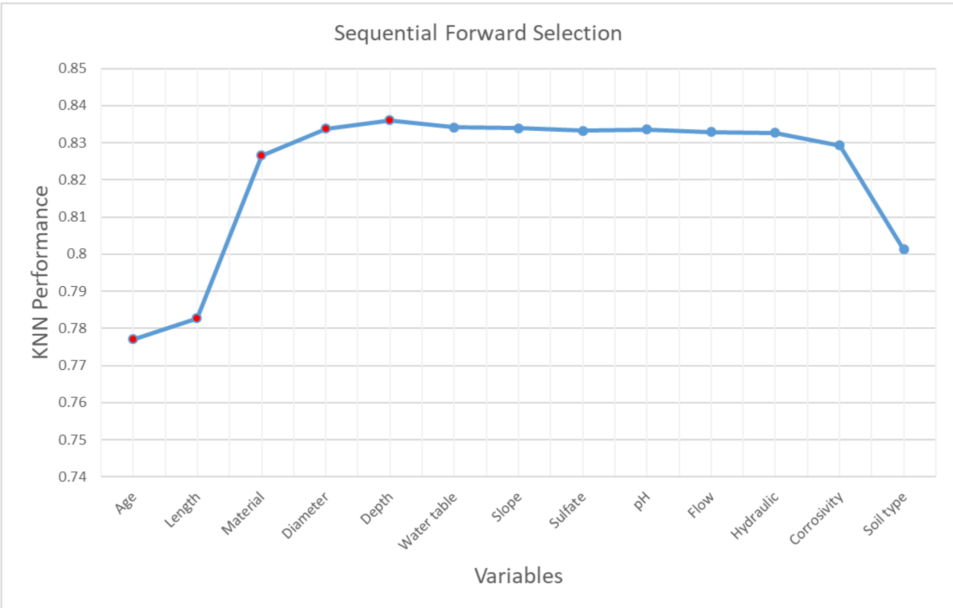


Figure 6-16 KNN Sequential Forward Selection

According to the result of SFS, pipe age, length, material, diameter and depth improved the performance of the model, while other variables were not powerful enough to increase the accuracy. Some variables such as, soil type and corrosivity reduced the performance of the model. Table 6-8 presents the performance of KNN model when variables are added to the model sequentially.

Table 6-8 Effect of Variables on Performance of KNN Model

Variable	Model Performance
Age	0.776966861
Length	0.782743138
Material	0.826573921
Diameter	0.833845959
Depth	0.836035664
Water table	0.834162134
Slope	0.833972429
Sulfate	0.833213608
pH	0.833593018
Flow	0.832960667
Hydraulic	0.832707727
Corrosivity	0.829293031
Soil type	0.801334260

In addition to SFS evaluation, best feature combination algorithm was generated to determine the performance of KNN model when different combination of variables is used to develop the model. Figure 6-17 illustrates the result of best feature combination for KNN model.

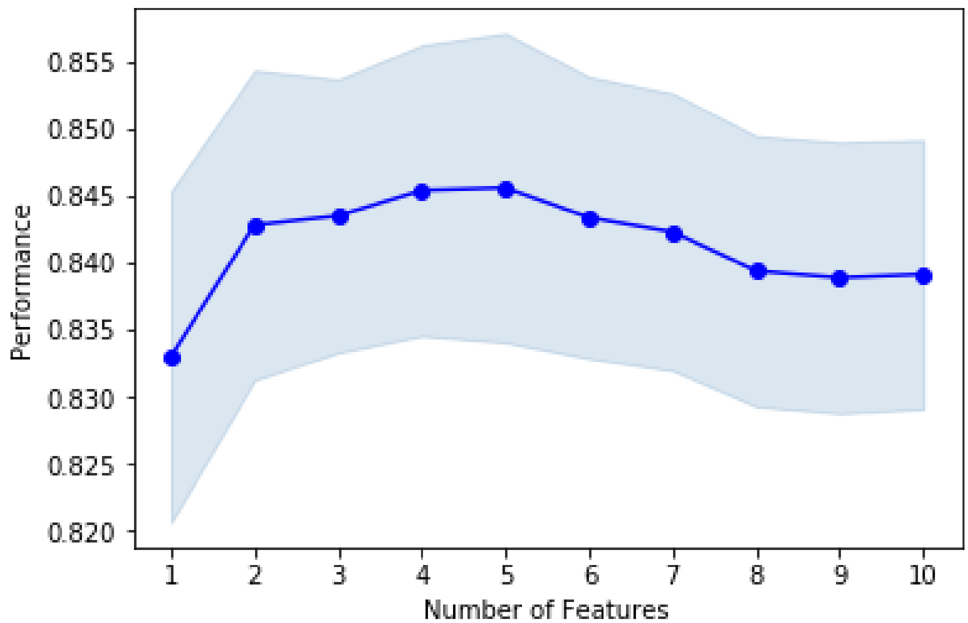


Figure 6-17 KNN Best Feature Combination

The results showed that combination of 5 variables provide the best performance for KNN model. Pipe age, length, material, diameter and depth are variables which provide more powerful prediction results. This finding supports the outcome of sequential forward selection method.

6.5 Discussions

Sanitary sewers as a part of wastewater infrastructure systems, are designed to collect sewage from domestic, industrial, and commercial users and convey to treatment plants. Sewer pipes constitute a major portion of wastewater systems, as they form the pathway between points of wastewater generation and treatment plants. As sewer system become older, the structural and operational performance degrade. The aging of sewer pipes increases the rate of pipe deterioration and failure of sewer pipes can result serious social and environmental impacts.

Maintenance and rehabilitation strategies are important factors to keep the performance of the sewer systems at an acceptable level of service and to provide cost-effective solutions for avoiding unforeseen failures. In the past, repair or rehabilitation of sewer pipes were only done once a pipe collapsed or failed. However, the current trend is to maintain and manage pipe systems before failure time. It is obvious that monitoring and inspection of all sewer pipes is almost impossible due to limited budget, time and assessment technologies. Therefore, more attention is needed to develop pipe deterioration models that can predict the current and future condition of sewer pipelines.

Over the past few years, several statistical and artificial intelligence models were developed to predict condition of sewer pipes, however there is still a high demand to implement more advanced models with more critical input variables. This study developed three different statistical and artificial intelligence models to predict condition level of sanitary sewer pipes based on historical inspection dataset obtained from City of Tampa. The dataset consisted of 19,766 individual pipe segments with different physical and environmental variables. Thirteen independent variables including piped age, material, diameter, flow, length, depth, slope, soil type, soil sulfate, soil pH, water table, soil hydraulic group and soil corrosivity were used to build prediction models. The target variable was condition levels of sanitary sewer pipes which were assessed based on PACP method.

The first model was developed to predict all five condition levels of sanitary sewer pipes, but the result was not enough acceptable. Therefore, the condition levels of pipes were transformed into binary class to investigate whether the pipes are in good or poor conditions. Three different models involving binary logistic regression, gradient boosting tree and KNN were developed to predict condition of sanitary sewer pipes.

Development of statistical models was performed using SPSS software and artificial intelligence models were implemented by Python. Numerous advance techniques such as cross validation and feature importance were used during model development to reduce risk of overfitting and uncertainty. All the models were validated using two or three different validation techniques, such as confusion matrix and ROC curve. Figure 6-18 illustrates the performance of the models used in this study. The multinomial logistic regression obtained the lowest accuracy (65.8%) and gradient boosting tree model showed the best result with 87.4% accuracy.

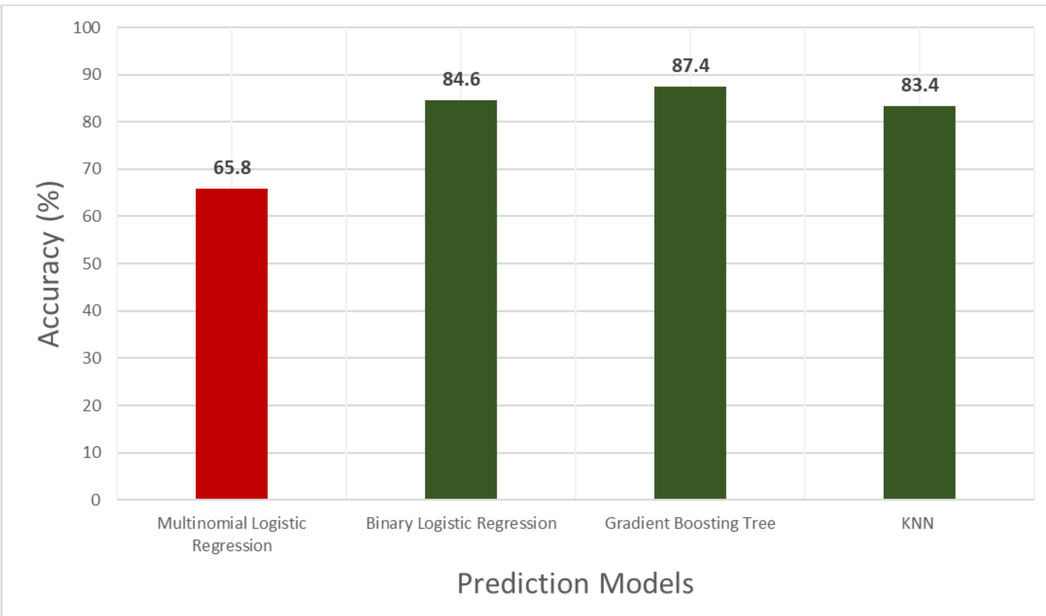


Figure 6-18 Comparison of Models Accuracy

The prediction performance of the models for predicting each condition level (pipes in good condition or poor condition) is shown in Figure 6-19. The results revealed that pipes in condition 0 (good condition) could be predicted better in all three models, while the condition level 1 (poor condition) had different percent correct values. In overall, number of pipes in condition level zero was approximately four times more than pipes in condition level one and that is one probable reason of better prediction results in condition zero.

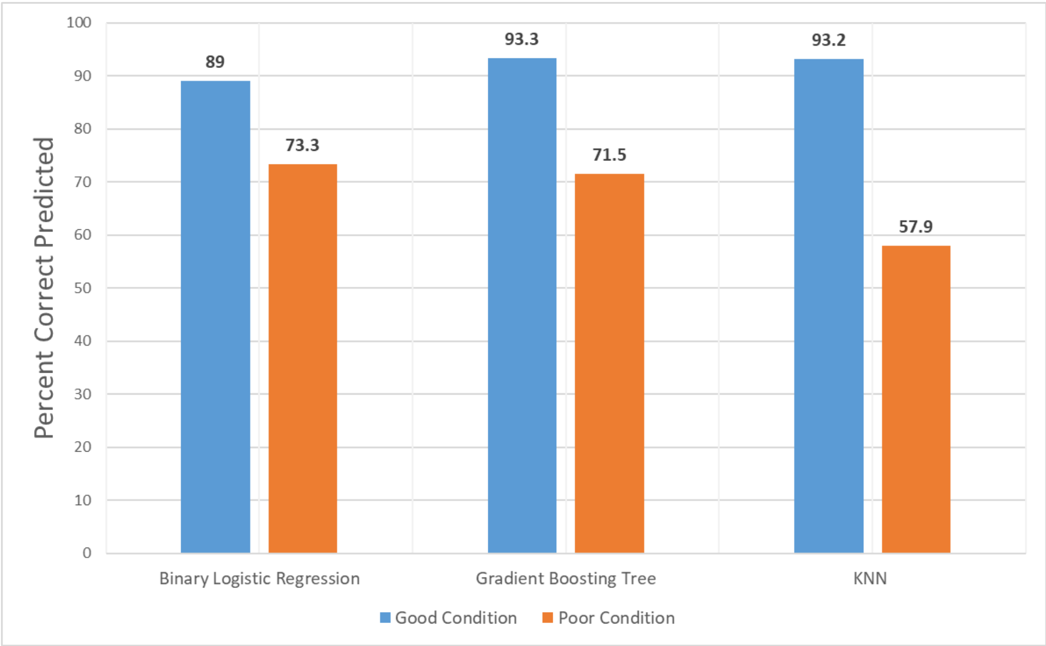


Figure 6-19 Comparison of Models Performance

The influence variables that affect deterioration of sanitary sewer pipes were identified in each model and comprehensively evaluated in logistic regression model. Identifying the influence variables is one important part of condition prediction models. Table 6-9 shows the results of selecting important variables in all developed models in this study. These variables have strong relationship with condition of sanitary sewer pipes and excluding them from the model could decrease the final accuracy. All the models identified five influence variables. Binary logistic regression and gradient boosting tree reflected same results, while KNN identified that pipe depth has more critical role than water table during training and validation of the model.

Table 6-9 Influence Variables Affecting Deterioration of Sewer Pipes

Variables	Binary Logistic Regression	Gradient Boosting Tree	KNN
Age	✓	✓	✓
Material	✓	✓	✓
Diameter	✓	✓	✓
Length	✓	✓	✓
Water table	✓	✓	X
Depth	X	X	✓
Flow	X	X	X
Slope	X	X	X
Soil Type	X	X	X
Soil Sulfate	X	X	X
Soil pH	X	X	X
Soil Hydraulic Group	X	X	X
Soil Corrosivity	X	X	X

6.6 Contribution to the Sewer Pipeline Industry

The results of this dissertation can help utility companies and municipalities to manage and optimize their sewer and stormwater systems. The developed models showed that condition prediction models for sewer pipes can be a part of pipeline asset management. Both statistical and artificial intelligence models are capable to predict future condition of sewer pipes and provide a guideline to monitor the short-term and long-term behavior of pipe network. Additionally, identifying the influence variables in the models can be an important outcome to optimize the useful life of sewer pipes during planning and installation procedures.

6.7 Chapter Summary

This chapter presented the detailed overview of validating sanitary sewer pipes condition prediction models, and the effect of input variables on deterioration of pipes. It was observed that the model displayed a good learning trend towards the facts presented. Both statistical and artificial intelligence models could predict condition of sanitary sewer pipes with more than 80% accuracy. Several advance techniques such as cross validation and feature importance were used during model development to reduce risk of overfitting and uncertainty. Influence variables that affect deterioration of sewer pipes were identified in this chapter to be used for optimizing the useful life of sewer pipes.

Chapter 7 Conclusions and Recommendations for Future Research

7.1 Conclusions

The following conclusions were drawn from the development of logistic regression, gradient boosting and KNN models. The conclusions of each model were provided separately for better understanding the performance of the models.

- Logistic Regression:
  - Application of multinomial logistic regression resulted 65.8% overall accuracy for predicting condition of sanitary sewer pipes, however pipes in condition levels 2, 3 and 4 were not estimated properly in this model. Only 4.4% of pipes in condition 2, 0.6% in condition 3 and 6.6% in condition 4 were predicted correctly. Low number of appropriate pipe datapoints in condition levels 2, 3 and 4 might be the probable reason of low prediction rate. The results of multinomial logistic regression were not enough appropriate to be accepted as a reliable prediction model.
  - Binary logistic regression resulted an overall correct prediction percentage of 84.6% for test dataset. 89% of the pipes in condition level 0 and 73.3% in condition level 1 were estimated correctly which indicates a high accuracy. Therefore, binary logistic regression equation can be used to predict the condition of pipes which have not been inspected yet. Additionally, the area under ROC curve was 0.903 which showed high reliability of the model.
  - Results of binary logistic regression indicated that pipe age, material, diameter, length and water table are significant variables affecting deterioration of sanitary sewer pipes.

- The binary logistic regression results identified that pipe age affects condition of sanitary sewer pipes strongly. The coefficient of pipe age in binary logistic regression equation is positive, and a unit increase in age results in an increase in the probability that the pipe is in poor condition level.
- Pipe diameter was also found to affect deterioration of sanitary sewer pipes largely. According to the binary logistic regression equation, the coefficient of pipe diameter is negative, and a unit increase in pipe size results in a decrease in probability of pipe being in poor condition. Therefore, larger pipes are more resistant to pipe deterioration.
- Sewer manhole to manhole length was also found to be a significant variable in the model. The results of binary logistic regression revealed that as sewer reach increased in length, the probability of pipe being in poor condition increased. The coefficient of pipe length is positive in binary logistic regression equation, therefore longer pipe is deteriorated faster than shorter one.
- The results of binary logistic regression indicated that sanitary sewer pipes are deteriorated faster when the water table is higher around the pipe. The coefficient of this variable is negative in binary logistic regression equation, therefore larger numbers (lower water table) decrease the risk of sewer pipes being in poor condition.
- Logistic regression deterioration curve showed a moderate difference in deterioration of PVC and VCP pipes. PVC pipes seems to degrade slower resulting in a delayed poor condition score. In general, VCP pipes would

have a shorter life than PVC pipes since they have more brittle qualities than PVC.

- Pipe flow, depth, slope, soil type, soil sulfate, soil pH, soil hydraulic group and soil corrosivity were identified insignificant variable in binary logistic regression model.

* Gradient Boosting Tree:
    - Gradient boosting tree model achieved 87.4% overall accuracy for predicting condition of sanitary sewer pipes. 93.3% of the pipes in condition level 0 and 71.5% in condition level 1 were estimated correctly which indicates a high accuracy. Additionally, the area under ROC curve was 0.93 which showed high reliability of the model. Based on the result of gradient boosting tree, it is an acceptable model developed in this study.
    - One benefits of gradient boosting tree models is that they are capable to rank the importance of the independent variables. According to the results of gradient boosting tree model, pipe age, length, diameter, water table and material were the most important variables in this model for predicting condition of sanitary sewer pipes in Tampa dataset.
    - The results of gradient boosting tree revealed that soil type, soil corrosivity, soil hydraulic group, sulfate, and pH have lower prediction power in gradient boosting tree model.
    - The results of gradient boosting tree supported the outcomes of binary logistic regression model. In general, the older pipes had more chance of being in poor condition in both logistic and tree models. Additionally, the probability of being in poor condition is higher in longer pipes like logistic regression results. Water table was also an influence variable in gradient

boosting tree model and sanitary sewer pipes are deteriorated faster when the water table is higher around the pipe. Moreover, influence of pipe diameter on pipe condition demonstrated that smaller diameter pipes had more probability of being in poor condition rather than the larger pipes.

- K-Nearest Neighbors
    - K-Nearest Neighbors resulted an overall correct prediction percentage of 83.4% for test dataset. 93.2% of the pipes in condition states 0 and 57.9% in condition state 1 were estimated correctly which indicates a good accuracy. The area under ROC curve was 0.89 which shows KNN has acceptable result. Therefore, K-Nearest Neighbors model can be used to predict the condition of pipes which have not been inspected yet.
    - KNN model does not support feature importance and sequential forward selection was used to identify the important variables in this model. According to the result of SFS, pipe age, length, material, diameter and depth improved the performance of the model, while other variables were not powerful enough to increase the accuracy.
    - Best feature combination algorithm was generated to determine the performance of KNN model when different combination of variables is used to develop the model. The results showed that combination of 5 variables provide the best performance for KNN model. Pipe age, length, material, diameter and depth are variables which provide more powerful prediction results. This finding supports the outcome of sequential forward selection method.

7.2 Contribution to the Body of Knowledge

The major contributions of this study are:

- Several statistical models have been developed in previous studies to predict the future condition of sewer pipelines. This study was added the diversity of models for used of artificial intelligence models, such as, gradient boosting trees and k-nearest neighbors (k-NN) to investigate deterioration behavior of sewer pipes.
- In this study some independent variables, such as soil hydraulic groups and soil pH were used to develop the prediction models and investigate the significant factors. These important environmental factors have not been used in previous studies to assess the deterioration of sewer pipes.

7.3 Limitations of this Research

As indicated previously, this research is undertaken mainly to demonstrate the possibility of using statistical and artificial intelligence models to predict future condition of sanitary sewer pipes. The main limitation of condition prediction models is availability of appropriate dataset to generate the models. Environmental parameters affecting condition of sanitary sewer pipes, such as bedding material, overburden pressure, soil water content, traffic flow and other factors identified in the literature were omitted due to lack of proper dataset. Additionally, pipe length was manhole to manhole length of sewer pipe segments in Tampa dataset. Lack of information regarding number of joints or length of pipe section was the other important limitation of this study. In other hand, population of sanitary sewer pipes in condition levels 2, 3 and 4 caused transforming the target variable to binary classes. Development of the models with all five condition levels could provide more effective results during development of the models.

7.4 Recommendations for Future Research

Additional research can build upon the work presented in this dissertation. Areas of potential future development include:

- The deterioration models developed in this dissertation can be improved by utilization of other independent variables, such as backfill type, bedding material, soil moisture, overburden pressure, installation method, pipe shape, previous maintenance, overflow and blockage history.
- Only PVC and VCP pipes were used to develop prediction models in this dissertation. An important component of future research is investigating the behavior of more pipe material such as steel and concrete pipes in sewer network and compare the results.
- Most of the prediction models developed in previous studies employed manhole to manhole length of sewer pipes to implement the models. Considering number of joints or length of pipe sections as an independent variable can provide better understanding about the effect of pipe length on deterioration of sewer pipes.
- Logistic regression developed in this study provided the probability of pipes being in poor condition level. One important potential research is considering consequence of pipe failure to develop risk assessment tools for sanitary sewer pipes.
- Prediction models developed in this dissertation can be used to make inspection timeline for sanitary sewer pipes. A cost-benefit analysis can be implemented to investigate the potential cost saving of condition prediction models toward regular yearly inspection plans.
- A particularly important component of future work is further investigation of deep learning algorithms to develop condition prediction models.
- Condition prediction models can be used to forecast the condition of lined or repaired pipeline to investigate the performance of lining material.

- If enough inspection dataset is available, an important component of future work can be comparing the results of prediction models developed for different cities.

References

A glossary of computing terms. (1998). Longman, on behalf of the British Computer Society, Harlow, Essex, England.

Abraham, D. M., Wirahadikusumah, R., Short, T. J., and Shahbahrami, S. (1998). "Optimization Modeling for Sewer Network Management." *Journal of Construction Engineering and Management*, 124(5), 402–410.

Agresti, A. (2007). An introduction to categorical data analysis. 3rd Ed., Wiley, Hoboken, New Jersey.

Al-Barqawi, H., and Zayed, T. (2006). "Condition rating model for underground infrastructure sustainable water mains." *Journal of Performance Constructed Facilities*, 20(2), 126-135.

Allouche, E. N., and Freure, P. (2002). "Management and maintenance practices of storm and sanitary sewers in Canadian municipalities." Institute of Catastrophic Loss Reduction (ICLR) Research Paper Series No. 18, Dept. of Civil and Environmental Engineering, Univ. of Western Ontario-Geotechnical Research Centre, London, Ontario, Canada.

Altarabsheh, A. G. (2016). "Managing Urban Wastewater System Using Complex Adaptive System Approach." Doctoral Dissertation. Purdue University. IN, USA.

American Water Works Association (AWWA), (2012). Buried no longer: confronting America's water infrastructure challenge, Boulder, CO, Retrieved from: https://www.awwa.org/Portals/0/files/legreg/documents/BuriedNoLonger.pdf, July 2018.

Ana, E., Bauwens, W., Pessemier, M., Thoeye, C., Smolders, S., Boonen, I., and Gueldre, G. D. (2009). "An investigation of the factors influencing sewer structural deterioration." *Urban Water Journal*, 6(4), 303–312.

Ariaratnam, S. T., El-Assaly, A., and Yang, Y. (2001). "Assessment of Infrastructure Inspection Needs Using Logistic Models." *Journal of Infrastructure Systems*, 7(4), 160–165.

ASCE (American Society of Civil Engineers) (2017). Report Card for America's Infrastructure. ASCE, Reston, VA, USA. Retrieved from: https://www.infrastructurereportcard.org/ , March 2018.

Ashoori, T., Najafi, M., Tabesh, A., Korky, S. M. S. J., and Malek Mohammadi, M. (2017). "Use of Cut-and-Cover Method for Pipeline Installation in Underground Freight Transportation." Pipelines 2017, ASCE, Phoenix, Arizona, August 6–9, 2017.

Atique, F., (2016). "Analysis of urban pipe deterioration using Copula method." Doctoral Dissertation. University of Delaware. Newark, DE, USA.

Ayoub, G. M., et al., 2004. "Assessment of hydrogen sulphide corrosion of cementitious sewer pipes: a case study." *Urban Water Journal*, 1 (1), 39–53.

Baik, H. S., Jeong, H. S., and Abraham, D. M. (2006). "Estimating Transition Probabilities in Markov Chain-Based Deterioration Models for Management of Wastewater Systems." *Journal of Water Resources Planning and Management*, 132(1), 15–24.

Bakry, I., Alzraiee, H., Kaddoura, K., Masry, M. E., and Zayed, T. (2016a). "Condition Prediction for Chemical Grouting Rehabilitation of Sewer Networks." *Journal of Performance of Constructed Facilities*, 30(6), 04016042.

Bakry, I., Alzraiee, H., Masry, M. E., Kaddoura, K., and Zayed, T. (2016b). "Condition Prediction for Cured-in-Place Pipe Rehabilitation of Sewer Mains." *Journal of Performance of Constructed Facilities*, 30(5), 04016016.

Baur, R., and Herz, R. (2002). "Selective inspection planning with ageing forecast for sewer types." *Journal of Water Science and Technol*ogy, 46(6-7), 389–396.

Bergman, M. J., Green, W., and Donnangelo, L. J. (2002). "Calibration of Storm Loads in The South Prong Watershed, Florida, Using Basins." *Journal of the American Water Resources Association*, 38(5), 1423–1436.

Bhattarai, J. (2013). "Study on the Corrosive Nature of Soil Towards the Buried-Structures." *Journal of Scientific World*, 11(11), 43–47.

Biau, G., and Scornet, E. (2016). "A random forest guided tour." *Test*, 25(2), 197–227.

Bishop, C. (2016). Pattern Recognition and Machine Learning. Springer, New York, U.S.

Burian, S. J., Pitt, R., Nix, S. J., and Durrans, S. R. (2000). "Urban Wastewater Management in the United States: Past, Present, and Future." *Journal of Urban Technology*, 33–62.

Burian, S. J., Streit, G. E., Mcpherson, T. N., Brown, M. J., and Turin, H. (2001). "Modeling the atmospheric deposition and stormwater washoff of nitrogen compounds." *Journal of Environmental Modelling & Software*, 16(5), 467–479.

Chae, M. J. and Abraham, D. M. (2001). “Neuro-fuzzy approaches for sanitary sewer pipeline condition assessment.” *Journal of Computing in Civil Engineering*, 15(1), 4 – 14.

Charniak, E. and McDermott, D. (1985). Introduction to Artificial Intelligence. Addison-Wesley, Reading, MA.

Chatterjee, S., and Simonoff, J. S. (2013). Handbook of regression analysis. John Wiley & Sons, Hoboken, NJ.

Chughtai, F. and Zayed, T. (2007a). “Sewer pipeline operational condition prediction using multiple regression.” *Journal of Pipelines*, Reston, VA., July 8-11, 2007.

Chughtai, F. R. M. (2008). “Integrated condition assessment models for sustainable sewer pipelines.” Thesis, Dept. of Building, Civil and Environmental Engineering, Concordia University, Montreal, Canada.

Chughtai, F., and Zayed, T. (2007b). “Structural condition models for sewer pipeline.” *In Proceedings of Pipelines 2007: Advances and Experiences with Trenchless Pipeline Projects*, Reston, VA., July 8-11, 2007.

Chughtai, F., and Zayed, T. (2011). “Integrating WRc and CERIU Condition Assessment Models and Classification Protocols for Sewer Pipelines.” *Journal of Infrastructure Systems*, 17(3), 129–136.

Chughtai, F., Zayed, T. (2008). “Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines.” *Journal of Constructed Facilities*, 22 (5),333-341.

Coles, S. (2011). An introduction to statistical modeling of extreme values. Springer, London, 978-1-84996-874-4.

Dasu, T., and Johnson, T. (2003). Exploratory data mining and data cleaning. John Wiley & Sons, Hoboken, NJ.

Davies, J. P., Clarke, B. A., Whiter, J. T., Cunningham, R. J. (2001). "Factors influencing the structural deterioration and collapse of rigid sewer pipes." *Journal of Urban Water* 3 73-89.

Doyle, G., Seica, M.V., Grabinsky, M.W. (2003). “The role of soil in the external corrosion of cast iron water mains in Toronto, Canada.” *Geotechnical Journal*, 40(2): 225-236.

EPA (2002). “The Clean Water and Drinking Water Infrastructure Gap Analysis.” EPA-816-R-02-020. Washington DC: US Environmental Protection Agency, Office of Water. August 2018, <www. Ep.gov/nrmrl>.

EPA, (2004), “Report to Congress; Impacts and Control of CSOs and SSOs.” August 2018, <www. Ep.gov/nrmrl>.

EPA, (2007), “Innovation and Research for Water Infrastructure for the 21st Century.” April 2018, <www. Ep.gov/nrmrl>.

EPA, (2009), “Condition Assessment of Watewater Collection Systems.” March 2018, <www. Ep.gov/nrmrl>.

EPA, (2010a), “State of Technology for Rehabilitation of Wastewater Collection Systems.” July 2018, <www. Ep.gov/nrmrl>.

EPA, (2010b), “Report on Condition Assessment of Wastewater Collection Systems.” July 2018, <www. Ep.gov/nrmrl>.

EPA, (2014), “Demonstration of Innovative Sewer System Inspection Technology: SL-RAT™.” Jun 2018, <www. Ep.gov/nrmrl>.

EPA, (2015), “Condition Assessment of Underground Pipes.” April 2018, <www. Ep.gov/nrmrl>.

Flintsch, G. W., and Chen, C. (2004). “Soft Computing Applications in Infrastructure Management.” *Journal of Infrastructure Systems*, 10(4), 157–166.

Friedman, J. (2001). “Greedy function approximation: A gradient boosting machine.” *Annals of Statistics*, 29 (2001), No. 5, 1189--1232.

Gat, Y. L. (2008). “Modelling the deterioration process of drainage pipelines.” *Urban Water Journal*, 5(2), 97–106.

Gedam, A., Mangulkar, S., and Gandhi, B. (2016). “Prediction of Sewer Pipe Main Condition Using the Linear Regression Approach.” *Journal of Geoscience and Environment Protection*, 04(05), 100–105.

Gheytaspour, M., and Habibzadeh Bigdarvish, O., (2018). “Forecasting Oxygen Demand in Treatment Plant Using Artificial Neural Networks.” *International Journal of Advanced Engineering Research and Science*, 5(3), 050-057.

Habibzadeh-Bigdarvish, O., Yu, X., Lei, G., Li, T., and Puppala, A. J. (2019). "Life-Cycle Cost-benefit Analysis of Bridge Deck De-icing using Geothermal Heat Pump System: A Case Study of North Texas." *Sustainable Cities and Society*, 101492.

Hagan, M. T., Demuth, H. B., Beale, M. H., and Jesús Orlando De., (2016). Neural network design. PWS Publishing, Boston.

Harrell, F., (2016). Regression Modeling Strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer, London.

Harvey, R. R., and Mcbean, E. A. (2014a). "Predicting the structural condition of individual sanitary sewer pipes with random forests." *Canadian Journal of Civil Engineering*, 41(4), 294–303.

Harvey, R. R., and Mcbean, E. A. (2014b). "Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure." *Journal of Hydro informatics*, 16(6), 1265–1279.

Hastie, T., Tibshirani, R., Friedman, J., (2017) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, London.

Hawari, A., Alkadour, F., Elmasry, M. and Zayed, T., (2016), "Simulation-Based Condition Assessment Model for Sewer Pipelines", *Journal of Performance of Constructed Facilities*, 04016066.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied logistic regression. Wiley, Hoboken, NJ.

Hou, Y., Deqing, L., Li, S., Yang, W., Li, C. (2016), "Experimental investigation on corrosion effect on mechanical properties of buried metal pipes." *International Journal of Corrosion*, vol. 2016, Article ID 5808372, 13 pages, 2016.

International Society for Trenchless Technology (ISTT). (1990). Trenchless technology guidelines, ISTT, London, UK.

Jalalediny Korky, S., Aaron, H., Najafi, M., Malek Mohammadi, M., Serajiantehrani, R. (2019). "Design and Installation of Water Pipeline Renewals.", North American Society for Trenchless Technology (NASTT) NASTT's 2019 No-Dig Show, Chicago, Illinois, U.S.

Jalalediny Korky, S., Najafi, M., Malek Mohammadi, M., Ashoori, T., Tabesh, A., (2017). "An Analysis of Soil Hauling Operations for Tunneling and Pipeline Installation." ASCE Pipeline Conference, 2019, Nashville, TN, U.S.

Jeong, H. S., Baik, H. S., and Abraham, D. M. (2005). “An ordered probit model approach for developing Markov chain-based deterioration model for wastewater infrastructure systems.” *Proc. Pipelines 2005: Optimizing Pipeline Design, Operations, and Maintenance in Today’s Economy*, Reston, VA., 649 - 661.

Kabir, G., Balek, N. B. C., and Tesfamariam, S. (2018). “Sewer Structural Condition Prediction Integrating Bayesian Model Averaging with Logistic Regression.” *Journal of Performance of Constructed Facilities*, 32(3), 04018019.

Kaushal, V., and Guleria, S. P. (2015). “Geotechnical Investigation of Black Cotton Soils.” *International Journal of Advances in Engineering Sciences*, Vol. 5, Issue 2, pp. 15-22.

Kaushal, V., Najafi, M. and Love, J. (2018). "Qualitative Investigation of Microbially Induced Corrosion of Concrete in Sanitary Sewer Pipe and Manholes." Proc. ASCE Pipelines 2018, Toronto, Canada, pp. 768-775.

Kawabata, T., Uchida, K., Ariyoshi, M., Nakase, H., Mohri, Y., and Ling, H. I. (2003). “D.E.M. Analysis on Behavior of Shallowly Buried Pipe Subject to Traffic Loads.” *New Pipeline Technologies, Security, and Safety*.

Khan, Z., Zayed, T., and Moselhi, O. (2010). “Structural Condition Assessment of Sewer Pipelines.” *Journal of Performance of Constructed Facilities*, 24(2), 170–179.

Khazraeializadeh, S. (2012). “A comparative analysis on sewer structural condition grading systems using four sewer condition assessment protocols.” Master’s thesis, Edmonton: Faculty of Graduate Studies and Research, The University of Alberta, Canada.

Kleiner, Y., Sadiq, R., and Rajani, B. (2004). “Modeling failure risk in buried pipes using fuzzy Markov deterioration process.” ASCE International Conference on Pipeline Engineering and Construction, San Diego, CA., 1-12.

Kley, G., and Caradot, N. (2013). “D1.2. REVIEW OF SEWER DETERIORATION MODELS.” KWB report, project SEMA, Berlin, Germany.

Koehn, P. (1994). “Combining genetic algorithms and neural networks: the encoding problem.”, Master’s thesis, University of Erlangen and The University of Tennessee, Knoxville, 1994.

Konstantopoulos, T. (2009). Markov Chains and Random Walks. Springer, New York.

Koo, D. H., and Ariaratnam, S. T. (2006). “Innovative method for assessment of underground sewer pipe condition.” *Automation in Construction*, 15(4), 479–488.

Kozak, J. (2019). Decision tree and ensemble learning based on ant colony optimization. Springer, Cham.

Kramer, O. (2016). Dimensionality reduction with unsupervised nearest neighbors. Springer, London.

Kulandaivel, G. (2004). “Sewer pipeline condition prediction using neural network models.” Master’s thesis, Michigan State University, USA, 2004.

Kumar, S. S., Abraham, D. M., Jahanshahi, M. R., Iseley, T., and Starr, J. (2018). “Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks.” *Journal of Automation in Construction*, 91, 273–283.

Kurzweil, R. (1990). The Age of Intelligent Machines. Zone Books, New York.

Laakso, T., Kokkonen, T., Mellin, I., and Vahala, R. (2018). “Sewer Condition Prediction and Analysis of Explanatory Factors.” *Journal of Water*, 10(9), 1239.

Law, T. C. M. and Moore, I. D. (2007). “Numerical modeling of tight-fitting flexible liner in damaged sewer under earth loads.” *Tunneling and Underground Space Technology*, Vol. 22, pp. 655–665.

Lewis, N. D. C. (2016). Deep learning made easy with R: a gentle introduction for data science. CreateSpace Independent Publishing Platform, England.

Lindner, R. (2008). “Effectively Managing Sewer Pipeline Infrastructure.” *Journal of Underground Infrastructure Management*, December 2008, 23-25.

Lubini, A. T., and Fuamba, M. (2011). “Modeling of the deterioration timeline of sewer systems.”, *Canadian Journal of Civil Engineering*, 2011, 38(12): 1381-1390.

Luger, G. F. (2009). Artificial Intelligence: structures and Strategies for Complex Problem Solving. Addison Wesley, Harlow, England.

Madanat, S., and Ibrahim, W. H. W. (1995). “Poisson Regression Models of Infrastructure Transition Probabilities.” *Journal of Transportation Engineering*, 121(3), 267–272.

Mailhot, A., Duchesne, S., Musso, E., and Villeneuve, J. P. (2000). “Modélisation de l’évolution de l’état structural des réseaux d’égout: application à une municipalité du Québec.” *Canadian Journal of Civil Engineering*, 27(1): 65–72. doi:10.1139/l99-054.

Malek Mohammadi, M., Aaron, H., Najafi, M., Jalalediny Korky, S., Serajiantehrani, R. (2019). “Development of a Framework for Design and Installation of a New Lining Material for Water Pipeline Renewal,” ASCE Pipeline Conference, 2019, Nashville, TN, U.S.

Malek Mohammadi, M., Najafi, M., Tabesh, A., Riley, J. and Gruber, J. (2019). “Condition Prediction of Sanitary Sewer Pipes,” ASCE Pipeline Conference, 2019, Nashville, TN, U.S.

Mashford, J., Marlow, D., Tran, D., and May, R. (2011). “Prediction of Sewer Condition Grade Using Support Vector Machines.” *Journal of Computing in Civil Engineering*, 25(4), 283–290.

McDonald, J. H. (2009). Handbook of biological statistics. Sparkey House, Baltimore.

McDonald, S. E., Zhao, J. Q. (2001). “Condition Assessment and Rehabilitation of Large Sewers.” Proceedings of International Conference on Underground Infrastructure Research, University of Waterloo, Ontario, 10–13 June 2001, pp. 361–369.

Melosi, M. V. (2000). “The Sanitary City: Urban Infrastructure in America from Colonial Times to the Present.”, *The Johns Hopkins University Press*, Volume 76, Number 2, Summer 2002.

Meyers, L. S., Gamst, G., and Guarino, A. J. (2017). Applied multivariate research: design and interpretation. Thousand Oaks, CA: Sage Publication.

Micevski, T., Kuczera, G., and Coombes, P. (2002). “Markov Model for Storm Water Pipe Deterioration.” *Journal of Infrastructure Systems*, 8(2), 49–56.

Modica, G.R., (2007). The history of the Newark sewer system. Cranbury, Richard Glubb & Associates, Inc, New Jersey.

Moore, I. D. (2008) “Assessment of Damage to Rigid Sewer Pipes and Erosion Voids in The Soil, And Implications for Design of Liners.”, North American Society for Trenchless Technology, No-Dig Conference & Exhibition, Dallas, Texas, April 27 – May 2.

Morcous, G., Rivard, H., and Hanna, A. M. (2002). “Modeling Bridge Deterioration Using Case-based Reasoning.” *Journal of Infrastructure Systems*, 8(3), 86–95.

Moteleb, M. (2010). “Risk based decision making tools for sewer infrastructure management.” Doctoral Dissertation. University of Cincinnati. OH, USA.

Najafi, M., and Gokhale, S. B. (2005). Trenchless technology: pipeline and utility design, construction, and renewal. McGraw-Hill, New York.

Najafi, M., and Kulandaivel, G. (2005). “Pipeline Condition Prediction Using Neural Network Models.” Proceedings of the American Society of Civil Engineers (ASCE) Pipeline Division Specialty Conference, Houston, TX, USA.

Najafi, M. (2016). Pipeline infrastructure renewal and asset management. McGraw-Hill Education, New York.

National Association of Sewer Service Companies (NASSCO) (2015), “Pipeline Assessment Certification Manual.”, Marriottsville, MD, USA, March 2018.

National engineering handbook: Section 7: Part 630, Hydrologic engineering. (2007). U.S. Dept. of Agriculture, Soil Conservation Service, Washington, D.C.

National Research Council Canada (NRC). (2004). “Municipal Infrastructure Planning (MIIP) Report: An Evaluation of Condition Assessment Protocols for Sewer Management.” Report No. B – 5123.6. Ottawa, ON, Canada.

National asset management steering group (2006), “International infrastructure management manual. 3rd. ed”. Association of Local Government Engineering N.Z. Inc., Thames, New Zealand.

Oguzie, E. E., Agochukwu, I. B., Onuchukwu, A. I. (2004). “Monitoring the corrosion susceptibility of mild steel in varied soil textures by corrosion product count technique”. *Materials chemistry and physics*. Vol. 84, (1), pp. 1-6.

Opila, M. C. (2011). “Structural Condition Scoring of Buried Sewer Pipes for Risk-Based Decision Making.” Doctoral Dissertation. University of Delaware. Newark, DE, USA.

O'Reilly, M. P., Prosbrook, R. B., Cox, G. C., McCloskey A, (1989). "Analysis of defects in 180 km of pipe sewers in southern water authority", Transport and Road Research Laboratory (TRRL), RR 172, Wokingham, United Kingdom.

Park, H. A. (2013). “An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain.” *Journal of Korean Academy of Nursing*, 43(2), 154.

Park, T. (2009). “A comprehensive asset management system for sewer infrastructures.” Dissertation, Pennsylvania State University, Pennsylvania.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830.

Pyle, D. (2007). Data preparation for data mining. Morgan Kaufmann Publishers Inc., San Francisco, CA.

Rahman, S. and Vanier, D.J (2004). NRC Report: An Evaluation of Condition Assessment Protocols for Sewer Management, National Research Council Canada (https://nparc.nrc-cnrc.gc.ca/eng/view/object/?id=93ed3e91-be5e-452d-b79d-c20d4ac77002), July 2018.

Rajani, B. and Makar, J. (2000) “A methodology to estimate remaining service life of grey cast iron water mains.” *Canadian Journal of Civil Engineering,* 27(6):1259–1272.

Rawlings, J. O. (1989). Applied regression analysis: a research tool. Wadsworth & Brooks, Cole Advanced Books & Software, Pacific Grove, CA.

Rokach, L., and Maimon, O. (2015). Data mining with decision trees: theory and applications. World Scientific Pub. Co, Singapore.

Russell, S. J., and Norvig, P. (2010). Artificial intelligence: a modern approach. Pearson, Boston.

Salman, B. (2010). “Infrastructure management and deterioration risk assessment of wastewater collection systems.” Doctoral Dissertation. University of Cincinnati. OH, USA.

Salman, B., and Salem, O. (2012). “Modeling Failure of Wastewater Collection Lines Using Various Section-Level Regression Models.” *Journal of Infrastructure Systems*, 18(2), 146–154.

Scheidegger, A., Hug, T., Rieckermann, J., and Maurer, M. (2011). “Network condition simulator for benchmarking sewer deterioration models.” *Journal of Water Research*, 45(16), 4983–4994.

Schulting, F.L. and Alegre, H. (2007). Global developments of strategic asset management. In Reports and Abstracts Book, 2nd Leading Edge Conference on Strategic Asset Management, 17-19 October 2007, Lisbon, Portugal.

Seo, S. (2006). “A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets.” Thesis, Pittsburgh, PA, USA.

Simon, P. (2015). Too big to ignore: the business case for big data. Wiley, Hoboken, NJ.

Singh, A., and Adachi, S. (2013). "Bathtub curves and pipe prioritization based on failure rate." *Built Environment Project and Asset Management*, 3(1), 105–122.

Sinha, S. K., and Mckim, R. A. (2007). "Probabilistic based integrated pipeline management system." *Tunneling and Underground Space Technology*, 22(5-6), 543–552.

Sjogren, L., Camitz, G., Peultier, J., Jacques, S., Baudu, V., Barrau, F., Chareyre, B., Bergquist, A., Pourbaix, A., Carpentiers, P. (2011). "Corrosion resistance of stainless-steel pipes in soil." *Materials and corrosion*. Vol. 62, (4), pp. 299-309.

Sterpi, D., Angelotti, A., Habibzadeh Bigdarvish, O., and Jalili, D. (2017). "On the modelling of thermo-active diaphragm walls based on monitoring data." In 15th International Conference of the International Association for Computer Methods and Advances in Geomechanics, IACMAG (pp. 1-7). CHN.

Sterpi, D., Angelotti, A., Habibzadeh-Bigdarvish, O., & Jalili, D. (2018). "Assessment of thermal behaviour of thermo-active diaphragm walls based on monitoring data." *Journal of Rock Mechanics and Geotechnical Engineering*, 10(6), 1145-1153.

Soil Science Division Staff. (2017). Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Sousa, V., Matos, J. P., and Matias, N. (2014). "Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition." *Automation in Construction*, 44, 84–91.

Spasojevic, A., Mair, R.J., and Gumbel, J.E. (2007) "Centrifuge modelling of the effects of soil loading on flexible sewer liners." *Journal of Geotechnique*, Volume 57, No. 4, pp. 331-341.

Syachrani, S. (2010). "Advanced sewer asset management using dynamic deterioration models." Doctoral dissertation, Oklahoma State University, Oklahoma, U.S

Syachrani, S., Jeong, H. S. D., and Chung, C. S. (2013). "Decision Tree–Based Deterioration Model for Buried Wastewater Pipelines." *Journal of Performance of Constructed Facilities*, 27(5), 633–645.

Tan, Z. and Moore, I.D. (2007). "Effect of backfill erosion on moments in buried rigid pipes," Transportation Research Board Annual Conference, Washington D.C. January, 29pp.

Temprano, J., Arango, Ó., Cagiao, J., Suárez, J., and Tejero, I. (2007). "Stormwater quality calibration by SWMM: A case study in Northern Spain." *Journal of Water Planning* SA, 32(1).

Thornhill, R and Wildbore, P, (2005), "Sewer Defect Codes: Origin and Destination.", *Journal of Underground Construction*. pp. 32-36.

Tran, D. H., Ng, A. W. M., Perera, B. J. C., Burn, S., and Davis, P. (2006). "Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes." *Urban Water Journal*, 3(3), 175–184.

Tran, D. H., Perera, B. J. C., and Ng, A. (2009). "Comparison of Structural Deterioration Models for Stormwater Drainage Pipes." *Computer-Aided Civil and Infrastructure Engineering*, 24(2), 145–156.

Tran, D. H., Ng, A., and Perera, B. (2007). "Neural networks deterioration models for serviceability condition of buried stormwater pipes." *Engineering Applications of Artificial Intelligence*, 20(8), 1144–1151.

Tran, D. H. (2007). "Investigation of Deterioration Models for Stormwater Pipe Systems." Doctoral Dissertation. Victoria University. Melbourne, Australia.

Vanier, D.J. (2001). "Why industry needs asset management tools." *Journal of Computing in Civil Engineering*, 15(1), 35-43.

Wasim, M., Shoaib, S., Mubarak, N. M., Inamuddin, and Asiri, A. M. (2018). "Factors influencing corrosion of metal pipes in soils." *Environmental Chemistry Letters*, 16(3), 861–879.

Water Research Centre (WRc), Manual of sewer condition 4th Edition classification. (2004). WRc plc, Swindon, Wiltshire, July 2018.

"Web Soil Survey." (2018). Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at the following link: https://websoilsurvey.sc.egov.usda.gov/. Accessed (April/10/2018).

WEF/ASCE. (2009). "Existing sewer evaluation and rehabilitation." WEF Manual of Practice No. FD-6 and ASCE Manuals and Reports on Engineering Practice No.62, WEF, Alexandria, VA; ASCE, Reston, VA, March 2018.

Wirahadikusumah, R., Abraham, D., and Iseley, T. (2001). "Challenging issues in modeling deterioration of combined sewers." *Journal of Infrastructure System*, 10.1061/(ASCE)1076-0342(2001)7:2(77), 77–84.

WRC, 1983 & 1986, Sewerage Rehabilitation Manual, Water Research Center, UK, London, July 2018.

WSAA, (2008), Sewer Inspection Reporting Code of Australia, Water Service Association of Australia (WSAA), Melbourne, June 2018.

Yajima, A., (2015). “Assessment of soil corrosion in underground pipelines via statistical Inference.” Doctor of Philosophy, University of Akron, Ohio, U.S.

Yan, J., and Vairavamoorthy, K. (2003). “Fuzzy approach for pipe condition assessment.” *Proc., New Pipeline Technologies, Security, and Safety, ASCE*, Reston, Va., 466–476.

Yang, J. (2004). “Road crack condition performance modeling using recurrent Markov chains and artificial neural networks.” Doctoral Dissertation. Department of Civil and Environmental Engineering, College of Engineering, University of South Florida, U.S.

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., and Si, Y. (2018). “A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost.” *IEEE Access*, 21020–21031, 2169-3536.

Appendix A

Abbreviations

3D – Three Dimensional

AC – Asbestos-cement

AET – Acoustic Emission Testing

AI – Artificial Intelligence

ASCE – American Society of Civil Engineering

AUC – Area Under Curve

AWWA – American Water Works Association

CAS – Cast Iron

CCTV – Closed-Circuit Television

CIPP – Cured-in-place Pipe

CMP – Corrugated Metal Pipe

CSS – Combined Sewer Systems

DIP – Ductile Iron Pipe

ECT – Eddy Current Testing

EPA – Environmental Protection Agency

FANP – Fuzzy Analytical Network Process

FHWA – Federal Highway Administration

FN – False Negative

FP – False Positive

FPR – False Positive Rate

FRP – Fiberglass Pipe

GA – Genetic Algorithm

GIS – Geographic Information Systems

HDPE – High-density Polyethylene

I/I – Infiltration/Inflow

IIMM – International Infrastructure Management Manual

JCW – Johnson County Wastewater

KNN – K-Nearest Neighbors

Log – Logarithm

MCMC – Markov Chain Monte Carlo

MDA – Mean Decrease Accuracy

MDI – Mean Decrease Impurity

MFL – Magnetic Flux Leakage

MLE – Maximum Likelihood Estimation

NASSCO – National Association of Sewer Service Companies

NetCoS – Network Condition Simulator

NNs – Neural Networks

NRC – National Research Council Canada

O&M – Operation and Maintenance

OR – Odds Ratio

OR – Overall Pipe Rating

PACP – Pipeline Assessment and Certification Program

PCCP – Prestressed Concrete Cylinder Pipe

PE – Plastic Pipe

PVC – Polyvinyl Chloride Pipes

RCP – Reinforced Concrete Pipes

RFEC – Remote Field Eddy Current

RI – Pipe Rating Index

ROC – Receiver Operating Characteristic

SBFS – Sequential Backward Floating Selection

SBS – Sequential Backward Selection

SE – Standard Error

SFAs – Sequential Feature Algorithms

SFFS – Sequential Forward Floating Selection

SFS – Sequential Forward Selection

SG – Segment Grade Score

SPSS – IBM SPSS Statistics Packages

SRM – Sewerage Rating Manual

SSS – Separate Sanitary Sewer and Storm Sewer System

SVM – Support Vector Machine

TN – True Negative

TP – True Positive

TPR – True Positive Rate

U.S. – United States

VCP – Vitrified Clay Pipes

WEF – Water Environment Federation

WRc – Water Research Centre

WSAA – Water Services Association of Australia

Appendix B

Data Sample (1,000 pipe segments)

If you need more data please contact Mohammadreza.malekmohammadi@mavs.uta.edu

N	Age	Material	Diameter	Flow	Depth	Slope	Length	Sulfate	pH	Soil	Water Table	Hydraulic	Corrosivity
1	63	1	8	366	3.29	0.45	679.99	0.02	5.3	Silty Gravel and Sand	145	0	2
2	50	1	24	2523	8.08	0.10	657.95	0.05	5.5	Fine Sand	84	0	2
3	12	1	10	4559	8.17	-0.50	611.32	0.02	5.3	Fine Sand	145	0	2
4	44	1	8	328	2.95	0.37	594.50	0.02	5.5	Silty Gravel and Sand	31	0	1
5	44	1	8	284	3.12	0.27	569.50	0.02	7.0	Silty Gravel and Sand	8	2	1
6	46	1	8	489	4.53	0.81	563.70	0.02	5.3	Fine Sand	145	0	2
7	66	1	8	817	3.59	2.27	547.63	0.17	5.3	Fine Sand	145	0	2
8	54	1	8	343	5.71	0.40	545.32	0.02	4.8	Silty Gravel and Sand	31	1	2
9	57	1	18	1664	10.8	0.12	544.61	0.02	5.3	Silty Gravel and Sand	145	0	2
10	42	1	8	201	6.16	0.14	531.98	0.02	5.3	Fine Sand	145	0	2
11	50	1	30	5204	6.64	0.08	525.61	0.10	7.5	Silty Gravel and Sand	8	0	2
12	50	1	30	5480	9.59	0.09	520.00	0.02	7.9	Silty Gravel and Sand	31	1	2
13	44	1	10	384	3.85	0.16	518.79	0.02	5.5	Silty Gravel and Sand	31	0	1
14	43	1	8	429	4.45	0.63	518.00	0.10	5.3	Fine Sand	8	0	2
15	44	1	8	182	3.53	0.11	505.00	0.02	5.3	Fine Sand	145	0	2
16	57	1	24	7631	8.68	0.56	500.00	0.02	5.3	Fine Sand	145	0	2
17	50	1	30	28100	10.06	-0.50	500.00	0.05	5.5	Fine Sand	15	0	1
18	57	1	24	7084	13.29	0.49	500.00	0.02	5.3	Fine Sand	145	0	2
19	51	1	10	467	4.26	0.23	495.42	0.02	5.3	Fine Sand	145	0	2
20	50	1	30	6221	12.45	0.11	494.68	0.02	5.5	Fine Sand	31	0	1
21	57	1	24	7711	10.17	0.58	482.70	0.10	5.3	Fine Sand	8	0	2
22	50	1	30	4390	12.37	0.06	478.73	0.02	5.5	Fine Sand	31	0	1
23	14	2	18	2601	10.09	0.30	476.96	0.02	5.3	Fine Sand	145	0	2
24	50	1	30	5091	8.32	0.08	471.46	0.02	7.9	Fine Sand	31	1	2
25	49	1	8	296	13.31	0.30	465.84	0.10	5.4	Fine Sand	8	0	2
26	9	2	42	0	16.52	0.00	464.78	0.02	7.0	Fine Sand	69	0	1
27	42	1	8	192	3.94	0.13	455.00	0.02	5.3	Fine Sand	145	0	2
28	46	1	24	3819	10.56	0.14	453.00	0.02	5.5	Fine Sand	31	0	1
29	56	1	30	3775	11.58	0.04	452.69	0.02	5.5	Fine Sand	31	0	1
30	20	2	8	0	5.96	0.00	451.81	0.02	5.5	Fine Sand	31	0	1
31	50	1	30	4669	6	0.06	451.65	0.02	5.8	Silty Gravel and Sand	15	3	1

32	48	1	8	250	10.65	0.21	450.98	0.02	7.0	Silty Gravel and Sand	8	2	1
33	64	1	8	297	4.53	0.30	450.00	0.02	4.8	Silty Gravel and Sand	31	1	2
34	53	1	15	1664	16.32	0.00	449.98	0.05	5.5	Silty Gravel and Sand	84	0	2
35	43	1	12	434	7.47	0.07	449.39	0.02	5.3	Silty Gravel and Sand	145	0	2
36	50	1	30	0	9.25	0.00	448.00	0.05	5.5	Silty Gravel and Sand	15	0	1
37	44	1	8	344	5.87	0.40	447.14	0.10	5.5	Silty Gravel and Sand	31	0	1
38	44	1	8	311	4.4	0.33	447.00	0.02	5.5	Silty Gravel and Sand	31	0	1
39	56	1	30	4773	12.5	0.07	446.98	0.02	5.5	Silty Gravel and Sand	31	0	1
40	59	1	8	293	4.43	0.29	444.64	0.02	7.0	Silty Gravel and Sand	8	2	1
41	51	1	8	280	8.71	0.27	442.26	0.02	7.0	Silty Gravel and Sand	69	0	1
42	56	1	30	6139	12.7	0.11	441.30	0.02	5.5	Silty Gravel and Sand	31	0	1
43	49	1	8	627	4.65	1.33	440.03	0.02	7.3	Silty Gravel and Sand	31	3	1
44	55	1	15	1424	10.82	0.24	439.70	0.02	7.0	Silty Gravel and Sand	8	2	1
45	50	1	18	1135	12.84	0.12	437.50	0.05	5.5	Silty Gravel and Sand	84	0	2
46	56	1	30	3737	15.62	0.04	437.50	0.02	5.5	Silty Gravel and Sand	31	0	1
47	56	1	30	5359	12.5	0.08	437.37	0.02	5.5	Silty Gravel and Sand	31	0	1
48	56	1	24	2836	15.11	0.08	436.55	0.02	5.3	Silty Gravel and Sand	145	0	2
49	55	1	8	360	3.21	0.44	435.00	0.02	5.5	Silty Gravel and Sand	31	0	1
50	54	1	18	1631	4.89	0.12	435.00	0.02	4.8	Silty Gravel and Sand	31	1	2
51	37	1	8	567	14.57	0.00	434.83	0.02	7.3	Silty Gravel and Sand	31	3	1
52	68	1	8	286	3.64	0.28	433.34	0.02	7.3	Silty Gravel and Sand	31	3	1
53	49	1	18	2084	13.09	0.12	431.58	0.05	5.5	Silty Gravel and Sand	84	0	2
54	25	2	8	382	5.16	0.50	431.29	0.02	5.3	Silty Gravel and Sand	145	0	2
55	43	1	12	550	7.66	0.12	431.19	0.02	5.3	Silty Gravel and Sand	145	0	2
56	11	2	8	0	2.95	0.00	430.68	0.02	5.3	Silty Gravel and Sand	145	0	2
57	57	1	8	420	3.76	0.60	430.00	0.05	5.5	Silty Gravel and Sand	31	0	1
58	54	1	8	342	6.75	0.40	429.56	0.02	7.0	Silty Gravel and Sand	69	0	1
59	51	1	18	1658	7.47	0.12	429.13	0.02	4.8	Silty Gravel and Sand	31	1	2
60	50	1	30	6596	7.95	0.13	429.10	0.02	5.5	Silty Gravel and Sand	31	0	1
61	55	1	8	2203	5.05	-0.50	427.01	0.02	5.3	Silty Gravel and Sand	145	0	2
62	57	1	24	7795	7.48	0.59	426.48	0.10	5.3	Silty Gravel and Sand	8	0	2
63	41	1	8	248	18.41	0.21	426.42	0.02	7.0	Silty Gravel and Sand	8	2	1
64	44	1	8	285	4.85	0.28	426.22	0.02	7.0	Silty Gravel and Sand	8	2	1
65	56	1	48	19281	13.64	0.09	425.70	0.02	5.8	Silty Gravel and Sand	15	3	1

66	57	1	24	7971	10.45	0.62	425.70	0.10	5.3	Silty Gravel and Sand	8	0	2
67	55	1	8	451	4.59	0.69	425.40	0.02	5.5	Silty Gravel and Sand	8	0	2
68	54	1	18	1596	5.38	0.11	425.03	0.02	7.3	Silty Gravel and Sand	31	3	1
69	43	1	8	464	4.32	0.73	424.83	0.10	5.3	Silty Gravel and Sand	8	0	2
70	49	1	18	1651	10.93	0.12	424.73	0.05	5.5	Silty Gravel and Sand	84	0	2
71	50	1	30	4734	2.92	0.07	424.14	0.02	5.5	Silty Gravel and Sand	31	0	1
72	39	1	8	556	3.5	0.00	423.80	0.02	5.3	Silty Gravel and Sand	145	0	2
73	49	1	8	307	9.06	0.32	420.74	0.02	7.0	Silty Gravel and Sand	69	0	1
74	50	1	30	3113	6.9	0.03	420.27	0.10	7.5	Silty Gravel and Sand	8	0	2
75	57	1	24	10137	11.57	1.00	420.00	0.02	5.3	Silty Gravel and Sand	145	0	2
76	46	1	8	343	4.5	0.40	420.00	0.05	5.5	Silty Gravel and Sand	84	0	2
77	50	1	30	5279	8.76	0.08	419.66	0.02	5.5	Silty Gravel and Sand	31	0	1
78	50	1	8	343	8.55	0.40	419.00	0.02	7.0	Silty Gravel and Sand	69	0	1
79	37	1	8	621	4.26	0.00	419.00	0.02	7.3	Silty Gravel and Sand	31	3	1
80	28	2	21	5902	6.81	0.69	419.00	0.02	5.4	Silty Gravel and Sand	8	0	2
81	55	1	8	357	4.11	0.43	419.00	0.02	5.5	Silty Gravel and Sand	31	0	1
82	49	1	8	344	6.7	0.40	416.80	0.02	4.8	Silty Gravel and Sand	31	1	2
83	53	1	8	848	4.59	2.44	416.78	0.10	5.3	Silty Gravel and Sand	8	0	2
84	53	1	8	569	3.95	1.10	416.20	0.10	5.5	Silty Gravel and Sand	31	0	1
85	54	1	8	276	5.41	0.26	415.75	0.02	7.0	Silty Gravel and Sand	8	2	1
86	46	1	8	353	7.02	0.42	415.64	0.10	5.1	Silty Gravel and Sand	31	0	2
87	13	2	8	286	11.5	-0.28	415.28	0.02	7.0	Silty Gravel and Sand	8	2	1
88	54	1	18	1637	5.43	0.12	415.00	0.02	4.8	Silty Gravel and Sand	31	1	2
89	50	1	8	805	4.5	2.20	415.00	0.10	5.3	Silty Gravel and Sand	8	0	2
90	55	1	8	308	3.99	0.32	414.80	0.02	5.3	Silty Gravel and Sand	145	0	2
91	54	1	8	461	4.12	0.00	414.69	0.02	5.3	Silty Gravel and Sand	145	0	2
92	54	1	8	537	5.32	0.98	414.33	0.02	5.3	Silty Gravel and Sand	145	0	2
93	46	1	24	3803	12.2	0.14	414.00	0.02	5.5	Silty Gravel and Sand	31	0	1
94	46	1	8	594	4.86	1.20	413.25	0.02	5.3	Silty Gravel and Sand	145	0	2
95	48	1	8	355	4.42	0.43	413.00	0.02	7.3	Silty Gravel and Sand	31	3	1
96	49	1	8	305	4.98	0.32	412.40	0.05	5.5	Silty Gravel and Sand	31	0	1
97	47	1	8	543	4.28	1.00	412.00	0.02	5.3	Silty Gravel and Sand	145	0	2
98	56	1	24	28100	14.08	-0.50	411.60	0.02	5.3	Silty Gravel and Sand	145	0	2
99	54	1	8	459	6.55	0.71	411.40	0.02	5.3	Silty Gravel and Sand	145	0	2

100	51	1	8	343	4.4	0.40	411.00	0.05	5.5	Silty Gravel and Sand	84	0	2
101	51	1	8	343	4.04	0.40	411.00	0.05	5.5	Silty Gravel and Sand	84	0	2
102	31	1	8	193	4.38	0.00	410.68	0.02	7.3	Silty Gravel and Sand	31	3	1
103	54	1	8	351	6.53	0.42	410.25	0.02	5.3	Silty Gravel and Sand	145	0	2
104	42	1	8	343	4.68	0.40	410.00	0.05	5.5	Silty Gravel and Sand	15	0	1
105	51	1	8	343	4.18	0.40	409.82	0.05	5.5	Silty Gravel and Sand	84	0	2
106	56	1	24	4225	12.98	0.17	409.33	0.02	5.3	Silty Gravel and Sand	145	0	2
107	38	1	8	489	5.86	0.81	409.00	0.02	5.5	Silty Gravel and Sand	31	0	1
108	61	1	8	326	5.34	0.36	408.85	0.02	7.3	Silty Gravel and Sand	31	3	1
109	40	1	8	208	7.4	0.15	408.00	0.02	5.5	Silty Gravel and Sand	31	0	1
110	47	1	8	343	6.76	0.40	407.35	0.02	5.3	Silty Gravel and Sand	145	0	2
111	55	1	15	1007	16.93	0.12	406.89	0.02	7.0	Silty Gravel and Sand	8	2	1
112	54	1	8	331	11.72	0.37	406.32	0.10	5.3	Silty Gravel and Sand	8	0	2
113	20	2	8	364	5.85	0.45	406.16	0.10	5.1	Silty Gravel and Sand	31	0	2
114	55	1	8	370	4.67	0.47	406.00	0.05	5.5	Silty Gravel and Sand	84	0	2
115	66	1	8	257	9.78	0.22	405.90	0.05	5.5	Silty Gravel and Sand	15	0	1
116	54	1	8	342	4.32	0.40	405.50	0.02	7.0	Silty Gravel and Sand	8	2	1
117	51	1	8	326	6.84	0.36	405.00	0.02	7.0	Silty Gravel and Sand	69	0	1
118	52	1	8	343	3.91	0.40	405.00	0.02	5.5	Silty Gravel and Sand	31	0	1
119	43	1	8	680	6.81	1.57	405.00	0.10	5.3	Silty Gravel and Sand	8	0	2
120	41	1	8	0	3.84	0.00	405.00	0.05	5.5	Silty Gravel and Sand	84	0	2
121	28	2	21	6442	4.94	0.82	405.00	0.02	7.0	Silty Gravel and Sand	8	2	1
122	51	1	8	343	5.11	0.40	404.67	0.05	5.5	Silty Gravel and Sand	15	0	1
123	54	1	8	374	4.69	0.48	404.10	0.10	5.3	Silty Gravel and Sand	8	0	2
124	8	1	8	262	5.46	0.23	404.06	0.02	5.3	Silty Gravel and Sand	145	0	2
125	31	1	8	348	4.52	0.41	404.00	0.05	5.5	Silty Gravel and Sand	84	0	2
126	14	2	15	1768	4.35	0.37	404.00	0.02	5.3	Silty Gravel and Sand	145	0	2
127	54	1	8	498	4.8	0.84	404.00	0.02	7.3	Silty Gravel and Sand	31	3	1
128	54	1	8	310	6.68	0.33	403.98	0.02	7.0	Silty Gravel and Sand	8	2	1
129	43	1	8	297	4.55	0.30	403.96	0.10	5.3	Silty Gravel and Sand	8	0	2
130	47	1	8	324	7.38	0.36	403.12	0.02	7.0	Silty Gravel and Sand	69	0	1
131	50	1	8	343	3	0.40	403.00	0.02	7.0	Silty Gravel and Sand	8	2	1
132	31	1	8	342	6.38	0.40	403.00	0.05	5.5	Silty Gravel and Sand	84	0	2
133	54	1	8	468	5.36	0.74	402.60	0.02	7.0	Silty Gravel and Sand	8	2	1

134	50	1	24	3164	15.9	0.10	402.20	0.10	5.3	Silty Gravel and Sand	8	0	2
135	6	2	8	339	7.41	0.39	402.16	0.10	5.1	Silty Gravel and Sand	31	0	2
136	46	1	8	306	5.12	0.32	402.07	0.02	7.0	Silty Gravel and Sand	8	2	1
137	19	2	15	2661	4.6	0.84	402.07	0.05	5.5	Silty Gravel and Sand	84	0	2
138	35	2	8	341	5.18	0.40	402.00	0.05	5.5	Silty Gravel and Sand	84	0	2
139	43	1	8	531	4.42	0.96	402.00	0.10	5.3	Silty Gravel and Sand	8	0	2
140	43	1	8	636	5.18	1.38	402.00	0.10	5.3	Silty Gravel and Sand	8	0	2
141	44	1	8	298	4.68	0.30	402.00	0.10	5.3	Silty Gravel and Sand	8	0	2
142	54	1	8	340	6.44	0.39	401.90	0.10	5.3	Silty Gravel and Sand	8	0	2
143	54	1	8	870	4.71	2.57	401.80	0.10	5.3	Silty Gravel and Sand	130	0	2
144	36	2	8	337	4.21	0.39	401.76	0.05	5.3	Silty Gravel and Sand	8	0	2
145	54	1	8	382	4.52	0.50	401.70	0.10	5.3	Silty Gravel and Sand	8	0	2
146	54	1	8	347	8.69	0.41	401.60	0.05	5.5	Silty Gravel and Sand	15	0	1
147	54	1	8	339	5.08	0.39	401.60	0.10	5.3	Silty Gravel and Sand	8	0	2
148	54	1	8	413	5.05	0.58	401.50	0.10	5.3	Silty Gravel and Sand	8	0	2
149	39	1	8	285	4.16	0.28	401.50	0.02	7.0	Silty Gravel and Sand	8	2	1
150	46	1	8	432	1.88	0.63	401.50	0.10	5.1	Silty Gravel and Sand	31	0	2
151	54	1	8	313	9.61	0.33	401.47	0.02	5.3	Silty Gravel and Sand	145	0	2
152	67	1	8	486	4.47	0.80	401.40	0.02	5.3	Silty Gravel and Sand	145	0	2
153	54	1	8	349	5.54	0.41	401.40	0.02	7.3	Silty Gravel and Sand	31	3	1
154	54	1	8	429	4.18	0.63	401.30	0.10	5.3	Silty Gravel and Sand	8	0	2
155	21	1	8	318	7.9	0.34	401.30	0.02	4.6	Silty Gravel and Sand	84	0	2
156	53	1	8	302	9.98	0.31	401.21	0.02	5.3	Silty Gravel and Sand	145	0	2
157	9	2	8	0	4.93	0.00	401.21	0.02	5.3	Silty Gravel and Sand	84	0	2
158	54	1	8	576	5.08	1.13	401.20	0.02	5.3	Silty Gravel and Sand	145	0	2
159	57	1	24	7743	3.83	0.58	401.20	0.10	5.3	Silty Gravel and Sand	8	0	2
160	54	1	8	308	3.52	0.32	401.20	0.02	5.3	Silty Gravel and Sand	145	0	2
161	54	1	8	348	4.67	0.41	401.10	0.10	5.3	Silty Gravel and Sand	8	0	2
162	54	1	8	352	5.39	0.42	401.00	0.02	5.3	Silty Gravel and Sand	145	0	2
163	44	1	8	310	4.44	0.33	401.00	0.02	4.8	Silty Gravel and Sand	31	1	2
164	55	1	8	555	5.48	1.05	401.00	0.02	4.8	Silty Gravel and Sand	130	0	2
165	54	1	8	366	8.96	0.45	401.00	0.02	5.3	Silty Gravel and Sand	145	0	2
166	51	1	8	497	4.55	0.84	401.00	0.02	5.3	Silty Gravel and Sand	145	0	2
167	43	1	8	530	4.2	0.95	401.00	0.10	5.3	Silty Gravel and Sand	8	0	2

168	54	1	8	349	4.95	0.41	401.00	0.10	5.3	Silty Gravel and Sand	8	0	2
169	43	1	8	317	6.63	0.34	401.00	0.10	5.3	Silty Gravel and Sand	8	0	2
170	33	2	8	365	13.83	0.45	401.00	0.02	8.2	Silty Gravel and Sand	59	0	2
171	54	1	8	342	5.34	0.40	401.00	0.10	5.3	Silty Gravel and Sand	8	0	2
172	44	1	8	292	2.44	0.29	401.00	0.02	5.5	Silty Gravel and Sand	31	0	1
173	51	1	8	343	8.35	0.40	400.97	0.10	5.3	Silty Gravel and Sand	8	0	2
174	54	1	8	353	5.49	0.42	400.80	0.10	5.3	Silty Gravel and Sand	8	0	2
175	54	1	8	336	5.25	0.38	400.80	0.02	5.3	Silty Gravel and Sand	145	0	2
176	55	1	8	307	5.67	0.32	400.80	0.02	5.3	Silty Gravel and Sand	145	0	2
177	55	1	10	932	6.81	0.00	400.73	0.02	5.8	Silty Gravel and Sand	15	3	1
178	54	1	8	350	4.31	0.42	400.70	0.10	5.3	Silty Gravel and Sand	8	0	2
179	55	1	10	554	12.62	0.32	400.60	0.02	5.5	Silty Gravel and Sand	8	0	2
180	54	1	8	966	5.44	3.17	400.60	0.02	5.3	Silty Gravel and Sand	145	0	2
181	43	1	8	320	5.31	0.35	400.50	0.10	5.3	Silty Gravel and Sand	8	0	2
182	54	1	8	473	6.2	0.76	400.50	0.02	5.3	Silty Gravel and Sand	145	0	2
183	55	1	8	307	3.86	0.32	400.50	0.02	5.3	Silty Gravel and Sand	145	0	2
184	55	1	10	477	11.43	0.23	400.50	0.02	5.5	Silty Gravel and Sand	8	0	2
185	49	1	8	320	3.59	0.35	400.45	0.02	7.3	Silty Gravel and Sand	31	3	1
186	54	1	8	559	6.05	1.06	400.40	0.05	5.5	Silty Gravel and Sand	84	0	2
187	54	1	8	303	5.56	0.31	400.40	0.02	5.3	Silty Gravel and Sand	145	0	2
188	57	1	8	342	3.88	0.40	400.40	0.10	5.3	Silty Gravel and Sand	8	0	2
189	21	1	8	331	5.89	0.37	400.40	0.02	5.5	Silty Gravel and Sand	31	0	1
190	55	1	10	777	6.1	0.62	400.30	0.02	7.0	Silty Gravel and Sand	8	2	1
191	54	1	8	844	5.11	2.42	400.30	0.10	5.3	Silty Gravel and Sand	8	0	2
192	53	1	8	310	4.06	0.33	400.24	0.05	5.5	Silty Gravel and Sand	15	0	1
193	55	1	8	339	4.96	0.39	400.16	0.02	4.8	Silty Gravel and Sand	130	0	2
194	22	2	8	297	2.68	0.30	400.07	0.02	5.5	Silty Gravel and Sand	31	0	1
195	46	1	8	343	4.92	0.40	400.06	0.10	5.1	Silty Gravel and Sand	31	0	2
196	54	1	8	343	5.37	0.40	400.00	0.02	7.0	Silty Gravel and Sand	8	2	1
197	53	1	8	307	3.66	0.32	400.00	0.02	7.0	Silty Gravel and Sand	8	2	1
198	55	1	8	664	5	0.00	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1
199	45	1	8	316	4.22	0.34	400.00	0.05	5.5	Silty Gravel and Sand	84	0	2
200	48	1	8	368	8.15	0.46	400.00	0.02	7.0	Silty Gravel and Sand	8	2	1
201	44	1	8	308	5.1	0.32	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1

202	43	1	8	629	4.82	1.34	400.00	0.05	5.5	Silty Gravel and Sand	84	0	2
203	25	2	8	514	7.06	0.90	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
204	54	1	8	352	5.86	0.42	400.00	0.02	5.3	Silty Gravel and Sand	145	0	2
205	34	1	8	336	8.36	0.38	400.00	0.05	5.5	Silty Gravel and Sand	84	0	2
206	42	1	8	312	12.93	0.33	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
207	47	1	8	712	3.84	1.72	400.00	0.05	5.5	Silty Gravel and Sand	84	0	2
208	55	1	8	345	4.2	0.40	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1
209	55	1	8	636	5	0.00	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
210	43	1	8	312	10.72	0.33	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
211	54	1	8	816	4.78	2.26	400.00	0.02	5.3	Silty Gravel and Sand	145	0	2
212	55	1	8	395	6.88	0.00	400.00	0.02	4.8	Silty Gravel and Sand	130	0	2
213	49	1	8	420	5.4	0.00	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
214	56	1	8	399	4.57	0.54	400.00	0.02	5.4	Silty Gravel and Sand	8	0	2
215	54	1	12	905	6.12	0.32	400.00	0.05	5.5	Silty Gravel and Sand	84	0	2
216	42	1	8	483	4.58	0.79	400.00	0.02	7.3	Silty Gravel and Sand	31	3	1
217	46	1	8	343	6.16	0.40	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1
218	55	1	8	940	5	0.00	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
219	50	1	8	343	4.79	0.40	400.00	0.05	5.5	Silty Gravel and Sand	84	0	2
220	47	1	8	309	3.61	0.33	400.00	0.05	5.5	Silty Gravel and Sand	84	0	1
221	68	1	8	238	6.41	0.19	400.00	0.02	4.8	Silty Gravel and Sand	31	1	2
222	25	2	8	748	5.48	1.90	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
223	42	1	8	244	3.75	0.20	400.00	0.02	7.3	Silty Gravel and Sand	31	3	1
224	54	1	8	575	4.82	1.12	400.00	0.02	5.3	Silty Gravel and Sand	145	0	2
225	51	1	8	356	10.46	0.43	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1
226	44	1	8	326	4.49	0.36	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1
227	42	1	8	312	15.45	0.33	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
228	38	1	8	273	5.86	0.25	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1
229	54	1	8	416	5.43	0.59	400.00	0.10	5.4	Silty Gravel and Sand	8	0	2
230	43	1	8	318	7.57	0.34	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
231	46	1	18	3336	7.9	0.50	400.00	0.02	4.6	Silty Gravel and Sand	15	1	2
232	46	1	8	343	4.71	0.40	400.00	0.02	7.3	Silty Gravel and Sand	31	3	1
233	53	1	8	307	8.02	0.32	400.00	0.02	5.3	Silty Gravel and Sand	145	0	2
234	55	1	8	900	5.5	0.00	400.00	0.10	5.3	Silty Gravel and Sand	8	0	2
235	71	1	8	246	8.8	0.21	400.00	0.02	7.0	Silty Gravel and Sand	8	2	1

236	44	1	8	343	4.4	0.40	400.00	0.02	7.3	Silty Gravel and Sand	31	3	1
237	55	1	10	622	11.45	0.40	400.00	0.02	7.3	Silty Gravel and Sand	31	3	1
238	35	1	8	318	15.82	0.34	400.00	0.05	5.5	Silty Gravel and Sand	15	0	1
239	50	1	8	307	7.26	0.32	400.00	0.02	5.5	Silty Gravel and Sand	31	0	1
240	55	1	8	491	4.03	0.82	399.99	0.05	5.5	Silty Gravel and Sand	84	0	2
241	54	1	8	402	5.53	0.55	399.90	0.10	5.3	Silty Gravel and Sand	8	0	2
242	57	1	8	838	11.82	2.39	399.90	0.02	5.3	Silty Gravel and Sand	145	0	2
243	21	1	8	358	5.61	0.44	399.80	0.02	4.6	Silty Gravel and Sand	84	0	2
244	54	1	10	514	6.12	0.27	399.80	0.10	5.3	Silty Gravel and Sand	8	0	2
245	54	1	8	355	5.61	0.43	399.70	0.10	5.3	Silty Gravel and Sand	8	0	2
246	56	1	8	468	2.97	0.00	399.70	0.05	5.5	Silty Gravel and Sand	84	0	2
247	44	1	10	348	7.32	0.13	399.69	0.05	5.5	Silty Gravel and Sand	84	0	2
248	54	1	8	341	4.2	0.40	399.60	0.02	5.3	Fine Sand	145	0	2
249	54	1	8	339	5.92	0.39	399.60	0.02	7.0	Fine Sand	69	0	1
250	49	1	8	391	6.17	0.52	399.50	0.02	7.0	Fine Sand	8	2	1
251	53	1	8	659	5.85	1.47	399.50	0.02	5.3	Fine Sand	145	0	2
252	54	1	8	353	5.88	0.42	399.50	0.02	5.3	Fine Sand	145	0	2
253	54	1	10	504	10.75	0.26	399.50	0.10	5.3	Fine Sand	8	0	2
254	57	1	8	349	6.67	0.41	399.50	0.02	5.3	Fine Sand	145	0	2
255	42	1	8	215	5.5	0.16	399.50	0.02	7.3	Fine Sand	31	3	1
256	47	1	8	365	4.17	0.45	399.50	0.02	4.8	Fine Sand	31	1	2
257	21	1	8	353	5.61	0.42	399.50	0.02	4.6	Fine Sand	84	0	2
258	46	1	12	870	9.08	0.30	399.46	0.02	8.2	Fine Sand	59	0	2
259	53	1	8	349	5.58	0.41	399.45	0.02	5.6	Fine Sand	31	1	2
260	54	1	8	745	4.94	1.89	399.40	0.02	5.3	Fine Sand	145	0	2
261	54	1	8	544	4.56	1.00	399.40	0.10	5.4	Fine Sand	8	0	2
262	54	1	8	475	5.06	0.77	399.40	0.02	7.3	Fine Sand	31	3	1
263	14	2	8	673	5	1.54	399.38	0.02	7.0	Fine Sand	8	2	1
264	51	1	8	387	4.8	0.51	399.30	0.02	8.2	Fine Sand	59	0	2
265	57	1	8	613	5.13	1.28	399.30	0.10	5.3	Fine Sand	8	0	2
266	54	1	8	343	7.02	0.40	399.30	0.05	5.4	Fine Sand	8	0	2
267	55	1	8	357	3.18	0.43	399.30	0.02	5.5	Silty Gravel and Sand	31	0	1
268	57	1	8	313	5.08	0.33	399.30	0.10	5.3	Fine Sand	8	0	2
269	54	1	8	340	5.12	0.39	399.20	0.10	5.3	Fine Sand	8	0	2

270	54	1	8	339	3.99	0.39	399.20	0.02	5.3	Fine Sand	145	0	2
271	57	1	24	8074	10.66	0.63	399.10	0.10	5.3	Fine Sand	8	0	2
272	54	1	10	495	8.23	0.25	399.10	0.05	5.4	Silty Gravel and Sand	8	0	2
273	44	1	8	343	5.62	0.40	399.00	0.02	5.5	Silty Gravel and Sand	31	0	1
274	55	1	10	524	12.39	0.28	399.00	0.02	5.5	Silty Gravel and Sand	8	0	2
275	46	1	8	340	8.26	0.39	399.00	0.02	5.5	Silty Gravel and Sand	31	0	1
276	54	1	8	307	4.22	0.32	399.00	0.02	5.3	Fine Sand	145	0	2
277	43	1	8	365	4.06	0.45	399.00	0.10	5.3	Fine Sand	8	0	2
278	42	1	8	440	3.82	0.66	399.00	0.02	7.3	Fine Sand	31	3	1
279	49	1	8	337	5.15	0.39	399.00	0.02	7.0	Fine Sand	8	2	1
280	35	1	8	543	8.1	1.00	399.00	0.05	5.5	Fine Sand	15	0	1
281	50	1	8	470	5.24	0.75	399.00	0.02	5.3	Fine Sand	145	0	2
282	51	1	8	375	7.33	0.48	399.00	0.02	8.2	Fine Sand	59	0	2
283	43	1	8	311	13.58	0.33	399.00	0.10	5.3	Fine Sand	8	0	2
284	44	1	8	307	5.75	0.32	399.00	0.02	5.5	Fine Sand	31	0	1
285	54	1	8	656	4.7	1.46	399.00	0.10	5.3	Fine Sand	8	0	2
286	44	1	8	448	7.42	0.68	398.96	0.02	5.3	Fine Sand	145	0	2
287	20	2	8	384	11.33	0.50	398.95	0.10	5.1	Fine Sand	31	0	2
288	54	1	8	344	5.86	0.40	398.80	0.02	5.3	Fine Sand	145	0	2
289	43	1	8	595	6.6	1.20	398.70	0.10	5.3	Fine Sand	8	0	2
290	54	1	8	1006	5.19	3.44	398.70	0.17	5.3	Fine Sand	145	0	2
291	54	1	8	339	4.41	0.39	398.70	0.10	5.3	Fine Sand	8	0	2
292	51	1	8	335	4.84	0.38	398.60	0.10	5.3	Fine Sand	8	0	2
293	54	1	8	339	10.22	0.39	398.50	0.05	5.5	Fine Sand	84	0	2
294	54	1	8	363	7.17	0.45	398.41	0.02	5.3	Fine Sand	145	0	2
295	54	1	8	494	4.88	0.83	398.40	0.02	5.3	Silty Gravel and Sand	145	0	2
296	64	1	8	337	3.7	0.39	398.20	0.02	5.5	Silty Gravel and Sand	31	0	1
297	54	1	8	357	4.32	0.43	398.14	0.02	5.3	Silty Gravel and Sand	145	0	2
298	54	1	8	899	6.11	2.75	398.10	0.10	5.3	Silty Gravel and Sand	8	0	2
299	39	1	8	261	5.95	0.23	398.10	0.02	7.0	Silty Gravel and Sand	8	2	1
300	21	1	8	345	8.05	0.40	398.00	0.02	5.5	Silty Gravel and Sand	31	0	1
301	19	2	8	342	6.49	0.40	398.00	0.10	5.3	Silty Gravel and Sand	8	0	2
302	53	1	8	660	5.06	1.48	398.00	0.02	5.3	Silty Gravel and Sand	145	0	2
303	38	1	8	342	4	0.40	398.00	0.02	5.3	Silty Gravel and Sand	145	0	2

304	21	1	8	348	7.05	0.41	398.00	0.05	5.5	Silty Gravel and Sand	84	0	2
305	43	1	8	380	4.32	0.49	398.00	0.10	5.3	Silty Gravel and Sand	8	0	2
306	44	1	8	312	2.43	0.33	398.00	0.02	5.5	Fine Sand	31	0	1
307	44	1	8	309	6.64	0.32	398.00	0.10	5.3	Fine Sand	8	0	2
308	44	1	8	321	3.96	0.35	398.00	0.02	5.5	Fine Sand	31	0	1
309	35	1	8	350	7.22	0.42	398.00	0.10	5.3	Fine Sand	8	0	2
310	54	1	8	379	4.34	0.49	398.00	0.05	5.5	Silty Gravel and Sand	84	0	2
311	55	1	15	1199	15.52	0.17	397.75	0.10	5.1	Silty Gravel and Sand	31	0	2
312	41	1	8	505	6.28	0.87	397.69	0.02	5.5	Silty Gravel and Sand	31	0	1
313	53	1	8	307	7.62	0.32	397.68	0.02	5.3	Silty Gravel and Sand	145	0	2
314	46	1	8	344	8.93	0.40	397.68	0.10	5.1	Silty Gravel and Sand	31	0	2
315	43	1	8	95	4.63	0.03	397.65	0.02	4.8	Silty Gravel and Sand	31	1	2
316	54	1	8	425	4.25	0.61	397.60	0.10	5.3	Silty Gravel and Sand	130	0	2
317	42	1	8	313	4.37	0.33	397.60	0.05	5.5	Silty Gravel and Sand	84	0	2
318	42	1	8	313	7.03	0.33	397.57	0.10	5.3	Silty Gravel and Sand	8	0	2
319	55	1	8	303	4.93	0.31	397.57	0.02	5.3	Silty Gravel and Sand	145	0	2
320	41	1	8	325	5.6	0.36	397.50	0.05	5.5	Silty Gravel and Sand	84	0	2
321	45	1	8	342	8.25	0.40	397.12	0.05	5.5	Silty Gravel and Sand	84	0	2
322	44	1	8	354	3.96	0.43	397.00	0.02	5.5	Silty Gravel and Sand	31	0	1
323	55	1	10	820	8.29	0.70	397.00	0.02	5.5	Silty Gravel and Sand	31	0	1
324	54	1	8	350	5.78	0.42	397.00	0.02	5.3	Silty Gravel and Sand	145	0	2
325	15	2	10	536	7.33	0.30	397.00	0.05	5.1	Silty Gravel and Sand	31	0	2
326	43	1	8	307	7.18	0.32	397.00	0.10	5.3	Silty Gravel and Sand	8	0	2
327	43	1	8	321	9.82	0.35	397.00	0.02	5.5	Silty Gravel and Sand	31	0	1
328	53	1	8	495	5.15	0.83	396.86	0.10	5.3	Silty Gravel and Sand	8	0	2
329	54	1	8	347	5.54	0.41	396.71	0.05	5.5	Silty Gravel and Sand	84	0	2
330	54	1	8	351	4.86	0.42	396.70	0.05	5.5	Silty Gravel and Sand	84	0	2
331	53	1	8	353	8.53	0.42	396.40	0.05	5.5	Silty Gravel and Sand	84	0	2
332	54	1	8	352	8.72	0.42	396.30	0.10	5.3	Silty Gravel and Sand	8	0	2
333	52	1	8	294	5.21	0.00	396.30	0.02	5.3	Silty Gravel and Sand	145	0	2
334	54	1	8	336	7.74	0.38	396.20	0.02	7.0	Silty Gravel and Sand	69	0	1
335	44	1	8	228	5.65	0.18	396.18	0.02	4.8	Silty Gravel and Sand	31	1	2
336	53	1	8	342	6.34	0.40	396.10	0.10	5.3	Silty Gravel and Sand	8	0	2
337	54	1	8	334	6.11	0.38	396.00	0.05	5.5	Silty Gravel and Sand	84	0	2

338	23	2	8	386	4.8	0.51	396.00	0.02	7.3	Silty Gravel and Sand	31	3	1
339	55	1	8	302	6.14	0.31	395.97	0.02	5.5	Silty Gravel and Sand	31	0	1
340	37	1	24	4363	14.3	0.18	395.92	0.05	5.5	Silty Gravel and Sand	84	0	2
341	55	1	8	578	5.03	1.13	395.80	0.10	4.6	Silty Gravel and Sand	15	1	2
342	43	1	8	431	5.8	0.63	395.50	0.10	5.3	Silty Gravel and Sand	8	0	2
343	56	1	8	307	7.23	0.32	395.50	0.02	5.3	Silty Gravel and Sand	145	0	2
344	45	1	8	286	3.7	0.28	395.41	0.02	5.5	Silty Gravel and Sand	31	0	1
345	25	2	8	348	6.08	0.41	395.22	0.05	5.5	Silty Gravel and Sand	84	0	2
346	55	1	10	417	8.86	0.18	395.20	0.02	5.5	Silty Gravel and Sand	31	0	1
347	58	1	8	320	0.59	0.35	395.10	0.02	5.5	Silty Gravel and Sand	31	0	1
348	55	1	8	347	4.07	0.41	395.00	0.02	5.5	Silty Gravel and Sand	31	0	1
349	42	1	8	288	3.73	0.28	395.00	0.05	5.5	Silty Gravel and Sand	84	0	2
350	46	1	8	343	3.81	0.40	395.00	0.02	7.3	Silty Gravel and Sand	31	3	1
351	55	1	8	359	5.58	0.44	395.00	0.02	5.5	Silty Gravel and Sand	8	0	2
352	37	1	8	343	5.63	0.40	395.00	0.02	5.4	Silty Gravel and Sand	8	0	2
353	53	1	8	308	6.35	0.32	394.90	0.05	5.5	Silty Gravel and Sand	31	0	1
354	54	1	10	781	7.89	0.63	394.84	0.02	5.3	Silty Gravel and Sand	145	0	2
355	53	1	15	1032	9.69	0.00	394.80	0.05	5.5	Silty Gravel and Sand	84	0	2
356	55	1	8	348	3.45	0.41	394.64	0.02	5.5	Silty Gravel and Sand	31	0	1
357	42	1	15	1043	6.65	0.13	394.61	0.02	7.0	Silty Gravel and Sand	8	2	1
358	54	1	8	384	3.81	0.50	394.36	0.02	5.3	Silty Gravel and Sand	145	0	2
359	29	1	24	5925	10.07	0.30	394.32	0.05	5.5	Silty Gravel and Sand	84	0	2
360	55	1	8	352	5.34	0.42	394.20	0.02	5.5	Silty Gravel and Sand	31	0	1
361	43	1	8	632	4.69	1.35	394.13	0.10	5.3	Silty Gravel and Sand	8	0	2
362	53	1	8	342	3.77	0.40	394.10	0.10	5.3	Silty Gravel and Sand	8	0	2
363	44	1	8	309	4.74	0.32	394.00	0.02	5.5	Silty Gravel and Sand	31	0	1
364	51	1	8	423	4.62	0.61	394.00	0.02	7.0	Silty Gravel and Sand	69	0	1
365	54	1	8	351	6.64	0.42	393.80	0.10	5.3	Silty Gravel and Sand	8	0	2
366	53	1	8	539	4.6	0.99	393.60	0.05	5.5	Silty Gravel and Sand	84	0	2
367	54	1	8	313	4.76	0.33	393.56	0.02	5.5	Silty Gravel and Sand	31	0	1
368	55	1	8	543	5.06	1.00	393.22	0.02	7.3	Silty Gravel and Sand	31	3	1
369	54	1	8	359	4.9	0.44	393.20	0.02	5.3	Silty Gravel and Sand	145	0	2
370	47	1	8	343	3.93	0.40	393.20	0.02	5.5	Silty Gravel and Sand	31	0	1
371	68	1	8	368	3.5	0.46	393.15	0.02	5.3	Silty Gravel and Sand	145	0	2

372	56	1	21	28100	7.39	20.00	393.00	0.02	4.8	Silty Gravel and Sand	130	0	2
373	45	1	8	325	6.6	0.33	393.00	0.10	5.3	Silty Gravel and Sand	8	0	2
374	44	1	8	160	4.4	0.09	393.00	0.02	4.8	Silty Gravel and Sand	31	1	2
375	53	1	8	465	4.73	0.73	392.60	0.05	5.5	Silty Gravel and Sand	84	0	2
376	55	1	8	342	5.02	0.40	392.50	0.02	5.5	Silty Gravel and Sand	8	0	2
377	53	1	8	318	12.13	0.34	392.21	0.10	5.3	Silty Gravel and Sand	8	0	2
378	43	1	8	347	6.14	0.41	392.18	0.02	5.5	Silty Gravel and Sand	31	0	1
379	55	1	8	340	4	0.39	392.06	0.02	7.3	Silty Gravel and Sand	31	3	1
380	12	2	8	355	7.74	0.43	392.00	0.05	5.5	Silty Gravel and Sand	15	0	1
381	12	2	8	330	4.34	0.37	392.00	0.05	5.5	Silty Gravel and Sand	15	0	1
382	43	1	8	491	4.32	0.00	392.00	0.10	5.3	Silty Gravel and Sand	8	0	2
383	12	2	8	333	4.08	0.38	392.00	0.05	5.5	Silty Gravel and Sand	15	0	1
384	54	1	8	394	5.31	0.53	392.00	0.10	5.3	Silty Gravel and Sand	8	0	2
385	43	1	8	253	5.5	0.22	392.00	0.10	5.3	Silty Gravel and Sand	8	0	2
386	54	1	8	367	4.69	0.46	391.88	0.02	5.3	Silty Gravel and Sand	145	0	2
387	24	2	8	293	3.7	0.29	391.72	0.02	4.8	Silty Gravel and Sand	31	1	2
388	54	1	8	321	4.9	0.35	391.59	0.02	5.3	Silty Gravel and Sand	145	0	2
389	45	1	18	1713	8.99	0.15	391.53	0.05	5.5	Silty Gravel and Sand	84	0	2
390	53	1	8	553	4.83	0.00	391.50	0.05	5.5	Silty Gravel and Sand	84	0	2
391	55	1	8	339	4.33	0.39	391.50	0.05	5.5	Silty Gravel and Sand	84	0	2
392	54	1	8	360	4.13	0.44	391.45	0.02	5.5	Silty Gravel and Sand	31	0	1
393	57	1	8	343	6.93	0.40	391.30	0.02	5.3	Silty Gravel and Sand	145	0	2
394	53	1	8	639	6.8	1.39	391.30	0.02	4.8	Silty Gravel and Sand	130	0	2
395	54	1	8	344	5.13	0.40	391.20	0.10	5.3	Silty Gravel and Sand	8	0	2
396	29	2	8	332	6.81	0.37	391.00	0.02	5.1	Silty Gravel and Sand	31	0	2
397	43	1	8	310	6	0.33	391.00	0.02	4.8	Silty Gravel and Sand	31	1	2
398	55	1	8	537	5.98	0.98	391.00	0.02	5.3	Silty Gravel and Sand	84	0	2
399	54	1	8	372	4.25	0.47	390.92	0.02	5.5	Silty Gravel and Sand	31	0	1
400	54	1	15	1127	13.25	0.15	390.56	0.10	5.3	Silty Gravel and Sand	8	0	2
401	46	1	8	417	5.36	0.59	390.50	0.02	5.3	Silty Gravel and Sand	145	0	2
402	55	1	8	595	7.3	1.20	390.50	0.02	4.8	Silty Gravel and Sand	15	0	2
403	54	1	8	577	8.02	1.13	390.50	0.10	5.3	Silty Gravel and Sand	8	0	2
404	48	1	8	337	3.96	0.39	390.50	0.02	7.3	Silty Gravel and Sand	31	3	1
405	54	1	8	314	3.12	0.34	390.50	0.02	5.5	Silty Gravel and Sand	31	0	1

406	55	1	8	341	5.34	0.39	390.40	0.02	4.8	Silty Gravel and Sand	130	0	2
407	67	1	10	877	6	0.79	390.30	0.02	7.0	Silty Gravel and Sand	8	2	1
408	54	1	8	443	4.84	0.67	390.30	0.10	5.4	Silty Gravel and Sand	8	0	2
409	53	1	8	291	9.51	0.29	390.20	0.10	5.3	Silty Gravel and Sand	8	0	2
410	53	1	8	500	5.84	0.85	390.19	0.10	5.3	Silty Gravel and Sand	8	0	2
411	46	1	8	339	4.25	0.39	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
412	50	1	8	343	4.8	0.40	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
413	42	1	8	343	4.92	0.40	390.00	0.05	5.5	Silty Gravel and Sand	84	0	2
414	51	1	8	420	4.3	0.60	390.00	0.02	5.5	Silty Gravel and Sand	31	0	1
415	34	1	8	352	4.79	0.42	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
416	54	1	10	475	10.87	0.23	390.00	0.05	5.5	Silty Gravel and Sand	84	0	2
417	43	1	8	491	5.61	0.82	390.00	0.10	5.3	Silty Gravel and Sand	8	0	2
418	34	1	8	352	4.85	0.42	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
419	34	1	8	377	4.73	0.48	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
420	30	2	8	521	5	0.92	390.00	0.02	4.8	Silty Gravel and Sand	31	1	2
421	55	1	15	1128	12.58	0.15	390.00	0.02	7.0	Silty Gravel and Sand	8	2	1
422	43	1	8	371	10.3	0.47	390.00	0.10	5.3	Silty Gravel and Sand	8	0	2
423	42	1	8	456	3.81	0.71	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
424	51	1	8	420	4.4	0.60	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
425	34	1	8	353	4.86	0.42	390.00	0.02	7.3	Silty Gravel and Sand	31	3	1
426	54	1	8	543	6.17	1.00	390.00	0.02	7.0	Silty Gravel and Sand	8	2	1
427	46	1	8	211	3.09	0.15	390.00	0.02	5.3	Silty Gravel and Sand	145	0	2
428	55	1	8	398	3.72	0.54	390.00	0.02	5.5	Silty Gravel and Sand	31	0	1
429	50	1	8	228	4.04	0.18	389.90	0.02	7.3	Silty Gravel and Sand	31	3	1
430	54	1	8	342	5.6	0.40	389.90	0.02	7.3	Silty Gravel and Sand	31	3	1
431	57	1	8	329	8.68	0.37	389.80	0.02	5.3	Silty Gravel and Sand	145	0	2
432	63	1	8	292	8.13	0.29	389.70	0.02	5.5	Silty Gravel and Sand	31	0	1
433	51	1	18	1751	7.58	0.14	389.62	0.02	4.8	Silty Gravel and Sand	31	1	2
434	50	1	8	307	3.39	0.32	389.50	0.02	7.3	Silty Gravel and Sand	31	3	1
435	42	1	8	345	6.47	0.41	389.28	0.05	5.5	Silty Gravel and Sand	84	0	2
436	55	1	8	412	4.17	0.58	389.00	0.10	5.5	Silty Gravel and Sand	31	0	1
437	55	1	8	331	4.82	0.37	389.00	0.02	7.3	Silty Gravel and Sand	31	3	1
438	43	1	8	303	4.48	0.31	389.00	0.10	5.3	Silty Gravel and Sand	8	0	2
439	54	1	8	327	11.81	0.36	388.96	0.02	7.3	Silty Gravel and Sand	31	3	1

440	47	1	10	462	10.4	0.22	388.87	0.05	5.5	Silty Gravel and Sand	84	0	2
441	41	1	8	342	3.36	0.40	388.77	0.05	5.5	Silty Gravel and Sand	84	0	2
442	56	1	8	358	5.12	0.43	388.60	0.02	5.3	Silty Gravel and Sand	145	0	2
443	55	1	8	485	5.23	0.80	388.60	0.02	4.8	Silty Gravel and Sand	130	0	2
444	41	1	8	0	4.23	0.00	388.55	0.10	5.3	Silty Gravel and Sand	8	0	2
445	51	1	8	420	4.6	0.60	388.23	0.02	5.5	Silty Gravel and Sand	31	0	1
446	39	1	18	1333	11.22	0.08	388.00	0.02	4.8	Silty Gravel and Sand	31	1	2
447	43	1	8	334	6.52	0.38	388.00	0.02	4.8	Silty Gravel and Sand	31	1	2
448	54	1	8	355	3.24	0.43	387.90	0.02	7.3	Silty Gravel and Sand	31	3	1
449	40	2	8	808	4.5	2.22	387.78	0.02	5.5	Silty Gravel and Sand	31	0	1
450	56	1	30	5127	12.75	0.08	387.40	0.02	5.3	Silty Gravel and Sand	145	0	2
451	43	1	8	313	4.33	0.33	387.24	0.02	4.8	Silty Gravel and Sand	31	1	2
452	29	2	8	281	7.95	0.27	387.00	0.02	7.3	Silty Gravel and Sand	31	3	1
453	55	1	8	250	4.58	0.21	387.00	0.02	7.3	Silty Gravel and Sand	31	3	1
454	56	1	30	4776	14.99	0.07	387.00	0.02	5.3	Silty Gravel and Sand	145	0	2
455	43	1	10	378	15.03	0.15	387.00	0.05	5.5	Silty Gravel and Sand	84	0	2
456	44	1	8	316	3.31	0.34	387.00	0.02	5.5	Silty Gravel and Sand	31	0	1
457	43	1	8	498	5	0.84	387.00	0.10	5.3	Silty Gravel and Sand	8	0	2
458	58	1	12	1094	8.57	0.47	386.95	0.05	5.5	Silty Gravel and Sand	84	0	2
459	12	2	8	334	4.51	0.38	386.78	0.05	5.5	Silty Gravel and Sand	15	0	1
460	55	1	8	346	4.93	0.41	386.77	0.02	7.3	Silty Gravel and Sand	31	3	1
461	46	1	8	347	4.69	0.41	386.66	0.10	5.1	Silty Gravel and Sand	31	0	2
462	53	1	15	1094	4.6	0.14	386.50	0.05	5.5	Silty Gravel and Sand	84	0	2
463	55	1	10	456	9.43	0.21	386.50	0.02	5.3	Silty Gravel and Sand	84	0	2
464	21	1	8	259	8.21	0.23	386.50	0.20	6.8	Silty Gravel and Sand	8	3	0
465	55	1	8	340	10.92	0.39	386.49	0.02	5.3	Silty Gravel and Sand	84	0	2
466	41	1	8	343	6.62	0.40	386.36	0.02	7.0	Silty Gravel and Sand	8	2	1
467	30	2	8	314	5.5	0.33	386.29	0.10	5.1	Silty Gravel and Sand	31	0	2
468	29	2	8	320	3.16	0.35	386.27	0.02	4.8	Silty Gravel and Sand	31	1	2
469	21	2	8	304	5.9	0.31	386.09	0.05	5.5	Silty Gravel and Sand	15	0	1
470	43	1	8	452	6.76	0.69	386.04	0.10	5.3	Silty Gravel and Sand	8	0	2
471	54	1	8	690	4.49	1.62	386.00	0.02	5.3	Silty Gravel and Sand	145	0	2
472	55	1	8	351	4.82	0.42	386.00	0.02	7.3	Silty Gravel and Sand	31	3	1
473	54	1	10	623	7.81	0.40	386.00	0.02	5.3	Silty Gravel and Sand	145	0	2

474	45	1	8	496	4.48	0.83	386.00	0.05	5.5	Silty Gravel and Sand	84	0	2
475	55	1	8	345	5.15	0.40	386.00	0.02	4.8	Silty Gravel and Sand	130	0	2
476	55	1	8	371	4.51	0.47	386.00	0.02	7.3	Silty Gravel and Sand	31	3	1
477	53	1	15	1304	11.22	0.00	385.80	0.05	5.5	Silty Gravel and Sand	84	0	2
478	55	1	8	279	3.75	0.26	385.73	0.02	5.5	Silty Gravel and Sand	31	0	1
479	53	1	8	465	5.74	0.74	385.72	0.02	4.8	Silty Gravel and Sand	130	0	2
480	53	1	8	515	4.02	0.90	385.70	0.02	5.5	Silty Gravel and Sand	31	0	1
481	57	1	8	343	4.13	0.40	385.70	0.10	5.3	Silty Gravel and Sand	8	0	2
482	55	1	8	387	4.32	0.51	385.70	0.02	7.3	Silty Gravel and Sand	31	3	1
483	55	1	8	394	4.15	0.53	385.50	0.02	7.3	Silty Gravel and Sand	31	3	1
484	54	1	8	311	3.3	0.33	385.50	0.02	5.5	Silty Gravel and Sand	31	0	1
485	57	1	8	342	6.08	0.40	385.50	0.02	5.3	Silty Gravel and Sand	145	0	2
486	55	1	8	347	12.34	0.41	385.50	0.02	7.3	Silty Gravel and Sand	31	3	1
487	54	1	8	334	5.72	0.38	385.45	0.02	7.0	Silty Gravel and Sand	8	2	1
488	41	1	15	1066	3.83	0.13	385.35	0.02	7.0	Silty Gravel and Sand	8	2	1
489	54	1	8	341	4.55	0.39	385.30	0.02	5.3	Silty Gravel and Sand	145	0	2
490	63	1	8	347	4.4	0.41	385.30	0.02	7.3	Silty Gravel and Sand	31	3	1
491	49	1	10	558	8.9	0.32	385.24	0.05	5.5	Silty Gravel and Sand	15	0	1
492	68	1	15	1263	12.07	0.19	385.00	0.02	5.3	Silty Gravel and Sand	145	0	2
493	46	1	8	441	4.33	0.66	385.00	0.05	5.5	Silty Gravel and Sand	15	0	1
494	31	1	8	2544	8.83	0.29	385.00	0.02	4.8	Silty Gravel and Sand	31	1	2
495	33	1	8	393	4.67	0.52	385.00	0.02	5.5	Silty Gravel and Sand	31	0	1
496	25	2	8	358	5.22	0.00	385.00	0.10	5.7	Silty Gravel and Sand	30	0	1
497	54	1	8	343	7.22	0.40	385.00	0.02	7.0	Silty Gravel and Sand	8	2	1
498	44	1	8	303	3.29	0.31	385.00	0.02	4.8	Silty Gravel and Sand	31	1	2
499	42	1	8	250	4.63	0.21	385.00	0.02	5.5	Silty Gravel and Sand	31	0	1
500	50	1	8	485	3.66	0.80	385.00	0.02	7.0	Silty Gravel and Sand	69	0	1
501	55	1	8	410	7.09	0.57	385.00	0.02	7.3	Silty Gravel and Sand	31	3	1
502	55	1	8	299	4.32	0.30	385.00	0.02	5.3	Silty Gravel and Sand	145	0	2
503	43	1	8	910	6.59	2.81	384.99	0.10	5.3	Silty Gravel and Sand	8	0	2
504	42	1	8	114	4.16	0.04	384.93	0.02	5.5	Silty Gravel and Sand	31	0	1
505	57	1	8	512	5.07	0.89	384.70	0.10	5.3	Silty Gravel and Sand	8	0	2
506	55	1	8	348	3.98	0.41	384.67	0.02	7.3	Silty Gravel and Sand	31	3	1
507	60	1	8	639	3.68	1.39	384.64	0.02	7.0	Silty Gravel and Sand	69	0	1

508	22	2	8	422	4.55	0.61	384.63	0.10	5.1	Silty Gravel and Sand	31	0	2
509	42	1	8	343	6.06	0.40	384.60	0.05	5.5	Silty Gravel and Sand	84	0	2
510	55	1	8	339	4.12	0.39	384.50	0.02	7.3	Silty Gravel and Sand	31	3	1
511	53	1	8	341	6.2	0.40	384.46	0.02	5.3	Silty Gravel and Sand	145	0	2
512	43	1	15	4278	5.91	2.17	384.00	0.10	5.3	Silty Gravel and Sand	8	0	2
513	43	1	8	407	5.75	0.56	384.00	0.02	5.5	Silty Gravel and Sand	31	0	1
514	63	1	8	356	5.02	0.43	383.90	0.02	7.3	Silty Gravel and Sand	31	3	1
515	54	1	8	356	5.59	0.43	383.80	0.10	5.3	Silty Gravel and Sand	8	0	2
516	55	1	8	327	2.57	0.36	383.50	0.02	5.5	Silty Gravel and Sand	31	0	1
517	55	1	8	368	4.47	0.46	383.50	0.02	7.3	Silty Gravel and Sand	31	3	1
518	51	1	10	492	11.41	0.25	383.41	0.02	5.5	Silty Gravel and Sand	31	0	1
519	55	1	8	398	5.17	0.54	383.20	0.02	7.3	Silty Gravel and Sand	31	3	1
520	50	1	18	1913	10.59	0.12	383.20	0.05	5.5	Silty Gravel and Sand	84	0	2
521	54	1	8	348	10	0.41	383.10	0.02	7.3	Silty Gravel and Sand	31	3	1
522	55	1	8	333	8.67	0.38	383.00	0.02	5.3	Silty Gravel and Sand	145	0	2
523	55	1	8	370	3.71	0.46	382.80	0.02	7.3	Silty Gravel and Sand	31	3	1
524	54	1	8	315	6.52	0.34	382.70	0.02	5.3	Silty Gravel and Sand	145	0	2
525	53	1	8	929	4.23	2.93	382.63	0.10	4.8	Silty Gravel and Sand	130	0	2
526	43	1	8	295	7.51	0.30	382.60	0.02	5.3	Silty Gravel and Sand	84	0	2
527	54	1	8	372	5.28	0.47	382.50	0.05	5.5	Silty Gravel and Sand	84	0	2
528	43	1	8	311	6.89	0.33	382.50	0.02	4.8	Silty Gravel and Sand	31	1	2
529	55	1	8	502	5.57	0.86	382.47	0.05	5.5	Silty Gravel and Sand	84	0	2
530	57	1	8	551	5.7	0.00	382.43	0.10	5.3	Silty Gravel and Sand	8	0	2
531	40	1	8	543	4.2	1.00	382.04	0.02	5.3	Silty Gravel and Sand	145	0	2
532	43	2	8	486	6.01	0.80	382.00	0.10	5.3	Silty Gravel and Sand	8	0	2
533	29	2	8	318	5.71	0.34	382.00	0.02	7.0	Silty Gravel and Sand	8	2	1
534	20	2	8	502	6.12	0.86	382.00	0.02	5.8	Silty Gravel and Sand	15	3	1
535	20	2	24	3329	12.6	0.11	382.00	0.10	4.8	Silty Gravel and Sand	31	1	2
536	41	1	12	714	8.65	0.20	381.73	0.02	5.5	Silty Gravel and Sand	31	0	1
537	6	2	8	0	4.6	0.00	381.71	0.02	7.0	Silty Gravel and Sand	8	0	1
538	42	1	8	350	5.2	0.00	381.67	0.02	5.5	Silty Gravel and Sand	31	0	1
539	51	1	18	1758	7.65	0.14	381.50	0.02	4.8	Silty Gravel and Sand	31	1	2
540	55	1	10	483	11.51	0.24	381.40	0.02	5.3	Silty Gravel and Sand	84	0	2
541	52	1	8	298	6.99	0.00	381.35	0.10	5.3	Silty Gravel and Sand	8	0	2

542	16	1	8	1314	5.27	-0.50	381.30	0.02	5.5	Silty Gravel and Sand	31	0	1
543	55	1	10	456	12.93	0.22	381.10	0.02	4.5	Silty Gravel and Sand	145	0	2
544	53	1	8	420	5.1	0.60	381.00	0.02	5.3	Silty Gravel and Sand	145	0	2
545	55	1	8	358	4.49	0.40	381.00	0.05	5.5	Silty Gravel and Sand	84	0	2
546	55	1	8	397	4.08	0.54	381.00	0.02	7.3	Silty Gravel and Sand	31	3	1
547	33	2	8	347	15.64	0.41	380.99	0.02	8.2	Silty Gravel and Sand	59	0	2
548	29	1	24	5448	8.15	0.30	380.85	0.05	5.5	Silty Gravel and Sand	84	0	2
549	54	1	24	4357	20.1	0.18	380.71	0.10	5.3	Silty Gravel and Sand	8	0	2
550	68	1	15	1132	6.59	0.15	380.70	0.02	5.3	Silty Gravel and Sand	145	0	2
551	53	1	8	395	5.73	0.53	380.60	0.02	5.5	Silty Gravel and Sand	31	0	1
552	54	1	8	364	8.41	0.45	380.60	0.10	5.3	Silty Gravel and Sand	8	0	2
553	33	1	12	562	8.51	0.12	380.58	0.02	7.3	Silty Gravel and Sand	31	3	1
554	53	1	8	557	4.88	1.05	380.50	0.05	5.5	Silty Gravel and Sand	84	0	2
555	55	1	8	307	5.36	0.32	380.37	0.05	5.5	Silty Gravel and Sand	84	0	2
556	46	1	8	345	4.63	0.40	380.27	0.02	5.3	Silty Gravel and Sand	145	0	2
557	55	1	10	465	11.03	0.22	380.24	0.02	5.3	Silty Gravel and Sand	84	0	2
558	51	1	8	479	7.36	0.78	380.20	0.02	7.0	Silty Gravel and Sand	69	0	1
559	55	1	8	512	5.7	0.89	380.07	0.05	5.5	Silty Gravel and Sand	15	0	1
560	53	1	8	343	4.61	0.40	380.00	0.02	5.3	Silty Gravel and Sand	145	0	2
561	53	1	8	543	5.19	1.00	380.00	0.02	5.4	Silty Gravel and Sand	8	0	2
562	53	1	8	488	4.41	0.81	380.00	0.02	5.3	Silty Gravel and Sand	145	0	2
563	45	1	8	312	6.72	0.33	380.00	0.02	4.8	Silty Gravel and Sand	130	0	2
564	50	1	8	332	6.88	0.37	380.00	0.02	7.3	Silty Gravel and Sand	31	3	1
565	34	1	8	273	14.45	0.25	380.00	0.02	4.8	Silty Gravel and Sand	31	1	2
566	45	1	18	2082	18.39	0.18	380.00	0.05	5.5	Silty Gravel and Sand	84	0	2
567	53	1	8	374	5.48	0.48	380.00	0.02	5.3	Silty Gravel and Sand	145	0	2
568	46	1	8	352	5.69	0.42	380.00	0.02	5.5	Silty Gravel and Sand	31	0	1
569	47	1	8	343	8.73	0.40	380.00	0.05	5.5	Silty Gravel and Sand	84	0	2
570	43	1	8	273	4.37	0.25	380.00	0.02	4.8	Silty Gravel and Sand	31	1	2
571	53	1	8	365	5.38	0.45	380.00	0.02	5.3	Silty Gravel and Sand	145	0	2
572	32	1	30	9309	14.09	0.26	380.00	0.02	4.0	Silty Gravel and Sand	30	0	2
573	50	1	24	2211	8.49	0.10	379.78	0.05	5.5	Silty Gravel and Sand	84	0	2
574	54	1	8	925	5.3	2.91	379.71	0.10	5.3	Silty Gravel and Sand	8	0	2
575	54	1	8	347	6.29	0.41	379.70	0.02	5.3	Silty Gravel and Sand	145	0	2

576	47	1	8	353	7.34	0.42	379.67	0.05	5.5	Silty Gravel and Sand	15	0	1
577	21	2	8	295	8.75	0.40	379.60	0.05	5.5	Silty Gravel and Sand	84	0	2
578	53	1	8	343	4.96	0.40	379.50	0.02	5.3	Silty Gravel and Sand	145	0	2
579	53	1	8	195	5.09	0.13	379.47	0.02	5.3	Silty Gravel and Sand	145	0	2
580	53	1	8	0	7.13	0.00	379.28	0.02	7.0	Silty Gravel and Sand	8	2	1
581	54	1	8	360	7.17	0.44	379.00	0.02	5.3	Silty Gravel and Sand	145	0	2
582	33	1	8	254	6.2	0.22	379.00	0.02	7.0	Silty Gravel and Sand	69	0	1
583	54	1	8	354	5.84	0.42	379.00	0.02	5.3	Silty Gravel and Sand	145	0	2
584	53	1	8	378	4.38	0.49	379.00	0.02	5.3	Silty Gravel and Sand	145	0	2
585	48	1	8	283	8.16	0.27	378.93	0.02	7.0	Silty Gravel and Sand	8	2	1
586	43	1	8	312	3.92	0.33	378.74	0.10	5.3	Silty Gravel and Sand	8	0	2
587	51	1	8	0	5.39	0.00	378.36	0.10	5.3	Silty Gravel and Sand	8	0	2
588	51	1	8	301	6.52	0.31	378.34	0.05	5.5	Silty Gravel and Sand	15	0	1
589	39	1	8	231	8.36	0.18	378.04	0.02	7.0	Silty Gravel and Sand	8	2	1
590	33	2	10	304	11.07	0.10	378.03	0.02	8.2	Silty Gravel and Sand	59	0	2
591	43	1	8	383	10.52	0.50	378.00	0.24	4.8	Silty Gravel and Sand	31	1	2
592	50	1	8	487	4.25	0.80	378.00	0.02	5.5	Silty Gravel and Sand	8	0	2
593	34	1	8	344	4.6	0.40	378.00	0.02	5.5	Silty Gravel and Sand	31	0	1
594	43	1	8	600	4.16	1.22	378.00	0.10	5.3	Silty Gravel and Sand	8	0	2
595	50	1	8	343	6.68	0.40	378.00	0.02	7.0	Silty Gravel and Sand	69	0	1
596	53	1	8	336	4.61	0.38	377.63	0.02	5.3	Silty Gravel and Sand	145	0	2
597	31	1	8	438	5.54	0.00	377.56	0.02	5.3	Silty Gravel and Sand	145	0	2
598	60	1	8	335	4.41	0.38	377.35	0.02	5.8	Silty Gravel and Sand	8	3	1
599	43	1	8	312	4.25	0.33	377.00	0.02	4.8	Silty Gravel and Sand	31	1	2
600	43	1	8	444	6.4	0.67	377.00	0.02	4.8	Silty Gravel and Sand	31	1	2
601	43	1	8	311	4.17	0.33	377.00	0.02	5.3	Silty Gravel and Sand	145	0	2
602	15	2	10	555	5.11	0.32	376.94	0.02	7.0	Silty Gravel and Sand	8	2	1
603	46	1	8	376	4.97	0.48	376.81	0.02	5.3	Silty Gravel and Sand	145	0	2
604	57	1	8	847	13.92	2.44	376.80	0.02	5.3	Silty Gravel and Sand	145	0	2
605	53	1	8	550	4.63	0.00	376.80	0.05	5.5	Silty Gravel and Sand	84	0	2
606	53	1	8	638	4.82	0.00	376.69	0.05	5.5	Silty Gravel and Sand	84	0	2
607	25	2	10	515	13.99	0.27	376.52	0.10	5.3	Silty Gravel and Sand	8	0	2
608	43	1	8	313	5.64	0.33	376.50	0.02	4.8	Silty Gravel and Sand	31	1	2
609	53	1	8	306	4.65	0.32	376.08	0.02	5.5	Silty Gravel and Sand	31	0	1

610	43	1	8	249	5.4	0.21	376.00	0.02	4.8	Silty Gravel and Sand	31	1	2
611	39	1	10	340	8.21	0.00	376.00	0.02	4.8	Silty Gravel and Sand	31	1	2
612	43	1	8	312	3.61	0.33	376.00	0.02	4.8	Silty Gravel and Sand	31	1	2
613	55	1	8	249	12.53	0.00	376.00	0.02	5.5	Silty Gravel and Sand	8	0	2
614	41	1	8	278	7.89	0.26	375.99	0.02	5.3	Silty Gravel and Sand	145	0	2
615	55	1	8	461	4.19	0.72	375.80	0.02	7.3	Silty Gravel and Sand	31	3	1
616	54	1	8	614	4.87	1.28	375.80	0.02	7.0	Silty Gravel and Sand	8	2	1
617	54	1	8	358	4.58	0.44	375.70	0.10	5.3	Silty Gravel and Sand	8	0	2
618	43	1	8	311	3.56	0.33	375.50	0.10	5.3	Silty Gravel and Sand	8	0	2
619	55	1	8	462	4.87	0.72	375.49	0.05	5.5	Silty Gravel and Sand	84	0	2
620	68	1	8	299	4.02	0.30	375.43	0.02	7.3	Silty Gravel and Sand	31	3	1
621	51	1	8	318	8.16	0.34	375.30	0.02	5.5	Silty Gravel and Sand	31	0	1
622	21	2	8	303	12.1	0.31	375.12	0.05	5.5	Silty Gravel and Sand	15	0	1
623	57	1	8	342	5.01	0.40	375.07	0.10	5.3	Silty Gravel and Sand	8	0	2
624	56	1	8	384	3.96	0.50	375.00	0.02	4.8	Silty Gravel and Sand	130	0	2
625	41	1	8	312	7.32	0.33	375.00	0.02	5.5	Silty Gravel and Sand	31	0	1
626	55	1	8	295	12.67	0.30	375.00	0.02	5.5	Silty Gravel and Sand	31	0	1
627	55	1	8	343	3.24	0.40	375.00	0.17	5.3	Silty Gravel and Sand	145	0	2
628	44	1	8	309	4.67	0.33	375.00	0.02	4.8	Silty Gravel and Sand	31	1	2
629	54	1	8	340	4.9	0.39	375.00	0.02	7.0	Silty Gravel and Sand	69	0	1
630	50	1	8	343	8.79	0.40	375.00	0.02	7.0	Silty Gravel and Sand	69	0	1
631	55	1	10	521	4.53	0.28	375.00	0.02	7.3	Silty Gravel and Sand	31	3	1
632	55	1	8	371	5.83	0.47	375.00	0.10	5.1	Silty Gravel and Sand	31	0	2
633	53	1	8	343	4.7	0.40	375.00	0.02	5.3	Silty Gravel and Sand	145	0	2
634	55	1	10	313	4.15	0.10	375.00	0.02	7.3	Silty Gravel and Sand	31	3	1
635	54	1	8	343	3.96	0.40	375.00	0.02	4.8	Silty Gravel and Sand	31	1	2
636	44	1	8	407	2.56	0.56	375.00	0.02	4.8	Silty Gravel and Sand	31	1	2
637	47	1	8	337	5.55	0.39	375.00	0.02	5.3	Silty Gravel and Sand	145	0	2
638	55	1	8	497	5.5	0.84	375.00	0.02	4.8	Silty Gravel and Sand	130	0	2
639	40	1	8	351	5.8	0.42	375.00	0.24	4.8	Silty Gravel and Sand	31	1	2
640	47	1	8	343	10.2	0.40	375.00	0.10	5.4	Silty Gravel and Sand	8	0	2
641	56	1	8	515	3.92	0.90	375.00	0.02	4.8	Silty Gravel and Sand	130	0	2
642	53	1	8	398	4.39	0.54	375.00	0.02	5.3	Silty Gravel and Sand	145	0	2
643	25	2	8	343	3.74	0.40	375.00	0.02	4.8	Silty Gravel and Sand	31	1	2

644	41	1	8	626	5	1.33	375.00	0.05	5.5	Silty Gravel and Sand	84	0	2
645	69	1	8	343	4.83	0.40	374.50	0.02	7.0	Silty Gravel and Sand	8	2	1
646	51	2	8	543	5.45	1.00	374.40	0.10	5.1	Silty Gravel and Sand	31	0	2
647	36	1	8	449	4.44	0.00	374.38	0.05	5.5	Silty Gravel and Sand	84	0	2
648	55	1	8	473	5.5	0.00	374.35	0.02	7.3	Silty Gravel and Sand	31	3	1
649	13	2	8	423	6.36	0.61	374.33	0.10	5.1	Silty Gravel and Sand	31	0	2
650	37	1	24	4267	11.64	0.18	374.31	0.05	5.5	Silty Gravel and Sand	84	0	2
651	68	1	8	307	4.05	0.32	374.05	0.02	7.3	Silty Gravel and Sand	31	3	1
652	33	1	8	716	4.95	1.74	374.00	0.10	5.3	Silty Gravel and Sand	8	0	2
653	43	1	10	817	11.02	0.69	374.00	0.02	4.8	Silty Gravel and Sand	31	1	2
654	43	1	8	297	4.7	0.30	374.00	0.02	4.8	Silty Gravel and Sand	31	1	2
655	57	1	8	344	3.92	0.40	374.00	0.02	5.3	Silty Gravel and Sand	145	0	2
656	43	1	8	363	9.01	0.45	374.00	0.02	4.8	Silty Gravel and Sand	31	1	2
657	53	1	8	373	5.19	0.47	374.00	0.02	5.3	Silty Gravel and Sand	145	0	2
658	35	1	8	344	5.09	0.40	373.94	0.02	5.3	Silty Gravel and Sand	145	0	2
659	56	1	8	455	4.35	0.70	373.80	0.02	5.3	Silty Gravel and Sand	145	0	2
660	54	1	8	659	4.19	1.48	373.80	0.02	7.3	Silty Gravel and Sand	31	3	1
661	61	1	8	343	8.18	0.40	373.59	0.02	5.5	Silty Gravel and Sand	8	0	1
662	54	1	8	411	7.43	0.57	373.50	0.10	5.4	Silty Gravel and Sand	8	0	2
663	55	1	8	301	3.75	0.31	373.39	0.02	5.5	Silty Gravel and Sand	31	0	1
664	68	1	8	447	3.9	0.68	373.19	0.02	5.3	Silty Gravel and Sand	145	0	2
665	51	1	8	341	6.26	0.39	373.05	0.02	7.0	Silty Gravel and Sand	69	0	1
666	55	1	8	359	4.31	0.44	373.00	0.02	4.5	Silty Gravel and Sand	145	0	2
667	34	1	8	312	5.34	0.33	373.00	0.02	5.5	Silty Gravel and Sand	31	0	1
668	42	1	8	312	5.33	0.33	373.00	0.10	5.3	Silty Gravel and Sand	8	0	2
669	43	1	8	221	7.4	0.17	373.00	0.02	4.8	Silty Gravel and Sand	31	1	2
670	36	1	8	274	9.26	0.32	373.00	0.05	5.5	Silty Gravel and Sand	84	0	2
671	34	1	8	477	5.3	0.77	373.00	0.02	4.8	Silty Gravel and Sand	31	1	2
672	57	1	8	359	4.63	0.44	373.00	0.02	5.3	Silty Gravel and Sand	145	0	2
673	43	1	8	331	7.22	0.37	373.00	0.02	4.8	Silty Gravel and Sand	31	1	2
674	55	1	8	301	6.52	0.31	372.76	0.02	5.3	Silty Gravel and Sand	84	0	2
675	19	2	8	338	15.5	0.39	372.62	0.05	5.1	Silty Gravel and Sand	31	0	2
676	54	1	8	408	4.81	0.57	372.50	0.05	5.4	Silty Gravel and Sand	8	0	2
677	54	1	8	343	11.18	0.40	372.50	0.02	5.3	Silty Gravel and Sand	145	0	2

678	53	1	8	370	8.68	0.46	372.40	0.05	5.5	Silty Gravel and Sand	84	0	2
679	53	1	8	307	5.13	0.32	372.10	0.02	5.3	Silty Gravel and Sand	145	0	2
680	41	1	8	307	5.14	0.32	372.00	0.02	5.5	Silty Gravel and Sand	31	0	1
681	50	1	8	407	7.29	0.56	372.00	0.02	7.3	Silty Gravel and Sand	31	3	1
682	66	1	8	343	4.76	0.40	371.80	0.02	5.3	Silty Gravel and Sand	84	0	2
683	43	1	15	1115	5.92	0.15	371.78	0.02	8.2	Silty Gravel and Sand	59	0	2
684	54	1	8	354	6.47	0.43	371.73	0.02	5.3	Silty Gravel and Sand	145	0	2
685	53	1	8	341	9.6	0.40	371.62	0.05	5.5	Silty Gravel and Sand	15	0	1
686	57	1	8	342	5.67	0.40	371.60	0.10	5.3	Silty Gravel and Sand	8	0	2
687	51	1	8	342	7.88	0.40	371.50	0.02	7.0	Silty Gravel and Sand	69	0	1
688	17	2	8	328	16.01	0.37	371.41	0.10	5.1	Silty Gravel and Sand	31	0	2
689	24	2	8	338	6.58	0.39	371.36	0.02	5.8	Silty Gravel and Sand	15	3	1
690	51	1	8	340	8.9	0.39	371.30	0.02	5.5	Silty Gravel and Sand	31	0	1
691	54	1	8	339	7.62	0.39	371.21	0.02	7.3	Silty Gravel and Sand	31	3	1
692	55	1	8	539	5.43	0.99	371.20	0.05	5.5	Silty Gravel and Sand	84	0	2
693	51	1	8	383	6.05	0.50	371.19	0.02	7.0	Silty Gravel and Sand	69	0	1
694	15	2	8	0	5	0.00	371.13	0.02	5.5	Silty Gravel and Sand	8	0	2
695	55	1	8	559	5.75	0.00	370.93	0.02	5.3	Silty Gravel and Sand	84	0	2
696	40	1	8	565	4.2	1.08	370.93	0.02	5.5	Silty Gravel and Sand	31	0	1
697	37	1	8	0	4.28	0.00	370.63	0.17	5.3	Silty Gravel and Sand	145	0	2
698	55	1	8	283	4.03	0.27	370.50	0.02	5.3	Silty Gravel and Sand	145	0	2
699	46	1	8	355	7.73	0.43	370.48	0.02	5.3	Silty Gravel and Sand	145	0	2
700	45	1	8	343	6.49	0.40	370.42	0.05	5.5	Silty Gravel and Sand	84	0	2
701	67	1	8	792	3.7	2.13	370.40	0.02	5.3	Silty Gravel and Sand	145	0	2
702	33	1	8	322	3.68	0.35	370.00	0.02	5.3	Silty Gravel and Sand	145	0	2
703	25	2	8	330	7.37	0.37	370.00	0.02	5.4	Silty Gravel and Sand	8	0	2
704	55	1	8	345	5.6	0.00	370.00	0.02	5.5	Silty Gravel and Sand	8	0	2
705	53	2	8	275	7.34	0.00	370.00	0.05	5.5	Silty Gravel and Sand	84	0	2
706	55	1	8	365	4.92	0.45	370.00	0.02	5.5	Silty Gravel and Sand	31	0	1
707	42	1	8	441	4.77	0.66	370.00	0.10	5.3	Silty Gravel and Sand	8	0	2
708	60	1	8	505	3.75	0.87	370.00	0.02	7.0	Silty Gravel and Sand	69	0	1
709	34	1	8	349	4.58	0.41	370.00	0.02	7.3	Silty Gravel and Sand	31	3	1
710	35	1	8	586	4.28	1.17	370.00	0.10	5.3	Silty Gravel and Sand	8	0	2
711	48	1	8	420	7.31	0.60	370.00	0.10	5.3	Silty Gravel and Sand	8	0	2

712	55	1	8	558	5.5	0.00	370.00	0.02	7.3	Silty Gravel and Sand	31	3	1
713	55	1	8	359	6.52	0.44	369.96	0.10	4.6	Silty Gravel and Sand	15	1	2
714	54	1	8	183	16.77	0.11	369.80	0.02	7.3	Silty Gravel and Sand	31	3	1
715	54	1	8	335	11.11	0.38	369.64	0.02	5.3	Silty Gravel and Sand	145	0	2
716	40	1	8	318	14.68	0.34	369.60	0.02	4.8	Silty Gravel and Sand	31	1	2
717	54	1	15	990	15.1	0.00	369.50	0.10	5.3	Silty Gravel and Sand	8	0	2
718	46	1	8	898	2.51	2.74	369.10	0.02	5.3	Silty Gravel and Sand	145	0	2
719	43	1	10	478	7.37	0.24	369.00	0.10	5.5	Silty Gravel and Sand	31	0	1
720	55	1	8	368	5.28	0.46	369.00	0.02	5.5	Silty Gravel and Sand	8	0	2
721	53	1	8	307	3.18	0.32	369.00	0.02	7.0	Silty Gravel and Sand	8	2	1
722	54	1	8	312	3.51	0.33	368.93	0.02	5.5	Silty Gravel and Sand	31	0	1
723	43	2	8	0	12.22	0.00	368.92	0.10	5.3	Silty Gravel and Sand	8	0	2
724	54	1	12	736	14.81	0.21	368.73	0.02	5.3	Silty Gravel and Sand	145	0	2
725	41	1	8	344	7.94	0.40	368.65	0.05	5.5	Silty Gravel and Sand	84	0	2
726	21	2	8	353	7.31	0.42	368.52	0.05	5.5	Silty Gravel and Sand	84	0	2
727	14	1	15	1479	8.97	0.26	368.27	0.02	5.3	Silty Gravel and Sand	145	0	2
728	51	1	8	314	6.48	0.33	368.25	0.02	7.0	Silty Gravel and Sand	69	0	1
729	57	1	8	307	8.47	0.32	368.17	0.02	4.5	Silty Gravel and Sand	145	0	2
730	18	2	8	343	7.7	0.40	368.04	0.02	7.3	Silty Gravel and Sand	31	3	1
731	55	1	8	352	5.59	0.42	368.00	0.02	5.3	Silty Gravel and Sand	145	0	2
732	47	1	8	279	7.63	0.40	368.00	0.05	5.5	Silty Gravel and Sand	84	0	2
733	35	1	8	371	5.8	0.47	368.00	0.10	5.3	Silty Gravel and Sand	8	0	2
734	34	1	10	495	14.7	0.25	368.00	0.02	4.8	Silty Gravel and Sand	31	1	2
735	27	2	8	343	3.63	0.40	368.00	0.02	4.8	Silty Gravel and Sand	31	1	2
736	45	1	8	321	7.49	0.35	368.00	0.05	5.5	Silty Gravel and Sand	84	0	2
737	32	1	8	343	5.62	0.40	368.00	0.02	4.8	Silty Gravel and Sand	31	1	2
738	43	1	8	325	4.88	0.36	368.00	0.02	4.8	Silty Gravel and Sand	31	1	2
739	44	1	10	518	12.1	0.28	368.00	0.05	5.5	Silty Gravel and Sand	84	0	2
740	43	1	8	324	10.9	0.36	367.80	0.05	5.5	Silty Gravel and Sand	84	0	2
741	23	2	8	233	5.8	0.00	367.78	0.02	4.8	Silty Gravel and Sand	31	1	2
742	38	1	8	342	4.28	0.40	367.50	0.02	7.3	Silty Gravel and Sand	31	3	1
743	55	1	8	463	5.65	0.73	367.12	0.05	5.5	Silty Gravel and Sand	84	0	2
744	40	1	15	1081	7.81	0.14	367.00	0.02	4.8	Silty Gravel and Sand	31	1	2
745	55	1	8	375	6.54	0.48	367.00	0.10	5.1	Silty Gravel and Sand	31	0	2

746	35	1	8	525	6.35	0.93	367.00	0.10	5.3	Silty Gravel and Sand	8	0	2
747	35	1	8	609	6.19	1.26	366.76	0.10	5.3	Silty Gravel and Sand	8	0	2
748	51	1	8	342	8.14	0.40	366.70	0.10	5.1	Silty Gravel and Sand	31	0	2
749	41	1	8	342	4.13	0.40	366.69	0.05	5.5	Silty Gravel and Sand	84	0	2
750	40	1	8	512	5.8	0.89	366.47	0.10	5.3	Silty Gravel and Sand	8	0	2
751	14	2	8	340	8.32	0.39	366.43	0.02	4.8	Silty Gravel and Sand	130	0	2
752	39	1	8	266	6.77	0.24	366.32	0.02	5.3	Silty Gravel and Sand	84	0	2
753	17	2	8	0	15.35	0.00	366.04	0.05	5.5	Silty Gravel and Sand	15	0	1
754	58	1	12	796	5.4	0.25	366.00	0.02	4.8	Silty Gravel and Sand	130	0	2
755	34	1	8	321	4.05	0.35	366.00	0.02	5.5	Silty Gravel and Sand	31	0	1
756	43	1	8	311	4.22	0.33	366.00	0.10	5.3	Silty Gravel and Sand	8	0	2
757	36	1	8	320	5.74	0.35	366.00	0.02	6.5	Silty Gravel and Sand	15	3	1
758	38	1	8	343	6.19	0.00	366.00	0.02	7.3	Silty Gravel and Sand	31	3	1
759	43	1	8	309	6.17	0.33	365.80	0.05	5.5	Silty Gravel and Sand	84	0	2
760	55	1	8	307	6.94	0.32	365.60	0.02	7.0	Silty Gravel and Sand	69	0	1
761	43	2	12	1129	11.48	0.50	365.50	0.10	5.3	Silty Gravel and Sand	8	0	2
762	54	1	8	343	4.84	0.40	365.50	0.02	7.0	Silty Gravel and Sand	8	2	1
763	55	2	8	422	4.82	0.60	365.40	0.02	5.5	Silty Gravel and Sand	8	0	2
764	31	2	8	325	8.57	0.00	365.15	0.10	5.1	Silty Gravel and Sand	31	0	2
765	34	1	18	2208	14.62	0.22	365.07	0.05	5.5	Silty Gravel and Sand	84	0	2
766	31	1	8	308	2.98	0.32	365.00	0.02	5.3	Silty Gravel and Sand	145	0	2
767	23	2	10	605	5.67	0.00	365.00	0.02	7.3	Silty Gravel and Sand	31	3	1
768	54	1	8	306	5.24	0.32	365.00	0.02	5.3	Silty Gravel and Sand	145	0	2
769	55	1	8	348	5.65	0.41	365.00	0.02	7.3	Silty Gravel and Sand	31	3	1
770	43	1	8	532	7.23	0.96	365.00	0.02	4.8	Silty Gravel and Sand	31	1	2
771	59	1	8	321	4.24	0.35	365.00	0.05	5.5	Silty Gravel and Sand	84	0	2
772	57	1	8	808	7.68	2.22	365.00	0.10	5.3	Silty Gravel and Sand	8	0	2
773	53	1	8	492	5.79	0.82	365.00	0.02	7.3	Silty Gravel and Sand	31	3	1
774	53	1	8	343	4.37	0.40	365.00	0.02	5.5	Silty Gravel and Sand	31	0	1
775	43	1	8	343	4.24	0.40	365.00	0.02	4.8	Silty Gravel and Sand	31	1	2
776	34	1	12	763	15.82	0.23	365.00	0.02	4.8	Silty Gravel and Sand	31	1	2
777	40	1	8	343	8.48	0.40	364.94	0.02	5.5	Silty Gravel and Sand	31	0	1
778	50	1	24	3406	7.77	0.10	364.90	0.05	5.5	Silty Gravel and Sand	84	0	2
779	59	1	8	536	3.97	0.97	364.71	0.02	7.3	Silty Gravel and Sand	31	3	1

780	37	1	12	829	12.98	0.27	364.70	0.02	4.8	Silty Gravel and Sand	31	1	2
781	54	1	8	348	9.21	0.41	364.60	0.10	5.3	Silty Gravel and Sand	8	0	2
782	49	1	10	554	9.8	0.32	364.50	0.05	5.5	Silty Gravel and Sand	15	0	1
783	30	1	36	10492	6.42	0.12	364.27	0.05	5.5	Silty Gravel and Sand	84	0	2
784	23	2	8	392	6.49	0.52	364.22	0.10	5.1	Silty Gravel and Sand	31	0	2
785	54	1	8	361	11.67	0.44	364.21	0.02	5.3	Fine Sand	145	0	2
786	46	1	8	363	7.3	0.45	364.11	0.02	5.6	Fine Sand	31	1	2
787	46	2	8	276	4.71	0.26	364.10	0.02	5.3	Fine Sand	145	0	2
788	15	2	18	1492	8.33	0.10	364.06	0.02	5.3	Fine Sand	145	0	2
789	34	1	8	306	6.41	0.32	364.00	0.02	4.8	Fine Sand	31	1	2
790	43	1	8	325	6.12	0.36	364.00	0.02	5.3	Fine Sand	84	0	2
791	39	1	8	305	4.2	0.32	364.00	0.02	5.5	Fine Sand	31	0	1
792	47	1	10	461	11.87	0.22	363.98	0.02	5.3	Fine Sand	8	0	2
793	55	1	15	988	15.1	0.11	363.83	0.10	5.1	Fine Sand	31	0	2
794	55	1	8	347	4.32	0.41	363.75	0.02	5.5	Fine Sand	31	0	1
795	41	1	8	628	5.84	1.34	363.67	0.02	5.5	Fine Sand	31	0	1
796	46	1	8	351	5.24	0.42	363.65	0.02	7.0	Fine Sand	69	0	1
797	46	1	8	490	5.27	0.81	363.59	0.02	7.0	Fine Sand	8	2	1
798	45	1	8	300	6.49	0.31	363.54	0.10	5.3	Fine Sand	8	0	2
799	60	1	8	321	5.5	0.35	363.52	0.05	5.5	Fine Sand	84	0	2
800	55	1	8	287	7.24	0.28	363.50	0.02	7.3	Fine Sand	31	3	1
801	47	1	8	343	3.93	0.40	363.48	0.02	4.8	Fine Sand	130	0	2
802	45	1	8	307	4.64	0.32	363.30	0.05	5.5	Fine Sand	84	0	2
803	57	1	8	391	4.61	0.52	363.30	0.02	7.3	Fine Sand	31	3	1
804	54	1	8	545	4.85	1.01	363.14	0.10	5.3	Fine Sand	8	0	2
805	14	2	8	346	2.86	0.41	363.08	0.02	5.3	Fine Sand	145	0	2
806	45	1	8	309	10.08	0.33	363.07	0.10	5.3	Fine Sand	8	0	2
807	43	1	8	292	6.08	0.29	363.00	0.05	5.5	Fine Sand	15	0	1
808	48	1	8	307	6.7	0.32	363.00	0.10	5.3	Fine Sand	8	0	2
809	43	1	10	462	7.83	0.22	363.00	0.02	7.3	Fine Sand	31	3	1
810	39	1	8	337	4.2	0.39	363.00	0.02	4.8	Fine Sand	31	1	2
811	55	1	8	355	8.94	0.43	363.00	0.05	5.5	Fine Sand	84	0	2
812	43	1	8	305	6.81	0.32	363.00	0.02	4.8	Fine Sand	31	1	2
813	54	1	10	757	4.88	0.59	362.86	0.10	5.3	Fine Sand	8	0	2

814	54	1	8	326	5.04	0.36	362.80	0.10	5.3	Fine Sand	8	0	2
815	54	1	8	295	7.14	0.29	362.74	0.02	7.0	Fine Sand	8	2	1
816	57	1	8	332	9.35	0.37	362.63	0.02	5.3	Fine Sand	145	0	2
817	44	1	8	314	5.02	0.33	362.62	0.02	4.8	Fine Sand	31	1	2
818	54	1	8	1025	5.98	3.57	362.60	0.02	7.0	Fine Sand	8	2	1
819	64	1	8	745	4.65	1.89	362.58	0.05	4.8	Fine Sand	8	0	2
820	43	1	8	322	5.8	0.35	362.50	0.10	5.3	Fine Sand	8	0	2
821	40	1	8	525	5.28	0.94	362.50	0.10	5.3	Fine Sand	8	0	2
822	29	2	8	329	4.02	0.37	362.43	0.02	6.6	Fine Sand	31	0	2
823	44	1	24	5120	8.47	0.25	362.40	0.02	5.3	Fine Sand	145	0	2
824	38	1	8	316	4.87	0.34	362.11	0.02	7.3	Fine Sand	31	3	1
825	30	2	8	330	6.35	0.37	362.07	0.10	5.1	Fine Sand	31	0	2
826	33	1	8	153	6.66	0.08	362.04	0.10	5.6	Fine Sand	31	1	2
827	14	2	8	387	6.74	-0.51	362.01	0.02	5.5	Fine Sand	69	0	1
828	43	1	8	306	4.73	0.32	362.00	0.02	4.8	Fine Sand	31	1	2
829	38	1	8	771	4	2.02	362.00	0.02	6.5	Fine Sand	8	3	1
830	43	1	8	547	4.82	1.02	362.00	0.10	5.3	Fine Sand	8	0	2
831	45	1	8	296	11.44	0.33	362.00	0.05	5.5	Fine Sand	84	0	2
832	43	1	8	520	4.7	0.92	362.00	0.10	5.3	Fine Sand	8	0	2
833	46	1	8	296	5.03	0.30	361.90	0.02	5.5	Fine Sand	31	0	1
834	17	2	8	392	13.96	0.52	361.87	0.10	5.1	Fine Sand	31	0	2
835	57	1	8	823	4.84	2.30	361.80	0.02	5.3	Fine Sand	145	0	2
836	56	1	24	3250	13.55	0.10	361.73	0.02	5.3	Fine Sand	145	0	2
837	27	2	8	295	13.03	0.00	361.70	0.10	5.3	Fine Sand	8	0	2
838	54	1	8	354	6.06	0.43	361.60	0.02	4.8	Fine Sand	130	0	2
839	38	1	8	244	5.55	0.20	361.50	0.02	7.0	Silty Gravel and Sand	69	0	1
840	54	1	8	396	4.51	0.53	361.40	0.10	5.3	Silty Gravel and Sand	8	0	2
841	63	1	8	318	6.97	0.34	361.40	0.02	4.8	Silty Gravel and Sand	31	1	2
842	54	1	8	419	5.62	0.60	361.30	0.05	5.5	Silty Gravel and Sand	84	0	2
843	64	1	8	299	3.74	0.30	361.30	0.02	5.3	Silty Gravel and Sand	145	0	2
844	6	2	8	0	6.26	0.00	361.24	0.02	7.3	Silty Gravel and Sand	31	3	1
845	38	1	8	306	3.83	0.32	361.23	0.10	5.3	Silty Gravel and Sand	8	0	2
846	34	1	8	332	4.24	0.37	361.21	0.05	5.5	Silty Gravel and Sand	84	0	2
847	64	1	8	161	5.95	0.09	361.19	0.02	7.3	Silty Gravel and Sand	31	3	1

848	31	2	8	344	7.03	0.40	361.17	0.02	7.3	Silty Gravel and Sand	31	3	1
849	54	1	10	494	7.97	0.25	361.12	0.02	7.3	Silty Gravel and Sand	31	3	1
850	55	1	8	340	4.77	0.39	361.00	0.02	7.3	Silty Gravel and Sand	31	3	1
851	43	2	8	873	3.8	2.59	361.00	0.10	5.3	Silty Gravel and Sand	8	0	2
852	57	1	8	356	7.9	0.43	361.00	0.10	5.3	Silty Gravel and Sand	8	0	2
853	42	1	8	332	6.83	0.37	361.00	0.02	7.3	Silty Gravel and Sand	31	3	1
854	59	1	8	343	3.91	0.40	361.00	0.02	5.3	Silty Gravel and Sand	84	0	2
855	40	1	8	674	4.86	1.54	361.00	0.02	5.3	Silty Gravel and Sand	145	0	2
856	50	1	8	343	4.74	0.40	361.00	0.10	5.3	Silty Gravel and Sand	59	0	2
857	53	1	8	352	4.73	0.42	361.00	0.02	5.3	Silty Gravel and Sand	145	0	2
858	43	1	8	326	4.12	0.36	361.00	0.10	5.3	Silty Gravel and Sand	8	0	2
859	58	1	8	397	2.72	0.53	360.99	0.02	5.5	Silty Gravel and Sand	31	0	1
860	54	1	24	4441	11.73	0.19	360.99	0.10	5.3	Silty Gravel and Sand	8	0	2
861	64	1	8	291	7.48	0.29	360.91	0.02	7.0	Silty Gravel and Sand	69	0	1
862	15	2	8	174	6.29	0.10	360.90	0.02	5.4	Silty Gravel and Sand	8	0	2
863	20	2	24	3253	14.68	0.10	360.89	0.10	4.8	Silty Gravel and Sand	31	1	2
864	40	1	8	353	13.06	0.42	360.85	0.02	4.8	Silty Gravel and Sand	130	0	2
865	54	1	12	669	16.95	0.17	360.80	0.02	5.3	Silty Gravel and Sand	145	0	2
866	23	2	8	287	6.42	0.28	360.68	0.02	5.3	Silty Gravel and Sand	145	0	2
867	55	1	10	439	8.62	0.20	360.64	0.02	7.3	Silty Gravel and Sand	31	3	1
868	29	1	36	10489	8.2	0.12	360.60	0.02	5.4	Silty Gravel and Sand	8	0	2
869	54	1	8	355	5.25	0.43	360.60	0.02	7.3	Silty Gravel and Sand	31	3	1
870	21	2	10	545	10.15	0.31	360.58	0.05	5.5	Silty Gravel and Sand	84	0	2
871	9	2	8	319	4.03	0.35	360.52	0.02	5.3	Silty Gravel and Sand	84	0	2
872	54	1	8	339	10.75	0.39	360.50	0.02	5.3	Silty Gravel and Sand	145	0	2
873	45	1	24	5251	16.21	0.26	360.46	0.05	5.5	Silty Gravel and Sand	84	0	2
874	31	2	8	0	4.26	0.00	360.38	0.02	7.3	Silty Gravel and Sand	31	3	1
875	21	2	8	374	9.03	0.47	360.23	0.10	5.1	Silty Gravel and Sand	31	0	2
876	59	1	8	767	3.98	2.00	360.20	0.02	5.3	Silty Gravel and Sand	145	0	2
877	52	1	18	1980	12.85	0.18	360.18	0.02	7.0	Silty Gravel and Sand	8	2	1
878	42	1	8	944	5.33	3.03	360.14	0.10	5.3	Silty Gravel and Sand	8	0	2
879	46	1	18	3498	7.8	0.00	360.12	0.02	5.5	Silty Gravel and Sand	31	0	1
880	54	1	8	636	7.57	1.38	360.10	0.10	5.3	Silty Gravel and Sand	8	0	2
881	55	1	10	461	6.52	0.22	360.10	0.02	5.8	Silty Gravel and Sand	15	3	1

882	4	2	8	0	14.05	0.00	360.06	0.10	5.1	Silty Gravel and Sand	31	0	2
883	59	1	8	343	4	0.40	360.00	0.05	5.5	Silty Gravel and Sand	15	0	1
884	43	1	8	304	4.3	0.31	360.00	0.10	5.3	Silty Gravel and Sand	8	0	2
885	68	1	8	297	3.29	0.30	360.00	0.02	4.8	Silty Gravel and Sand	31	1	2
886	31	2	8	348	4.62	0.41	360.00	0.10	5.1	Silty Gravel and Sand	31	0	2
887	34	1	8	441	7.7	0.66	360.00	0.02	4.8	Silty Gravel and Sand	31	1	2
888	33	1	18	1827	9.5	0.15	360.00	0.02	5.3	Silty Gravel and Sand	145	0	2
889	33	1	8	459	10.2	0.00	360.00	0.02	4.8	Silty Gravel and Sand	130	0	2
890	54	1	6	279	4.71	0.26	360.00	0.02	5.5	Silty Gravel and Sand	31	0	1
891	42	1	8	343	9.24	0.40	360.00	0.05	5.5	Silty Gravel and Sand	84	0	2
892	42	1	8	343	10.25	0.40	360.00	0.02	5.5	Silty Gravel and Sand	31	0	1
893	32	1	8	325	5.91	0.36	360.00	0.05	5.5	Silty Gravel and Sand	15	0	1
894	58	1	8	342	4.09	0.40	360.00	0.10	5.3	Silty Gravel and Sand	8	0	2
895	48	1	8	429	5.21	0.63	360.00	0.02	7.0	Silty Gravel and Sand	8	2	1
896	55	1	8	307	3.45	0.32	360.00	0.02	5.3	Silty Gravel and Sand	145	0	2
897	57	1	8	341	6.88	0.39	360.00	0.02	5.5	Silty Gravel and Sand	31	0	1
898	55	1	8	346	6.88	0.41	360.00	0.10	5.1	Silty Gravel and Sand	31	0	2
899	42	1	8	343	5.12	0.40	360.00	0.05	5.5	Silty Gravel and Sand	84	0	2
900	36	1	8	307	4	0.32	360.00	0.02	7.3	Silty Gravel and Sand	31	3	1
901	54	1	8	858	5.26	0.00	360.00	0.02	7.3	Silty Gravel and Sand	31	3	1
902	54	1	8	343	5.93	0.40	360.00	0.02	5.5	Silty Gravel and Sand	31	0	1
903	45	1	15	1174	14.42	0.16	360.00	0.02	5.5	Silty Gravel and Sand	8	0	1
904	53	1	8	3030	4.74	0.00	359.90	0.02	5.3	Silty Gravel and Sand	145	0	2
905	60	1	8	755	3.46	1.93	359.89	0.10	5.3	Silty Gravel and Sand	8	0	2
906	43	1	8	385	4.25	0.50	359.68	0.10	5.3	Silty Gravel and Sand	8	0	2
907	35	1	10	709	7.53	0.52	359.67	0.10	5.3	Silty Gravel and Sand	8	0	2
908	22	2	8	394	5.59	0.53	359.65	0.10	5.1	Silty Gravel and Sand	31	0	2
909	30	2	8	336	12	0.38	359.60	0.10	5.1	Silty Gravel and Sand	31	0	2
910	68	1	15	1316	12.45	0.21	359.60	0.02	5.3	Silty Gravel and Sand	145	0	2
911	48	1	8	301	7.56	0.31	359.60	0.02	5.5	Silty Gravel and Sand	8	0	1
912	18	2	8	227	7.41	0.18	359.47	0.02	5.8	Silty Gravel and Sand	15	3	1
913	3	2	8	0	4.87	0.00	359.45	0.02	5.3	Silty Gravel and Sand	145	0	2
914	53	1	8	340	7.19	0.39	359.38	0.02	5.5	Silty Gravel and Sand	31	0	1
915	22	2	8	488	4.15	0.81	359.34	0.10	5.1	Silty Gravel and Sand	31	0	2

916	53	1	8	307	4.03	0.32	359.30	0.02	7.3	Silty Gravel and Sand	31	3	1
917	55	1	12	706	7.2	0.00	359.26	0.02	7.3	Silty Gravel and Sand	31	3	1
918	47	1	8	0	5.86	0.00	359.23	0.02	4.8	Silty Gravel and Sand	130	0	2
919	32	1	10	519	7.8	0.28	359.11	0.10	5.6	Silty Gravel and Sand	31	1	2
920	63	1	8	396	3.84	0.53	359.06	0.02	7.3	Silty Gravel and Sand	31	3	1
921	35	1	10	498	6.03	0.26	359.00	0.02	8.2	Silty Gravel and Sand	59	0	2
922	20	2	24	3173	15.97	0.10	359.00	0.10	4.8	Silty Gravel and Sand	31	1	2
923	27	2	27	4814	10.09	0.12	359.00	0.02	5.5	Silty Gravel and Sand	31	0	1
924	55	1	8	509	4.6	0.88	359.00	0.02	7.3	Silty Gravel and Sand	31	3	1
925	50	1	8	543	5.7	1.00	359.00	0.02	5.3	Silty Gravel and Sand	145	0	2
926	39	1	8	438	4.12	0.65	359.00	0.02	5.5	Silty Gravel and Sand	31	0	1
927	35	1	10	786	8.44	0.64	358.97	0.10	5.3	Silty Gravel and Sand	8	0	2
928	59	1	8	454	4.05	0.70	358.90	0.10	5.1	Silty Gravel and Sand	31	0	2
929	51	1	8	348	7.4	0.41	358.84	0.02	7.0	Silty Gravel and Sand	69	0	1
930	41	1	8	373	8.54	0.47	358.80	0.02	4.8	Silty Gravel and Sand	31	1	2
931	49	1	8	307	3.16	0.32	358.67	0.02	7.3	Silty Gravel and Sand	31	3	1
932	39	1	8	415	4.03	0.59	358.65	0.05	5.5	Silty Gravel and Sand	15	0	1
933	44	1	8	226	6.22	0.17	358.60	0.02	5.3	Silty Gravel and Sand	145	0	2
934	59	1	8	387	3.68	0.51	358.50	0.02	7.0	Silty Gravel and Sand	69	0	1
935	67	1	15	1592	6.62	0.30	358.46	0.02	7.0	Silty Gravel and Sand	8	2	1
936	13	2	8	352	6.05	0.42	358.35	0.02	7.0	Silty Gravel and Sand	8	2	1
937	50	1	8	338	5.75	0.39	358.30	0.02	5.3	Silty Gravel and Sand	84	0	2
938	53	1	15	1396	18.83	0.23	358.25	0.10	5.6	Silty Gravel and Sand	31	1	2
939	60	1	8	575	4.2	1.12	358.20	0.10	5.3	Silty Gravel and Sand	8	0	2
940	54	1	8	698	6.04	1.65	358.19	0.02	5.3	Silty Gravel and Sand	145	0	2
941	35	1	8	337	9.7	0.39	358.03	0.02	5.5	Silty Gravel and Sand	31	0	1
942	42	1	8	492	4.12	0.82	358.00	0.02	5.3	Silty Gravel and Sand	145	0	2
943	31	1	8	339	5.63	0.39	358.00	0.02	4.8	Silty Gravel and Sand	31	1	2
944	64	1	8	298	7.42	0.30	358.00	0.02	5.5	Silty Gravel and Sand	31	0	1
945	18	2	8	249	5.64	0.00	358.00	0.02	4.0	Silty Gravel and Sand	30	0	2
946	34	1	8	264	6.9	0.24	358.00	0.02	4.8	Silty Gravel and Sand	31	1	2
947	53	1	8	355	5.01	0.43	358.00	0.02	5.3	Silty Gravel and Sand	145	0	2
948	33	1	21	3079	9.06	0.19	358.00	0.05	5.5	Silty Gravel and Sand	15	0	1
949	64	1	8	297	9.3	0.30	358.00	0.02	5.5	Silty Gravel and Sand	31	0	1

950	36	1	10	435	8.39	0.20	358.00	0.02	8.2	Silty Gravel and Sand	59	0	2
951	18	2	8	294	9.69	0.29	357.98	0.02	7.0	Silty Gravel and Sand	8	2	1
952	31	1	15	1085	12.67	0.14	357.56	0.05	5.6	Silty Gravel and Sand	31	1	2
953	61	1	8	406	6	0.56	357.50	0.05	5.5	Silty Gravel and Sand	84	0	2
954	61	1	8	463	4	0.73	357.50	0.05	5.5	Silty Gravel and Sand	84	0	2
955	31	2	8	322	7.28	0.35	357.48	0.02	5.5	Silty Gravel and Sand	31	0	1
956	66	1	8	597	3.13	1.21	357.20	0.17	5.3	Silty Gravel and Sand	145	0	2
957	18	2	8	300	6.71	0.31	357.19	0.02	7.3	Silty Gravel and Sand	31	3	1
958	31	2	8	0	5.9	0.00	357.14	0.02	7.3	Silty Gravel and Sand	31	3	1
959	40	1	15	307	9.87	0.01	357.11	0.02	4.8	Silty Gravel and Sand	31	1	2
960	33	1	8	341	5.58	0.39	357.00	0.02	5.5	Silty Gravel and Sand	31	0	1
961	54	1	8	351	6.26	0.42	357.00	0.02	5.3	Silty Gravel and Sand	145	0	2
962	22	2	12	753	11.65	0.22	357.00	0.02	5.5	Silty Gravel and Sand	31	0	1
963	55	1	12	708	6	0.00	357.00	0.02	4.6	Silty Gravel and Sand	15	1	2
964	55	1	8	342	5.37	0.40	357.00	0.02	5.3	Silty Gravel and Sand	145	0	2
965	55	1	8	674	5.8	1.54	357.00	0.02	5.5	Silty Gravel and Sand	31	0	1
966	55	1	8	952	5	3.08	357.00	0.02	7.3	Silty Gravel and Sand	31	3	1
967	55	1	8	420	4.22	0.60	357.00	0.02	5.3	Silty Gravel and Sand	145	0	2
968	35	1	8	926	6.67	2.91	357.00	0.02	5.5	Silty Gravel and Sand	31	0	1
969	60	1	8	296	2.08	0.30	356.95	0.02	7.3	Silty Gravel and Sand	31	3	1
970	18	2	8	351	10.16	0.42	356.94	0.02	7.0	Silty Gravel and Sand	8	2	1
971	43	1	8	343	4.5	0.40	356.81	0.02	5.5	Silty Gravel and Sand	31	0	1
972	43	1	8	454	13.92	0.70	356.66	0.05	5.5	Silty Gravel and Sand	84	0	2
973	66	1	12	3488	8.59	4.75	356.60	0.24	5.0	Silty Gravel and Sand	61	2	1
974	40	1	8	355	3.8	0.43	356.60	0.02	5.3	Silty Gravel and Sand	145	0	2
975	54	1	8	342	6.58	0.40	356.60	0.05	5.5	Silty Gravel and Sand	84	0	2
976	38	1	8	308	4.28	0.32	356.57	0.05	5.5	Silty Gravel and Sand	84	0	2
977	54	1	8	423	4.72	0.61	356.54	0.02	5.3	Silty Gravel and Sand	145	0	2
978	54	1	8	314	6.41	0.33	356.40	0.02	7.3	Silty Gravel and Sand	31	3	1
979	33	1	8	299	8.29	0.00	356.37	0.02	4.8	Silty Gravel and Sand	31	1	2
980	13	2	8	388	6.74	0.51	356.34	0.02	7.3	Silty Gravel and Sand	31	3	1
981	55	1	8	307	5.15	0.32	356.30	0.02	7.0	Silty Gravel and Sand	69	0	1
982	16	2	8	442	3.26	0.66	356.25	0.10	5.1	Silty Gravel and Sand	31	0	2
983	55	1	8	312	5	0.33	356.12	0.02	5.5	Silty Gravel and Sand	31	0	1

984	57	1	8	400	8.56	0.54	356.10	0.10	5.3	Silty Gravel and Sand	8	0	2
985	53	1	8	633	5.12	1.36	356.09	0.10	5.3	Silty Gravel and Sand	8	0	2
986	43	1	8	516	6.6	0.00	356.07	0.02	4.8	Silty Gravel and Sand	130	0	2
987	53	1	8	345	6.78	0.40	356.01	0.02	5.5	Silty Gravel and Sand	31	0	1
988	68	1	8	748	5.22	1.90	356.00	0.02	5.3	Silty Gravel and Sand	145	0	2
989	42	1	8	343	8.23	0.40	356.00	0.05	5.5	Silty Gravel and Sand	84	0	2
990	23	2	8	302	10.93	0.31	356.00	0.02	5.8	Silty Gravel and Sand	15	3	1
991	34	1	8	310	7.07	0.33	356.00	0.02	5.5	Silty Gravel and Sand	31	0	1
992	30	2	8	404	6.18	0.55	355.96	0.02	6.6	Silty Gravel and Sand	31	0	2
993	56	1	8	540	5.49	0.99	355.90	0.02	5.3	Silty Gravel and Sand	145	0	2
994	54	1	8	519	4.4	0.92	355.80	0.02	5.3	Silty Gravel and Sand	145	0	2
995	7	2	8	335	3.84	0.38	355.80	0.02	4.8	Silty Gravel and Sand	31	1	2
996	45	1	10	516	13.42	0.28	355.79	0.05	5.5	Silty Gravel and Sand	84	0	2
997	54	1	12	700	11.26	0.19	355.70	0.05	5.5	Silty Gravel and Sand	84	0	2
998	61	1	8	419	7.18	0.60	355.69	0.02	5.5	Silty Gravel and Sand	31	0	1
999	46	1	24	5441	18.49	0.29	355.68	0.02	5.5	Silty Gravel and Sand	31	0	1
1000	53	1	8	306	5.6	0.32	355.60	0.02	5.5	Silty Gravel and Sand	31	0	1