ASSET MANAGEMENT OF WASTEWATER

INTERCEPTORS ADJACENT TO BODIES OF WATER

by

MOHAMMAD DAMEN BANI FAWWAZ

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfilment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2022

Dedication

This work is dedicated to my beloved family for their endless support. There are no words to express how grateful I am to them. I will always be grateful to you and owe you.

Acknowledgments

I would like to express my gratitude to Dr. Mohammad Najafi, PE, F. ASCE, Associate Professor and Director of the Center for Underground Infrastructure Research & Education (CUIRE), Department of Civil Engineering, The University of Texas at Arlington for his support, patience, and encouragements.

I am grateful to all my Ph.D. dissertation committee members for their insightful remarks, recommendations, and comments: Dr. Seyed Mohsen Shahandashti, Dr. Jay Michael Rosenberger, and Dr. Kyeong Rok Ryu.

Furthermore, many thanks to the City of Fort Worth for providing me with the wastewater system data that was the core of this research. I would like to express my gratitude to Mark Shell, Technical Services Coordinator, Water Department, and Sonnie M. LaForce, former Strategic Operations Supervisor, Development Services Department, for accepting and approving the use of wastewater pipeline data.

Abstract

ASSET MANAGEMENT OF WASTEWATER
INTERCEPTORS ADJACENT TO BODIES OF WATER

Mohammad Bani Fawwaz, Ph.D.

The University of Texas at Arlington, 2022

Supervising Professor: Mohammad Najafi

This dissertation presents challenges of managing a pipeline network adjacent to bodies of water to maximize asset life by evaluating the significant factors that affect the condition levels of pipeline assets. Pipeline asset management derives from pipelines' physical conditions, condition rating, and serviceability through investigating, monitoring, and analyzing rupture history. The remaining asset life and structural condition of the pipeline network running near and under bodies of water are often hard to predict. In case of a pipeline failure, major damages may occur to the surrounding environment, adding up to disruptions in service and repairing costs. This research develops Multinomial Logistic Regression (MLR) and Binary Logistic Regression models to predict how the bodies of water could affect the soil surrounding wastewater interceptors. The models were developed based on data from the City of Fort Worth, Texas. This dissertation concludes that pipe diameter, pipe age, location of the pipeline with reference to bodies of water (far or near), and the pipe material are the most significant variables that affect the surrounding conditions and remaining life of wastewater interceptors. In future, clearer perception through increased software development and machine learning for managing pipeline asset management would provide impacts of different parameters on pipeline expected life.

Table of Contents

CHAPTER 1 INTRODUCTION AND BACKGROUND 1
1.1 Background and Overview 1
1.2 Pipeline Condition Assessment 2
1.2.1 Phases of Condition Assessment Projects 2
1.2.2 Factors impacting a pipeline’s service life 3
1.2.3 Asset Management Strategies 3
1.3 Research Needs 4
1.4 Objectives 5
1.5 Scope of Work 5
1.6 Methodology 6
1.7 Hypotheses 6
1.8 Timeline for Completion 8
1.9 Chapter Summary 8
CHAPTER 2 LITERATURE REVIEW 9
2.1 Pipeline Asset Management 9
2.1.1 Asset Management Inspection Cycle 10
2.2 Condition Assessment and Rehabilitation 11
2.2.1 City of Fort Worth Wastewater Interceptors Condition Assessment 12
2.3 Pipelines Buried Underwater 14
2.3.1 Underwater Pipeline Inspection 14
2.3.2 Hydrological Surveying Technologies 15
2.4 GIS-Based Engineering Pipeline Management 16
2.4.1 GIS Database and Asset Management 16
2.5 Pipes’ Soil Interactions 17
2.6 Chapter Summary 18
CHAPTER 3 LOGISTIC REGRESSION 19
3.1 Data Sets and Model Setting 19
3.2 Logistic Regression 20
3.2.1 Assumptions for Logistic Regression Modeling 21
3.2.2 Binary Logistic Regression 21
3.2.3 Multinomial Logistic Regression 24
3.3 Chapter Summary 25

CHAPTER 4 CASE STUDY AND DATA ANALYSIS **26**
4.1 Background Information of Wastewater Interceptors 26
4.2 Data Collection 28
4.3 Data Preparation 28
4.3.1 Elevation Differences 29
4.3.2 Model Variances 31
4.4 Descriptive Statistics 33
4.5 Chapter Summary 34
CHAPTER 5 MODEL DEVELOPMENT **35**
5.1 Introduction 35
5.2 Multinominal Logistic Regression 35
5.2.1 Model Clarification 35
5.2.2 Parameters Estimation 36
5.2.3 Model Significance 40
5.3 Binary Logistic Regression 43
5.3.1 Model Clarification 43
5.3.2 Parameters Estimation 44
5.3.3 Model Significance 46
5.4 Chapter Summary 46
CHAPTER 6 RESULTS, CLASSIFICATION, AND DISCUSSION **47**
6.1 Binary Logistic Regression 47
6.1.1 Classification 47
6.1.2 True vs. False and Positive vs. Negative 49
6.1.3 Accuracy 50
6.1.4 ROC (Receiver Operating Characteristic) 50
6.1.5 Sensitivity and Specificity 51
6.2 Significant Variables 53
6.3 Insignificant Variables 53
6.4 Soil Erosion 54
6.5 Discussions 55
6.6 Practical Applications 58
6.7 Chapter Summary 59
CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH **60**
7.1 Conclusions 60

7.2 Dissertation Limitations 60
7.3 Recommendations for Future Research 61
REFERENCES 62

List of illustrations

Figure 1-1 Factors Impacting a Pipeline’s Service Life 3

Figure 1-2 Research Methodology 7

Figure 1-3 Research Timeline 8

Figure 2-1 Elements of Comprehensive Asset Management................... 9

Figure 2-2 Asset Management Inspection Cycle..................................... 11

Figure 2-3 Categories for Assessing Wastewater System Criticality 12

Figure 2-5 Triple Bottom Line Approach 17

Figure 3-1 Prediction Model Influencers.. 19

Figure 3-2 Logistic Regression Categories 20

Figure 4-1 The City of Fort Worth Wastewater Pipeline Diameters 27

Figure 4-2 The City of Fort Worth Service Area Elevations 2010 29

Figure 4-3 The City of Fort Worth Service Area Elevations 2015 30

Figure 4-4 The City of Fort Worth Pipelines Material 31

Figure 4-5 The City of Fort Worth Pipelines Installation Date 32

Figure 4-6 The City of Fort Worth Pipeline Age Frequencies 32

Figure 5-1 Model Parameters.. 36

Figure 5-2 MLR Overall Correct Percentages 42

Figure 6-1 Binary Logistic Overall Correct Percentages 48

Figure 6-2 True vs. False and Positive vs. Negative................................ 49

Figure 6-3 Significant Variables... 53

List of Tables

Table 1-1 Asset Management Strategies 4

Table 1-2 Scope of Work 6

Table 2-1 Costs and Benefits of Condition Assessment Program 13

Table 2-2 ICAP Condition Scores and RUL 14

Table 2-3 Underwater Sonar Inspection Technologies 15

Table 2-4 Hydrological Surveying Technologies 15

Table 3-1 Classification Table for Binary Regression (General Form)..... 23

Table 3-2 Classification Table for MLR 25

Table 4-1 Wastewater System in the City of Fort Worth 26

Table 4-2 The City of Fort Worth Descriptive Statistics............................ 28

Table 4-3 Descriptive Statistics for Binary Logistic Regression 33

Table 4-4 Descriptive Statistics for Multinomial Logistic Regression 34

Table 5-1 Dependent Variable Levels 35

Table 5-2 Surrounding Conditions Level 1 Parameters 37

Table 5-3 Surrounding Conditions Level 2 Parameters 38

Table 5-4 Surrounding Conditions Level 3 Parameters 39

Table 5-5 The Significance of Variables.. 40

Table 5-6 Multinomial Logistic Regression Significance 40

Table 5-7 Multinomial Classification Table 42

Table 5-8 Binary Logistic Regression Model Variables............................ 44

Table 5-9 Parameters Estimation for Binary Regression Model 45

Table 5-10 Parameters Estimation for Binary Backward Stepwise 46

Table 5-11 Significance for Binary Logistic Regression Model 46

Table 6-1 Classification for Binary Logistic Regression Model 48

Table 6-2 Confusion Matrix for Binary Logistic Regression Model 49

Table 6-3 Binary Logistic Regression Model Performance 52

Table 6-4 Models' Variables Comparison 56

Table 6-5 Models' Accuracy Comparison 56

Table 6-6 Binary Logistic Regression Variables' Performance 57

Chapter 1 Introduction and Background

1.1 Background and Overview

Utility and pipeline systems form one of the most capital-intensive infrastructure systems, and they are aging, overused, possibly mismanaged, and neglected (Najafi, 2022). Most wastewater systems are gravity systems; flow is transferred by natural forces rather than complicated pumping technology. The United States' wastewater network consists of over 800,000 miles of public sewers and 500,000 miles of private lateral sewers that connect homes and businesses to public sewer lines. The typical lifespan expected for wastewater pipes is 50 to 100 years (ASCE, 2021).

Condition assessment is an ongoing process. The biggest challenges of maintaining wastewater systems are that the process is out of view (EPA, 2015). The latest 2021 infrastructure report card published by the American Society of Civil Engineering (ASCE) reveals incremental progress toward restoring our nation's infrastructure. For the first time, our infrastructure GPA went up from D+ in 2017 to C- in 2021 (ASCE, 2021). Furthermore, most municipal sewer systems are at least 60 years old, and some utilities assume that newer pipes must be in good condition compared to older pipes, which is not the case since many examples show 80-year-old pipes in excellent condition and 30-years-old pipes near failure (EPA, 2015).

An estimation of how much pipe of each size in each region must be repaired and rehabbed in the coming 40 years is compiled by combining the demographically based pipelines inventories with the projected service lifetime for each region (AWWA, 2012). The effects associated with pipeline failures can be extended to impact other infrastructures, so many utilities have adopted new technologies in pipeline asset management to enhance proactive asset management strategies (Matthews et al., 2016). Moreover, the U.S. utilities must meet all National Pollutant Discharge Elimination System (NPDES) permit requirements and innovative Geographic Information System (GIS) cloud-based data combined with mapping technologies within utility asset management planning to begin the next step of risk analysis based on condition assessment (Harris, 2017).

Many pipelines that cross under bodies of water are buried deep underneath the soil must regularly be inspected and evaluated. Once the pipeline conditions are available, asset management, repair, and rehabilitation decisions will be made (Flynn et al., 2018).

Asset management strategies start with reviewing the available historical pipeline data and understanding failure and deterioration models (EPA, 2012). Repair and rehabilitation decisions control the continual performance of pipeline systems. A proactive asset management system will overwhelm the reactive system to stay within the cost-effective choices and keep the system at an acceptable level.

Local municipalities use geographical information systems (GIS) for archival, revenue, and information retrieval purposes, but the use of GIS varies among municipalities within each state. Effective asset management requires evaluating pipeline systems and identifying pipelines with a high risk of failure. A geographic information system (GIS) data set consisting of pipe age, length, material, and previous repairs will allow municipalities to make asset management decisions while continuously updating the GIS data set (Nardini et al., 2013).

1.2 Pipeline Condition Assessment

1.2.1 Phases of Condition Assessment Projects

Condition assessment projects typically have four phases: preliminary investigations, field investigations, integrity assessments, and post-processing condition assessments. Generally, various tools and techniques will be used since no single tool can provide all the required information for condition assessment. Once the main pipeline details such as diameter, length, age, and failure history are available, technique selection will be uncomplicated (Mahaffey, 2016).

1.2.2 Factors impacting a pipeline's service life

The ability of the pipeline to carry external and in-service loads forms the pipeline's structural integrity. Pipeline structural integrity must be assessed in condition assessment to determine the level of deterioration. For example, pipeline material could react with the environment, causing corrosion that can vary along the pipeline. The corrosion mechanism could act entirely differently inside and outside the pipe (Mahaffey, 2016). Figure 1-1 illustrates water seepage and movement of soil when the pipeline is installed by trenchless technology methods and is under or near bodies of water.

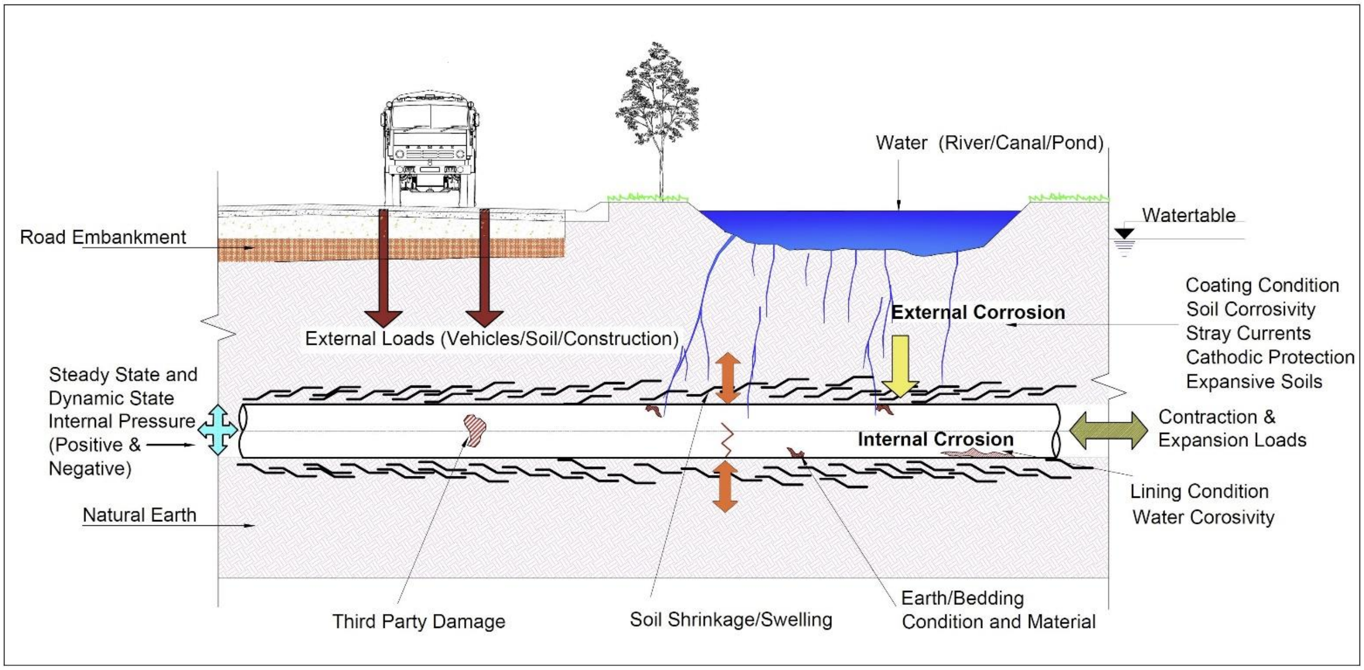


Figure 1-1 Factors Impacting a Pipeline's Service Life

1.2.3 Asset Management Strategies

Infrastructure asset management is the continual assessment of the operations and maintenance history and projected life expectancy, with a long-range plan for financing asset rehabilitation or replacement (R&R); it results in prioritizing infrastructure assets and incorporating assets into the annual capital improvement planning (AWWA, 2012).

Condition assessment will enable municipalities to understand the current structural condition of pipelines and implement the predictive level strategy (Ugarelli et al., 2008). The traditional asset management strategies are operative (reactive), inspection (condition-based), proactive (preventive), and predictive (advanced). Each strategy has a specific role in the asset management methodologies series. In general, asset management strategies are four main categories, as shown in Table 1-1.

Table 1-1 Asset Management Strategies (Ugarelli et al., 2010)

Operative (reactive)	• Municipalities often make decisions based on practical experience. • It is emergency repairs and rehabilitation. • In a simple approach, the pipe section will consume its full-service life. • It causes interruptions in traffic and service.
Inspection (condition-based)	• Municipalities monitor pipelines periodically. • Pipelines are classified based on their condition rating. • It recognizes the current pipeline condition without failure consequences.
Proactive (preventive)	• Repair and rehabilitation are done before failure. • It needs more time to choose the best cost-effective repair.
Predictive (advanced)	• Cities provide economic analysis support to the proactive approach. • It gives the availability to choose between regular maintenance and rehabilitation. • It indicates long-term implications on life cycle cost.

1.3 Research Needs

The focus of this dissertation is based on buried wastewater interceptors' asset management adjacent to bodies of water. Wastewater assets have long life cycles. Furthermore, a wide assortment of studies has been done to demonstrate asset management of wastewater pipelines. The following recent research highlights the needs for inspection and monitoring of pipelines:

- Harris (2017) encouraged municipalities to enhance affordable resources such as GIS in conducting asset management planning to plan well-conceived projects properly.
- Sever et al., (2017) indicated that surrounding soil condition has a vital role in the pipeline loads, which is more important to expose than the visible.
- Wade (2016) recommended that wastewater utilities use timely information technologies to address the most critical infrastructure needs since inspecting and rehabilitating large-diameter wastewater systems is expensive.
- Loganathan (2021) recommended that integrating GIS during inspection of pipe segment would help map the critical pipelines and condition assessment.

1.4 Objectives

The main objective of this dissertation is to evaluate the life of wastewater interceptors considering the long-term impacts of surrounding soil conditions for operational and maintenance tasks.

The secondary objective is to evaluate the significant factors that affect the condition levels of assets. Furthermore, to compare the wastewater interceptors surrounding soil elevations from 2010 through 2015. The comparison will be between wastewater interceptors adjacent (less than 10 ft) to bodies of water and the interceptors away (more than 10 ft) from bodies of water.

1.5 Scope of Work

The scope of this study is according to Table 1-2.

Table 1-2 Scope of Work

Included	Not Included
Wastewater interceptors within the city of Fort Worth service area	The sewer force main pipes are not considered
Evaluation will be based on the surrounding conditions	The stormwater pipes will not be investigated
Selected wastewater interceptors will be based on recommendations from the City of Fort Worth where the assets repairs and rehab are anticipated	Soil type
--	Pipe installation method
	Watertable

1.6 Methodology

The models developed in this study will be used to link dissimilarity between wastewater interceptors near and away from bodies of water by considering physical and environmental factors.

The following steps present an approach to developing the outcome of this research. Figure 1-2 presents the detailed research methodology.

- Step 1: Problem definition
- Step 2: Literature Review
- Step 3: Data collection
- Step 4: Data analysis
- Step 5: Model development
- Step 6: Model validation
- Step 7: Model comparison
- Step 8: Select the best model based on the results
- Step 9: Asset management strategy recommendations

1.7 Hypotheses

Based on the available historical data, it is conceived that bodies of water have significant effects on the soil elevation that surrounds the wastewater interceptors.

Moreover, pipe inspection and surrounding soil elevations for wastewater interceptors adjacent to bodies of water will need a different asset management strategy.

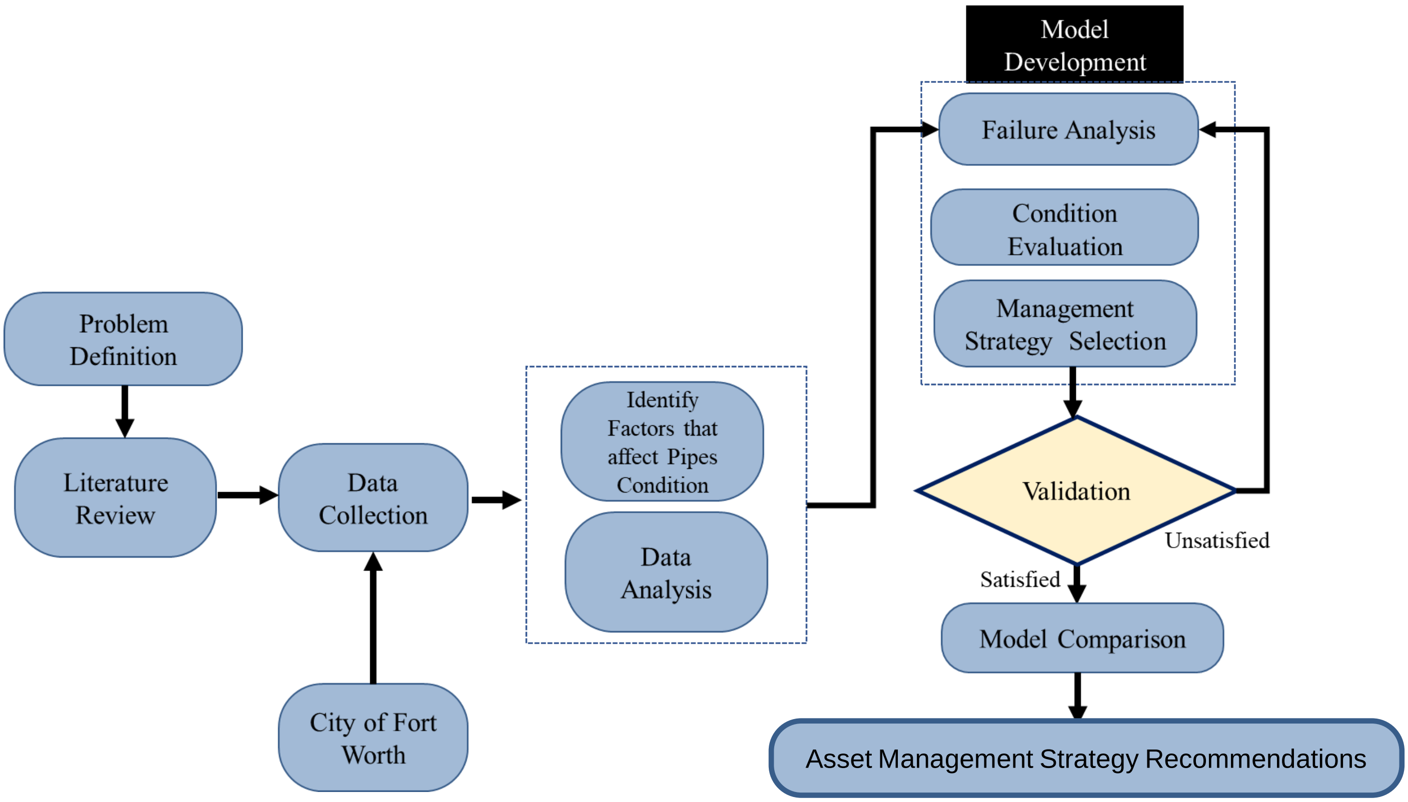


Figure 1-2 Research Methodology

1.8 Timeline for Completion

Figure 1-3 presents the timeline of the dissertation completion.

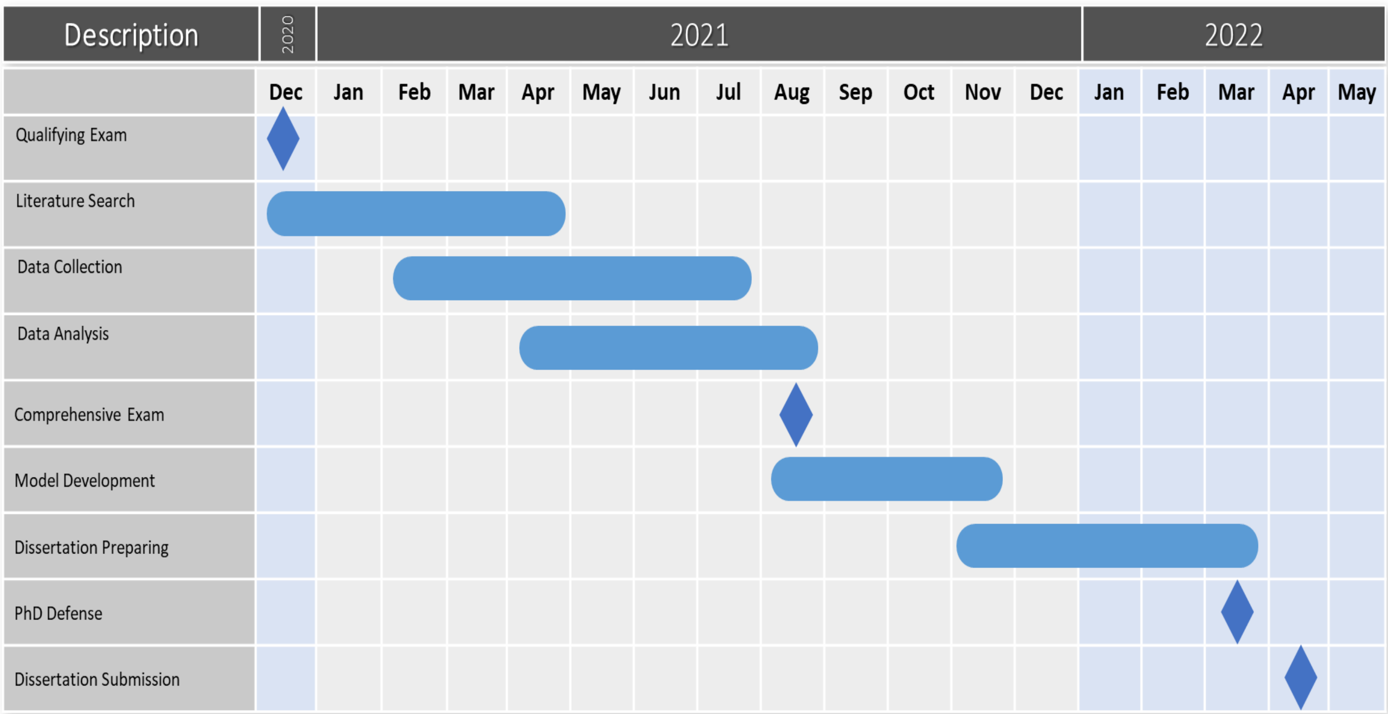


Figure 1-3 Research Timeline

1.9 Chapter Summary

This chapter discussed background information about wastewater systems and major asset management strategies. Research needs, objectives, scope of work, methodology, hypotheses, and the research schedule were provided.

Chapter 2 Literature Review

2.1 Pipeline Asset Management

The International Infrastructure Management Manual (INGENI-UM 2002) has defined asset management as: "The combination of management, financial, economic, engineering, and other practices applied to physical assets to provide the required level of service in the most cost-effective manner."

The wastewater pipeline industry continuously changes like other industries, which created the need for improvements in repair and rehabilitation efforts. Municipalities have adapted proactive pipeline asset management to reduce pipes' failures (Matthews et al., 2016). Figure 2-1 illustrates the main components of comprehensive asset management.

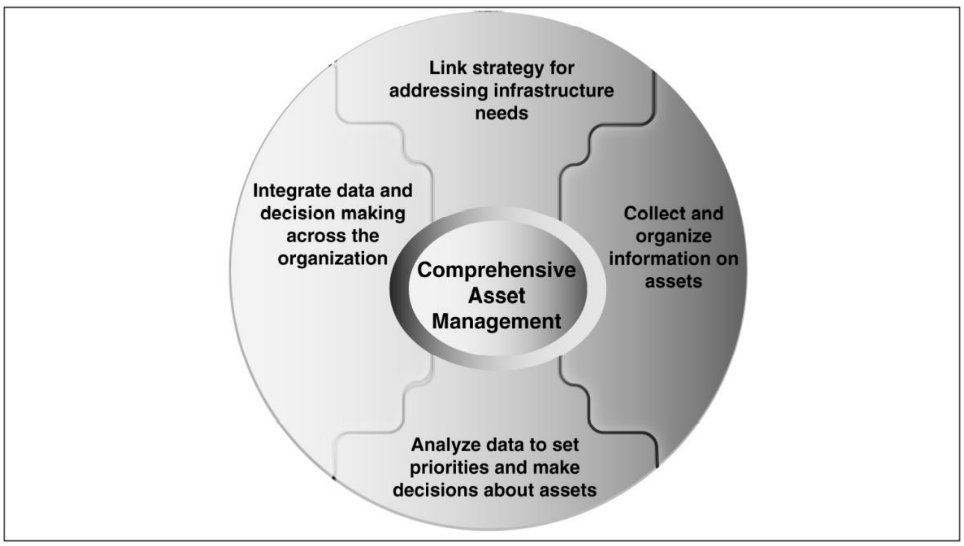


Figure 2-1 Elements of Comprehensive Asset Management (GAO, 2004)

According to the Government Accountability Office (GAO), the federal government spends billions of dollars to help municipalities finance wastewater infrastructure projects, raising concerns about the infrastructure's conditions and asset management plans. Comprehensive asset management will allow utility managers to have accurate information about the existing assets, such as pipe age, performance, required repairs and service, and condition (GAO, 2004).

However, evaluating pipe management strategies with Envision pre-assessment checklist resulted in three main strategies as below:

- A run-to-failure strategy is recommended only if the pipe failure consequences are minimal with financial shortfalls.
- A pre-emptive replacement strategy is shared and reduces the impact of pipe failures. Pipe replacement should be based on the actual condition; otherwise, some pipes will be replaced even when they have a remaining useful life.
- A balanced approach strategy, repair, and rehabilitation decisions will be based on the pipe condition factor (Matthews et al., 2016).

Asset management is a comprehensive plan for managing infrastructure assets to deliver a satisfying service level and minimize operating and ownership costs. A comprehensive asset management plan can help municipalities turn from a reactive approach into a proactive approach while providing life cycle cost analysis based on cost-benefit analysis (Najafi and Gokhale, 2022; EPA, 2002).

2.1.1 Asset Management Inspection Cycle

A stepwise asset management plan allows municipalities to follow top-down and bottom-up cycles, as shown in Figure 2-2. The plan will start with the assembly’s current condition (Tier 0), then decisions will be made to collect additional data (Tier 1). Once Tier 1 is completed, the outcome will be post-processed in the top-down phase of the cycle. Assets with a high rank based on Tier 1 will be selected for Tier 2 inspection. After that, the bottom-up and top-down cycles will be repeated until sufficient information has been reached to identify the municipalities’ asset register (Wade, 2016).

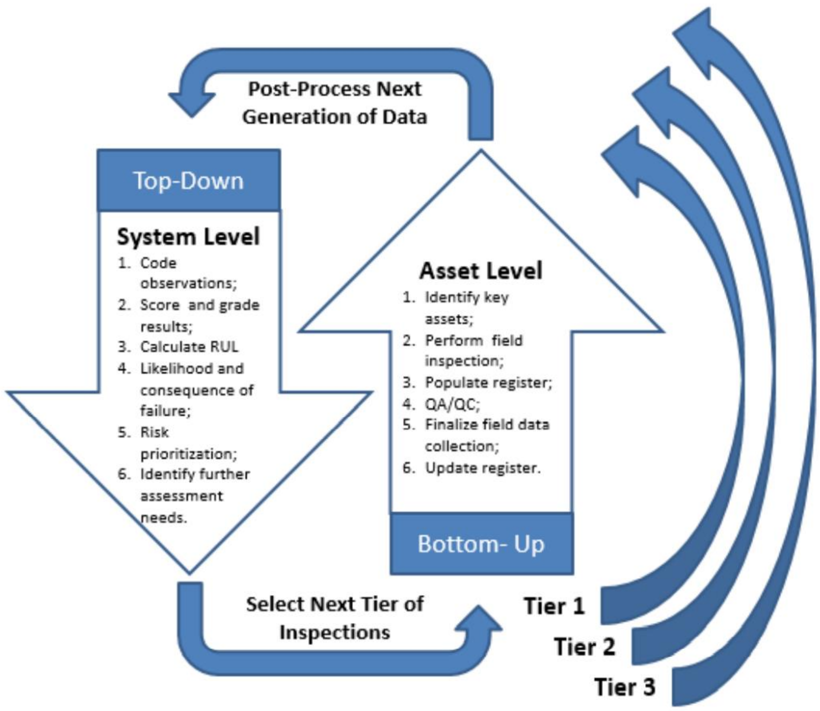


Figure 2-2 Asset Management Inspection Cycle (Wade, 2016)

2.2 Condition Assessment and Rehabilitation

Most wastewater systems are 60 years old, and majority are out of sight underground. Therefore, it is valuable to invest in condition assessment to reduce emergency repairs and replacements (EPA, 2015).

Condition assessment is based on the current pipe condition; a criticality score for pipes will be formulated based on the probability of failure, consequences of failure, and redundancy (Najafi, 2016). The primary parameters for consequences of failure are environmental and transportation impact, flow quantity, and cost of urgent repairs. Meanwhile, the primary parameters for the probability of failure are pipe age, hydraulic capacity, soil conditions, and loads (Sever et al., 2017).

Pipeline fractures in any location will have deleterious impacts on transportation, level of service, and the environment. Criticality rating is assigned to the asset as a one-time event and will help municipalities focus on high-risk and priority pipes (EPA, 2015).

The standard criteria categories for assessing the criticality of wastewater systems are shown in Figure 2-3.

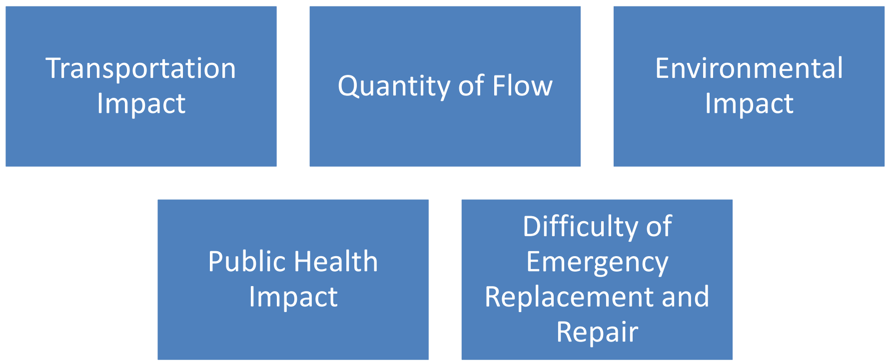


Figure 2-3 Categories for Assessing Wastewater System Criticality (EPA, 2015)

2.2.1 City of Fort Worth Wastewater Interceptors Condition Assessment

The City of Fort Worth, Texas, wastewater interceptor system has 262 miles of pipelines with diameters between 24 inches to 96 inches. The newly installed pipe materials are fiberglass, direct plastic burial, and liners. Meanwhile, the original pipe materials are vitrified clay and reinforced concrete. The City of Fort Worth Interceptor Condition Assessment Program (ICAP) is one of the world's most significant multi-sensor interceptor condition assessment projects. The program assessment techniques used are sonar, laser profiling, and high-definition TV inspection (Thornhill and Crumb, 2014).

According to EPA (2015), as for any project, condition assessment of wastewater system has costs and benefits. Table 2-1 presents some of the common cost parameters and benefits.

Table 2-1 Costs and Benefits of Condition Assessment Program (EPA, 2015)

Costs of Condition Assessment Program	**Benefits of Condition Assessment Program**
Equipment costs	Less expensive repairs when identifying problems early
Labor costs	Reduce operating and maintenance costs
Service distribution costs when conducting the inspection	Reduce emergency consequences and costs
–	More effective operation and maintenance
	Prediction of needed capital renewal

The three technologies used for ICAP inspection are High-Definition CCTV, sonar, and 3-D laser. High-Definition CCTV is an advanced technology for capturing video images with high-resolution quality. Sonar inspection technology is used for below the water level pipes; an image is created based on sonar signals that show broken pipes, deflection, and debris. 3-D laser inspection is a technology that creates a three-dimensional pipe model. The sonar technology data is combined with the laser data to develop a three-dimensional model for the whole pipe, then HD video is integrated to get a high-resolution pipe picture and a three-dimensional model for the pipe (Thornhill and Crumb, 2014).

The City of Fort Worth has been evaluating its wastewater interceptors within the Interceptor Condition Assessment Program (ICAP) since 2010; the data collected has been compared to standard respective pipe classes until pipe corrosion extent has been determined. After that, a rank scale between 1 and 5 (best to worst) was assigned to pipe segments, as shown in Table 2-2 (Kercho and Conlon, 2019).

Table 2-2 ICAP Condition Scores and RUL (Kercho and Conlon, 2019)

Score	**Pipe Condition**	**Remaining Useful Life (RUL)**
1	A material loss of 0 to 0.5 inches from the original inside wall	36 to 50 years
2	A material loss of 0.5 inches to the interior face of the first row of reinforcement steel	21 to 35 years
3	Material loss from the interior face of the first row of reinforcement steel to half the distance to the internal face of the second row of reinforcement steel	11 to 20 years
4	Material loss from half the distance to the interior face of the second row of reinforcement steel to the internal face of the second row of reinforcement steel	3 to 10 years
5	The interior face of the second row of reinforcement steel to the outer pipe wall surface	less than two years

2.3 Pipelines Buried Underwater

There are millions of miles of pipelines in the United States of America, and thousands of these pipelines cross under bodies of water (Flynn, et al. 2018). Assessing wastewater systems is challenging when the pipelines cross waterways and have a high risk of failure; these pipelines are usually not easy to manage since we have limited access (Forsyth, et al. 2018).

2.3.1 Underwater Pipeline Inspection

Underwater acoustic imaging (sonar) is a comprehensive concept involving diverse technologies that provide pipeline visual documentation. Sonar technologies are categorized based on the type of outcome data into two-dimensional sonar systems and three-dimensional sonar systems, as shown in Table 2-3 (Forsyth, et al. 2018):

Table 2-3 Underwater Sonar Inspection Technologies (Forsyth, et al. 2018)

Three-Dimensional Sonar Systems	**Two-Dimensional Sonar Systems**
Single-Beam Sonar	Side-Scan Sonar
Geophysical Sub-Bottom Profilers	Sector-Scanning Sonar
Multi-Beam Sonar	LiDAR (Laser Scanning)
Real-Time Multi-Beam Sonar	--

2.3.2 Hydrological Surveying Technologies

There is a remarkable difference between inspecting underground and underwater pipelines when the underground underwater pipeline leaks or needs inspection. Challenges such as restricted equipment and access to pipelines are present. However, different technologies could be used to overcome these challenges.

Table 2-4 Hydrological Surveying Technologies (Flynn et al., 2018)

ShapeAccelArrays (SAA)

- SAA is a geotechnical sensors that records pipelines deformation, it needs excavation so as to be installed before backfilling. ShapeAccelArrays are made up of rigid segments connected by flexible joints, every rigid segment is made up of three orthogonally mounted tilt sensors and one microprocessor to calculate its position in XYZ plane according to the segment length and measurements from tilt sensors.

Pipeline Inspection Gauges (PIG's)

- PIG is an advance inspection technology that could be used for corrosion inspections, pipe mapping, and deformation inspections. PIGs consist of sealant disks, sensors, GPS trackers, and data canisters. PIGs have to be inserted into closed and sealed pipes.

SmartBall

- SmartBall is an alternative technology to PIG, it can detect leakage, pipe mapping, temperature, and pressure. SmartBall needs two access point for insertion and extraction and collect data up to 21 hours. The difference between PIG and SmartBall is travelling ability, SmartBall is limited for up 3,000 ft meanwhile, PIG can travel to greater distance up to battery life.

To inspect underground underwater pipelines, diverse technologies of hydrological surveying could be used, such as ShapeAccelArrays (SAA), Pipeline Inspection Gauges (PIGs), and SmartBall technology which are briefly explained in Table 2-4 (Flynn et al., 2018).

2.4 GIS-Based Engineering Pipeline Management

A geographic information system (GIS) is a platform that satisfies the core engineering function’s needs. Hence, most municipalities have employed GIS for general administrative purposes such as data archival and information diffusion (Venigalla et al., 2007). A GIS containing water pipeline data such as pipe age, diameter, length, pipe material, and condition assessment contributes to asset management by aggregating data and promoting utilities to make risk-based statements (Nardini et al., 2013). Asset GIS-based management will develop a comprehensive water pipeline system view to enhancing planning (Nardini et al., 2015).

2.4.1 GIS Database and Asset Management

Maps can determine the installation date to reveal the pipeline age, and reports will reveal pipe sections and rehabilitation location history. However, considerable information is needed to develop a database associated with the planning process. Likelihood of failure (LOF) and consequence of failure (COF) are essential to establish the asset GIS database condition. Through the GIS process, relevant information could be scanned and organized.

The National Association of Sewer Service Companies (NASSCO) suggests that their Version 7 Pipeline Assessment and Certification Program (PACP) be utilized with the Triple Bottom Line Approach, as shown in Figure 2-5.

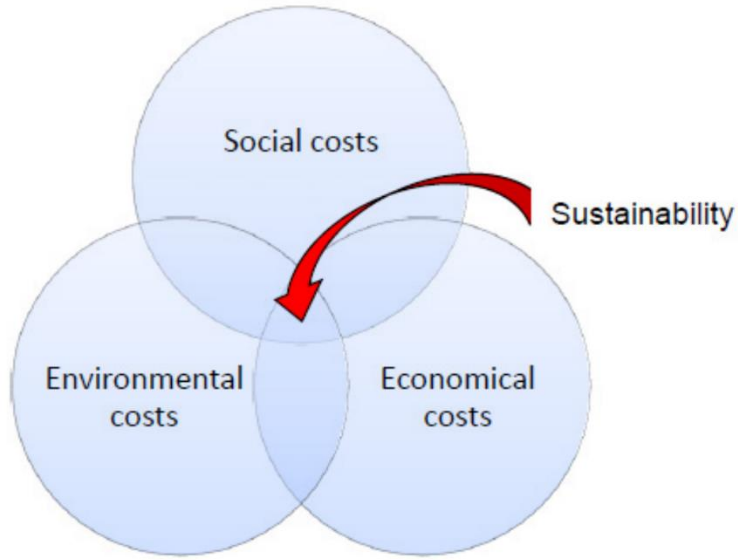


Figure 2-5 Triple Bottom Line Approach (NASSCO PACP V 7.0, 2015)

Economic costs comprise the costs of designing and conducting repairs; hence, it varies based on the pipe diameter, type of pipe, depth, pipe length, and the environmental costs surrounding the asset. Social costs comprise the indirect costs such as: customers affected, road closures, and public relations. Environmental costs comprise the costs of waterway pollution and public health costs. The three cost structures overlap (Harris, 2017).

2.5 Pipes' Soil Interactions

Pipeline damage caused by landslides is prevalent in different areas, where a continual monitoring and repair efforts are planned to ensure their serviceability (Calvetti and Di Prisco, 2004). Soil movement and erosion can significantly affect the infrastructures' service time. To analyze the pipes' resistance to such soil changes, it is necessary to quantify the interaction between the pipelines and the surrounding soil.

The primary design recommendations for pipelines provide a bilinear force-displacement relationship curve for the soil-pipeline interaction. However, Trautmann and O'Rourke's (1983) real experimental results indicated that the force steadily dropped when the relative displacement between soil and pipe was 0.1 m in the case of thick sand for backfill (Yoshizaki and Sakanoue, 2004).

2.6 Chapter Summary

Municipalities need a comprehensive proactive asset management plan to keep their assets with time and cost savings. Pipelines' location has a vital role in developing the management plan.

Chapter 3 Logistic Regression

3.1 Data Sets and Model Setting

In this chapter, binary logistic regressions and multinomial logistic regressions are discussed. Usually, municipalities and utility management use the available statistical data to evaluate the current condition and predict the future of their assets. Prediction models play a pivotal role in providing municipalities' asset management foundation. It is not a one-time project; the models are continually validated to reflect the data set changes. Logistic regression is a model developed using statistical data sets interpreted as independent and dependent variables and determined to be numerical and nominal variables.

Understanding the future of pipe networks necessitates the use of a thorough model (Malek Mohammadi, 2019). this dissertation develops a prediction model based on the statistical data set; the selection and validation for the model are influenced by several factors, as shown in Figure 3-1.

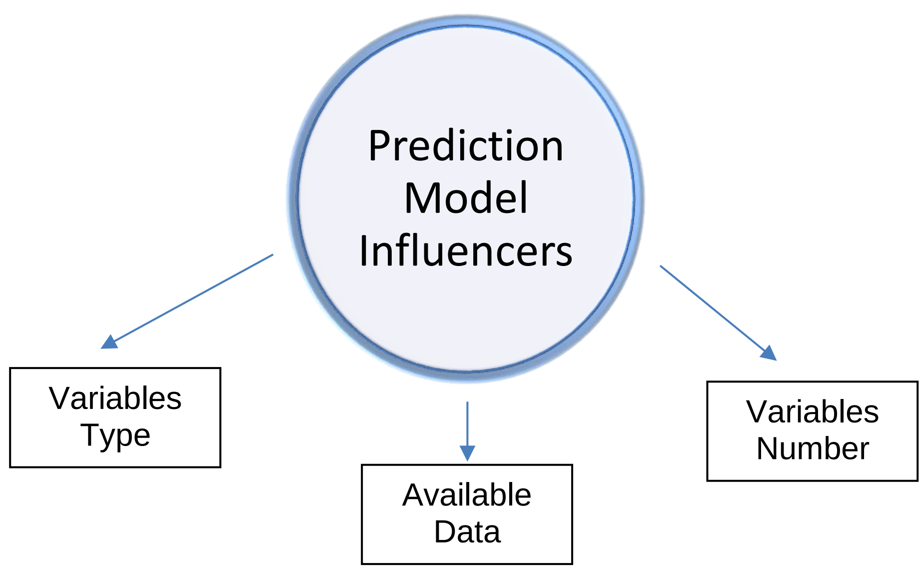


Figure 3-1 Prediction Model Influencers

3.2 Logistic Regression

Logistic regression is a statistical procedure that assists in determining the relationship between a dependent variable and a set of independent variables (Gall et al., 2007).

Logistic regression has three main categories, as shown in Figure 3-2 (Park, 2013). This dissertation model was developed using binary and multinomial logistic regression since ordinal logistic regression requires the statistical data to be categorical (data has natural ordered categories).

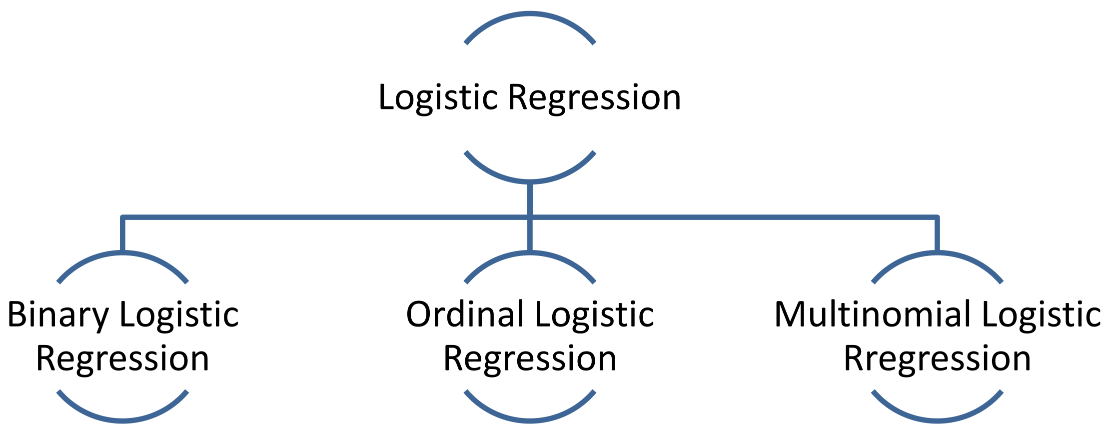


Figure 3-2 Logistic Regression Categories

Fundamentally, the logistic regression model does not have any specific conditions for the variables other than the model to be applicable. Moreover, the model is significantly based on the p-value for the case.

A logistic model gives an account of the relationship between a dependent and a set of variables. (Khashei and Bijari, 2010). According to Hawari, et al (2020), the statistical model for logistic regression is shown in Eq. 3.1.

$$\log\left(\frac{\pi}{1-\pi}\right)Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad \text{Eq. 3.1}$$

Where:

Y = dependent variable

α = intercept parameter

β_n = regression coefficients associated with p independent variables.

Probability of (y =1) determined using exponential transformation.

π = p (y = 1| x_1 …. x_n)

The logistic regression analysis helps determine which independent variables are the most significant in the dependent variable.

3.2.1 Assumptions for Logistic Regression Modeling

Binary and multinomial logistic regression share the same assumptions. Comprehensively, the binary logistic regression dependent variable should be binary. Meanwhile, the multinomial logistic regression model is developed when the dependent variable has three or more values. Logistic regression assumptions are not complicated. According to Meyers et al. (2006), the assumptions are as follows:

- Between independent variables, perfectly multicollinearity should not exist, which means independent variables are not highly correlated with each other.
- The model does not have specification errors.
- A linear relationship between continuous independent variables and the dependent variable transformation logit should exist.

3.2.2 Binary Logistic Regression

Logistic regression differentiates from linear regression when the outcome variable is binary (Hosmer Jr et al., 2013). The dependent variable must be only two different values (e.g., 0 and 1) regarding the binary logistic regression. If the dependent variable is categorical or numerical, the corresponding variable must be dummy coded into two values before employing the binary logistic regression.

Binary logistic regression generates a model to account for relationships between log odds of the dependent and independent variables (Hosmer and Lemeshow, 2000). According to Agresti (2007), Eq. 3.2 is the linear logistic regression model logit form:

$$\text{logit}\,[\pi(X)] = \log\left(\frac{\pi(X)}{1-\pi(X)}\right) = \alpha + \beta_x \qquad \text{Eq. 3.2}$$

If the model has multiple independent variables, Eq. 3.3 will be used (Hawari et al., 2020).

$$log\left(\frac{\pi}{1-\pi}\right) = log\left[\frac{P(Y=1 \mid x_1 \ldots. x_n)}{1-P(Y=1 \mid x_1 \ldots. x_n)}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad \text{Eq. 3.3}$$

Where:

α = intercept parameter

β_n= regression coefficient associated with n independent variables

$$P(Y = 1 \mid x_1 \ldots. x_n) = \frac{e^{\alpha+\sum_{j=1}^{n}\beta j x j}}{1+e^{\alpha+\sum_{j=1}^{n}\beta j x j}} \qquad \text{Eq. 3.4}$$

$$\pi(X) = \frac{exp\,(\alpha+\beta_x)}{1+exp\,(\alpha+\beta_x)} \qquad \text{Eq. 3.5}$$

3.2.2.1 Coefficients Significance

3.2.2.1.1 Log-Likelihood Test

Log-likelihood is used to determine whether the binary model is significant. In this methodology, the model is developed by including the variable of interest. By eliminating that variable, followed by comparing those data sets by using a chi-square distribution corresponding to the degree of freedom equals the number of eliminated variables. Meanwhile, if the independent variable is categorical and takes on more than one value, the degree of freedom will be the number of categorical values minus one (Salman 2010).

Typically, the log-likelihood test compares the predicted to observed values as in Eq. 3.6.

$$G = -2\ln\left(\frac{\text{likelihood without the variable}}{\text{likelihood with the variable}}\right) \qquad \text{Eq. 3.6}$$

3.2.2.1.2 Wald Test

Wald test is another available strategy to check if a determined coefficient of the developed binary logistic regression model is significant or not, using Eq. 3.7 and Eq. 3.8.

$$W = \left(\frac{\beta i - \beta o}{\text{Standard Error }(\beta i)}\right) \qquad \text{Eq. 3.7}$$

Usually, the parameter of interest is 0 (βo=0); the formula simplifies to

$$W = \left(\frac{\beta i}{\text{Standard Error }(\beta i)}\right) \qquad \text{Eq. 3.8}$$

Where βi presents the predictor variable coefficient.

3.2.2.2 Classification Table and Verification

After developing the binary logistic regression model, it must be verified, and the correct predicted values are calculated. Classification tables are one of the available techniques to verify the developed model. Table 3-1 presents a general form of the classification table.

Table 3-1 Classification Table for Binary Logistic Regression (General Form)

Observations	Predictions	
	0	**1**
0	A_{11}	A_{12}
1	A_{21}	A_{22}

A cut-off value is determined and then compared to the estimated probability. If it is greater than the cut-off value, it is assigned to class one. Otherwise, it will get a class zero. Usually, the cut-off value for the binary dependent variable is 0.5.

According to Salman (2010), the percentage of correct predictions is calculated based on a classification table using the formula Eq. 3.9.

$$\text{Percentage of correct predictions} = \frac{100(A_{11}+A_{22})}{(A_{11}+A_{12}+A_{21}+A_{22})} \qquad \text{Eq. 3.9}$$

3.2.3 Multinomial Logistic Regression

Multinominal logistic regression is a statistical tool used to develop a relationship between dependent and independent variables. Multinomial logistic regression is used if the dependent variable has more than two values (Hawari et al., 2020).

$$log\left(\frac{\pi}{1-\pi}\right) = log\left[\frac{P(Y=i|x_1,\dots,x_n)}{1-P(Y=k|x_1,\dots,x_n)}\right]$$

$$= \alpha + \beta_{i1}x_1 + \beta_{i2}x_2 + \dots + \beta_{in}x_n \qquad \text{Eq. 3.10}$$

Where:

α = intercept parameter for category i

β_{in}= regression coefficient associated with category i

i = 1, 2, 3, …., k-1

K = possible values associated with dependent variable

$x_1,\dots,x_n$ = independent variables

3.2.3.1 Coefficients Significance

Log-likelihood test and Wald test are common in determining the significance of the variables (Hosmer et al., 2013). Coefficient significance for multinomial logistic regression could be determined using the same techniques used in binary logistic regression.

Meanwhile, the multinomial logistic regression model has more than one equation, indicating that one variable could be significant in one equation rather than all of them. In other words, log-likelihood has an advantage on the Wald test.

3.2.3.2 Classification Table and Verification

The classification table for the multinomial logistic regression model is equivalent to binary logistic regression; the only difference is the number of values in the table. Since multinomial logistic regression has more than one equation, the classification table values will be more as in Table 3-2.

Table 3-2 Classification Table for Multinomial Logistic Regression

Observations	Predictions				
	1	**2**	**…**	**…**	**n**
1	A_{11}	A_{12}	…	…	A_{1n}
2	A_{21}	A_{22}	…	…	A_{2n}
…	…	…	…	…	…
…	…	…	…	…	…
n	A_{n1}	A_{n2}	…	…	A_{nn}

The percentage of correct predictions is as shown in Eq. 3.11.

$$\text{Percentage of correct predictions} = \frac{100(A_{11}+A_{22}+\ldots+A_{nn})}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\text{AijAij}\right)} \qquad \text{Eq. 3.11}$$

3.3 Chapter Summary

Multinominal logistic regression and binary logistic regression could be used to develop a prediction model based on the available historical data, and both share the same conditions and assumptions.

Chapter 4 Case Study and Data Analysis

4.1 Background Information of Wastewater Interceptors

The City of Fort Worth has been systematically evaluating its sanitary sewer interceptors as part of the Interceptor Condition Assessment Program (ICAP) (Thornhill and Crumb, 2014). ICAP was based on the pipe materials and conditions; landscape and surrounding conditions were not considered.

Based on condition assessment inspection data, statistical models build relationships between known pipe variables and wastewater pipelines conditions (Atambo, 2021). This research considers the wastewater interceptors and surrounding landscape elevations. It is based on the data collected from the City of Fort Worth, Texas, United States. The wastewater system constitutes 3,519 miles of pipelines with different diameters, pipe materials, and installation dates. Based on the City of Fort Worth data generated from the ArcGIS layers, Table 4-1 summarizes the wastewater system in the City of Fort Worth.

Table 4-1 Wastewater System in the City of Fort Worth

Discipline	**Commentary**
Miles of the wastewater system	3,519 miles
Range of pipe diameter	2 – 96 inches
Majority pipe size	6 in and 8 in
Burial depth range	0.50 – 29.64 ft
Pipe Material	Asbestos-cement (AC) Cast iron (CI) Corrugated metal pipe (CMP) Ductile iron pipe (DIP) Fiberglass pipe (FRP)

(continued)

Discipline	Commentary
	Prestressed concrete cylinder pipe (PCCP) High-density polyethylene (HDPE) Polyvinyl chloride (PVC) Reinforced concrete pipes (RCP) Vitrified clay (VC)
Majority of the Pipes	PVC ~ 58%

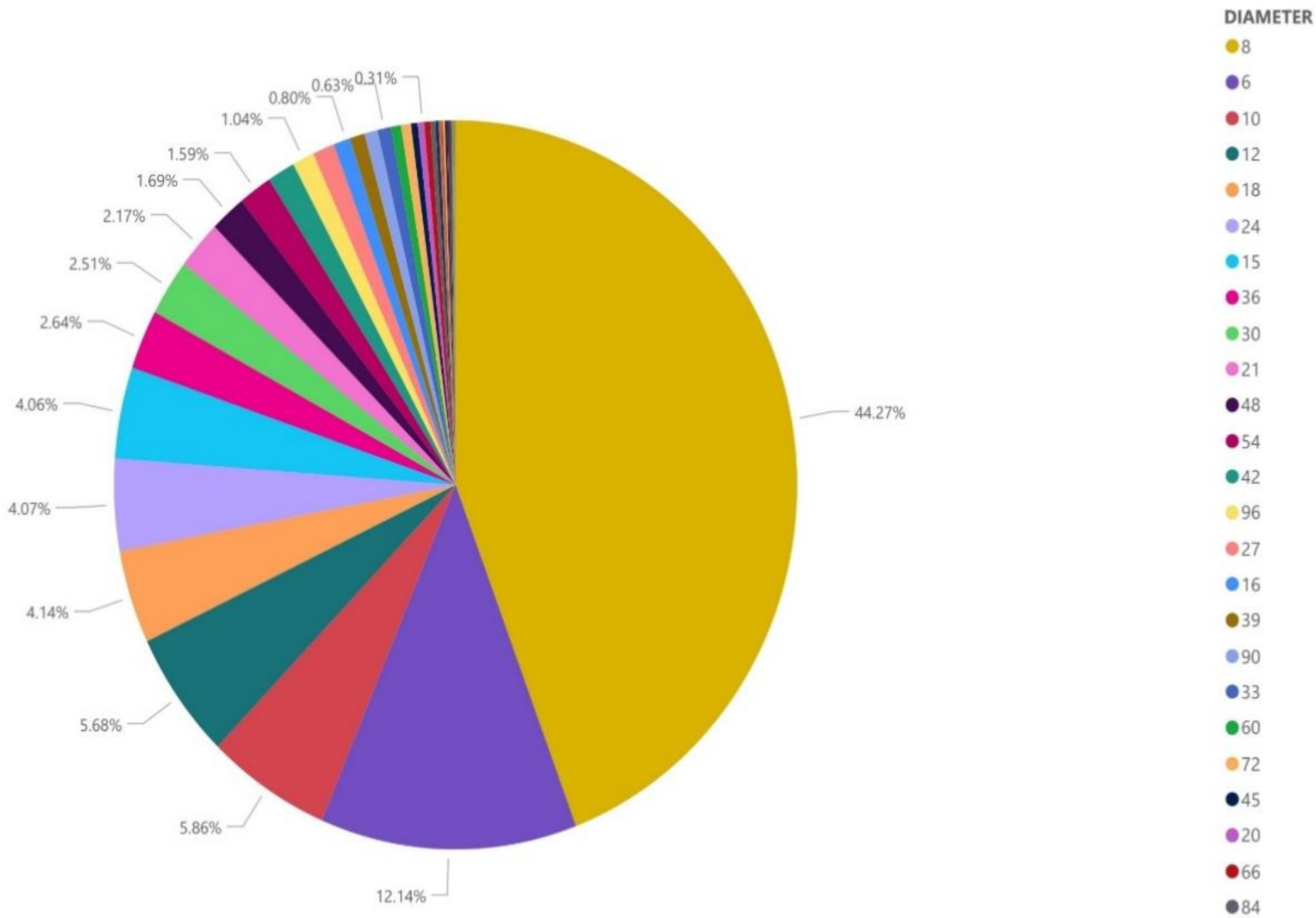


Figure 4-1 The City of Fort Worth Wastewater Pipeline Diameters

The City of Fort Worth wastewater pipeline diameters of 6-inch and 8-inch pose for 56.41% of the entire wastewater system. Meanwhile, the wastewater interceptors with pipeline diameters range from 24-inch to 96-inch from around 262 miles, as shown in Figure 4-1.

In the 1990s, the City of Fort Worth started taking wastewater systems' maintenance and renovation seriously under the Administrative Order (AO) from EPA. The City of Fort Worth has gone through various failures, especially for large wastewater interceptors.

4.2 Data Collection

The City of Fort Worth has broad dataset layers compatible with ArcGIS software created and periodically updated. In this research, the wastewater interceptors, and the surrounding elevation differences between the years 2010 and 2015 are the basis of the model. Table 4-2 represents the descriptive statistics for the data collected from the City of Fort Worth.

Table 4-2 The City of Fort Worth Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Installation Date	64528	1906	2018	1988.56	23.309
Pipe age	64528	3	115	32.44	23.309
Elevation Difference	64528	-2.99999770	3.999999099	.4927443750	2.024463392
DIAMETER	64528	1.25	96.00	10.0595	7.92412
Length	64528	.3299448462	6128.329095	260.9049112	200.8608658
Valid N (listwise)	64528				

4.3 Data Preparation

The first step of data preparation was combining the LiDAR layers for the two years and wastewater interceptors' layers with spatial and different functions.

4.3.1 Elevation Differences

Figures 4-2 and 4-3 show the elevation difference maps for 2010 and 2015.

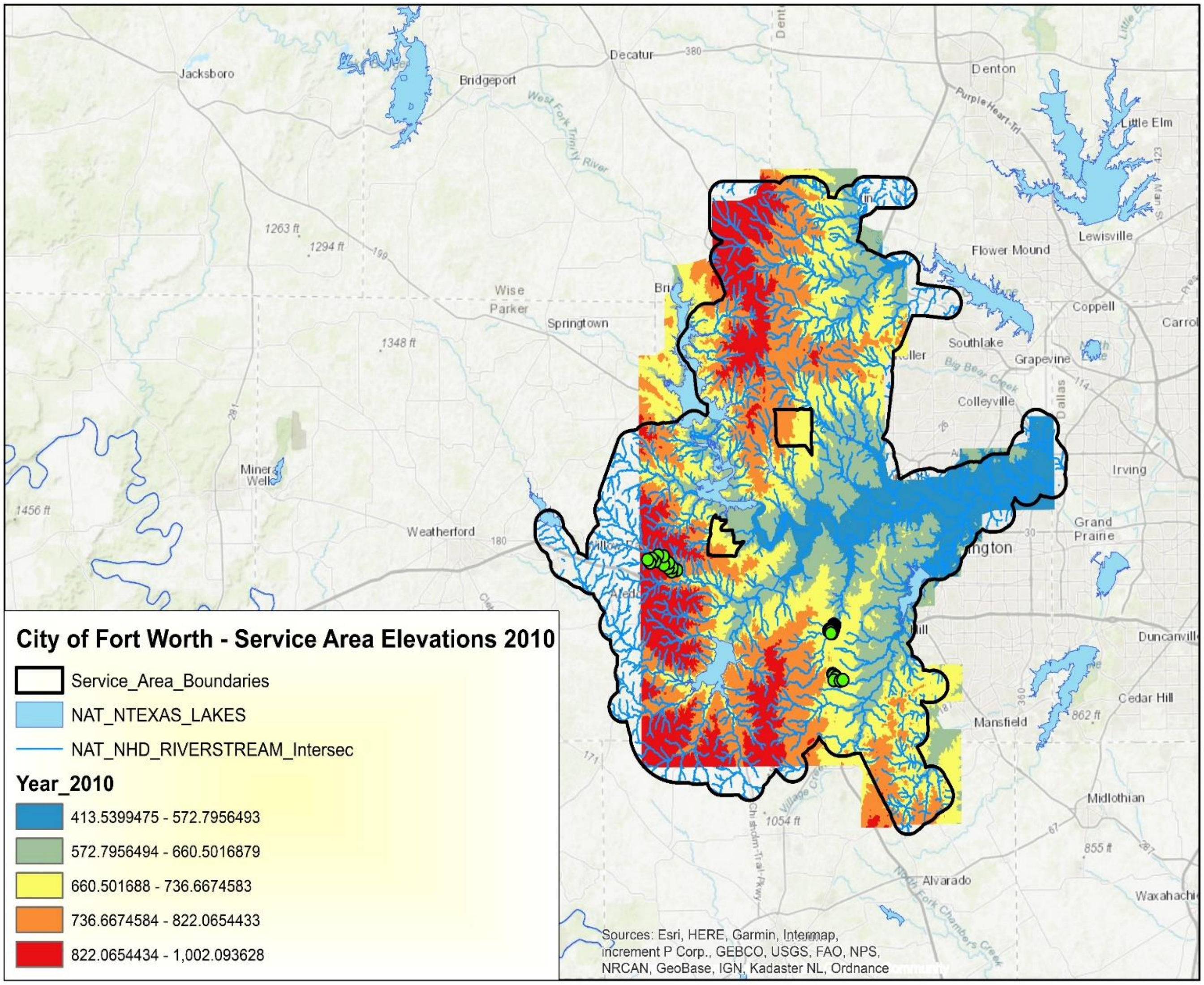


Figure 4-2 The City of Fort Worth Service Area Elevations 2010

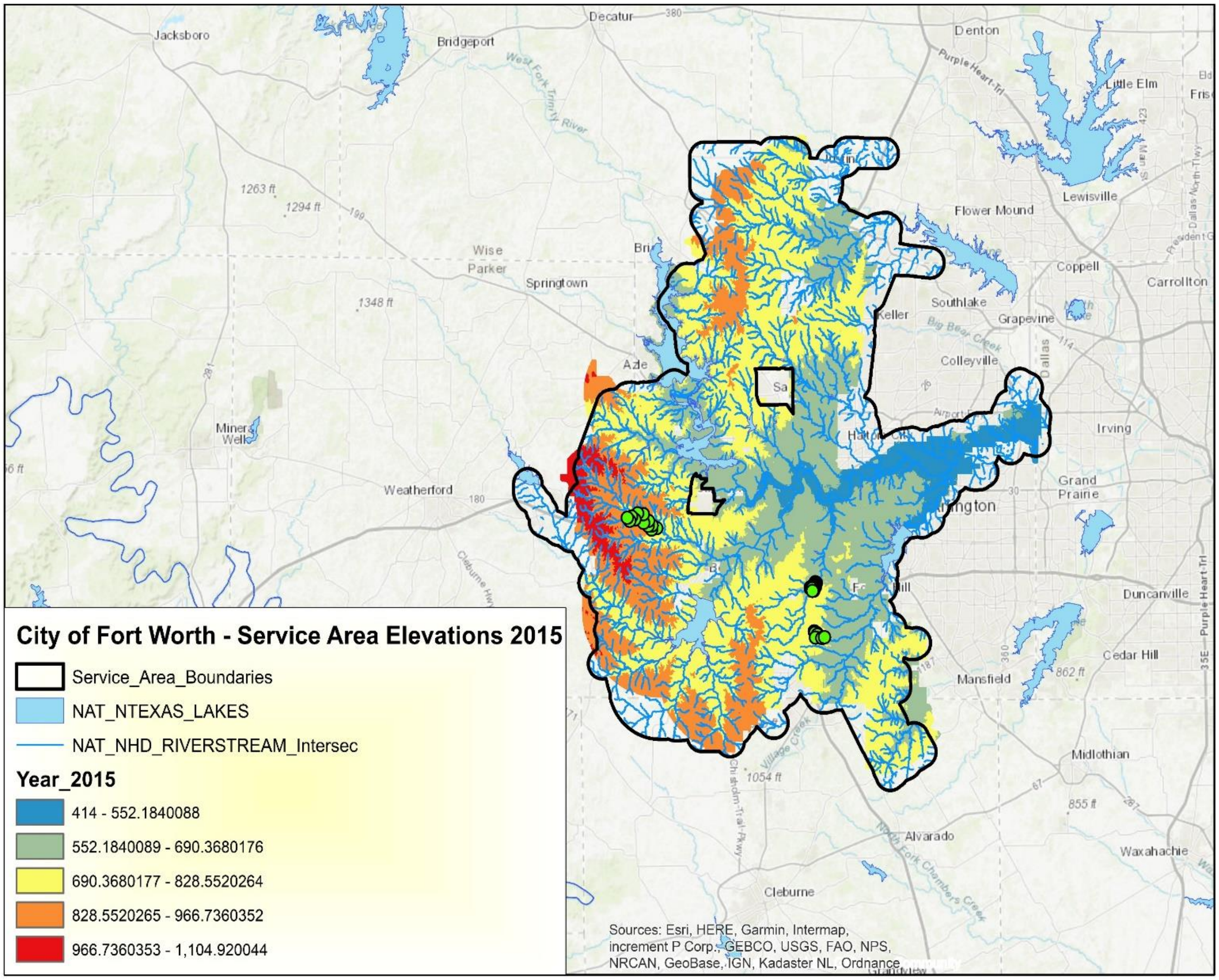


Figure 4-3 The City of Fort Worth Service Area Elevations 2015

4.3.2 Model Variances

The next step will be splitting the layers to get two separate datasets; one will be near bodies of water (rivers and lakes), and the second data set will be for wastewater interceptors away from the bodies of water. Then, the elevation difference between the years 2010 and 2015 is calculated.

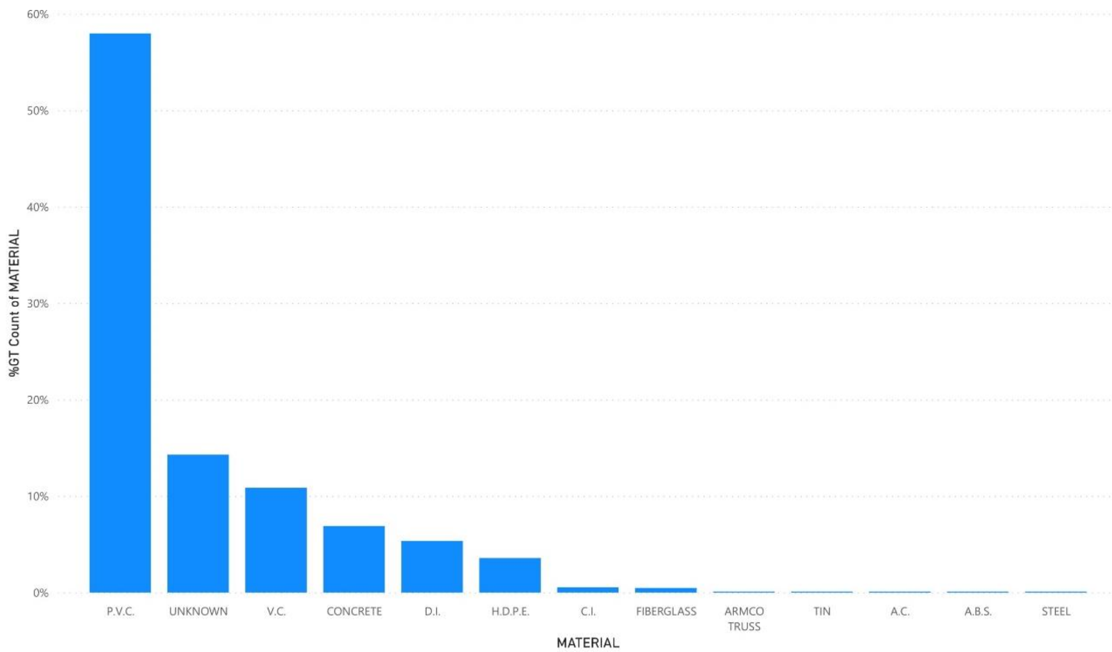


Figure 4-4 The City of Fort Worth Pipelines Material

Figure 4-4 shows all the pipe materials included in the City of Fort Worth GIS database. Moreover, the installation date for the pipelines was between the years 1948 and 2000, and the highest percentage of overall pipeline installation date was for the year 2000 with about 6%, as shown in Figures 4-5 and 4-6.

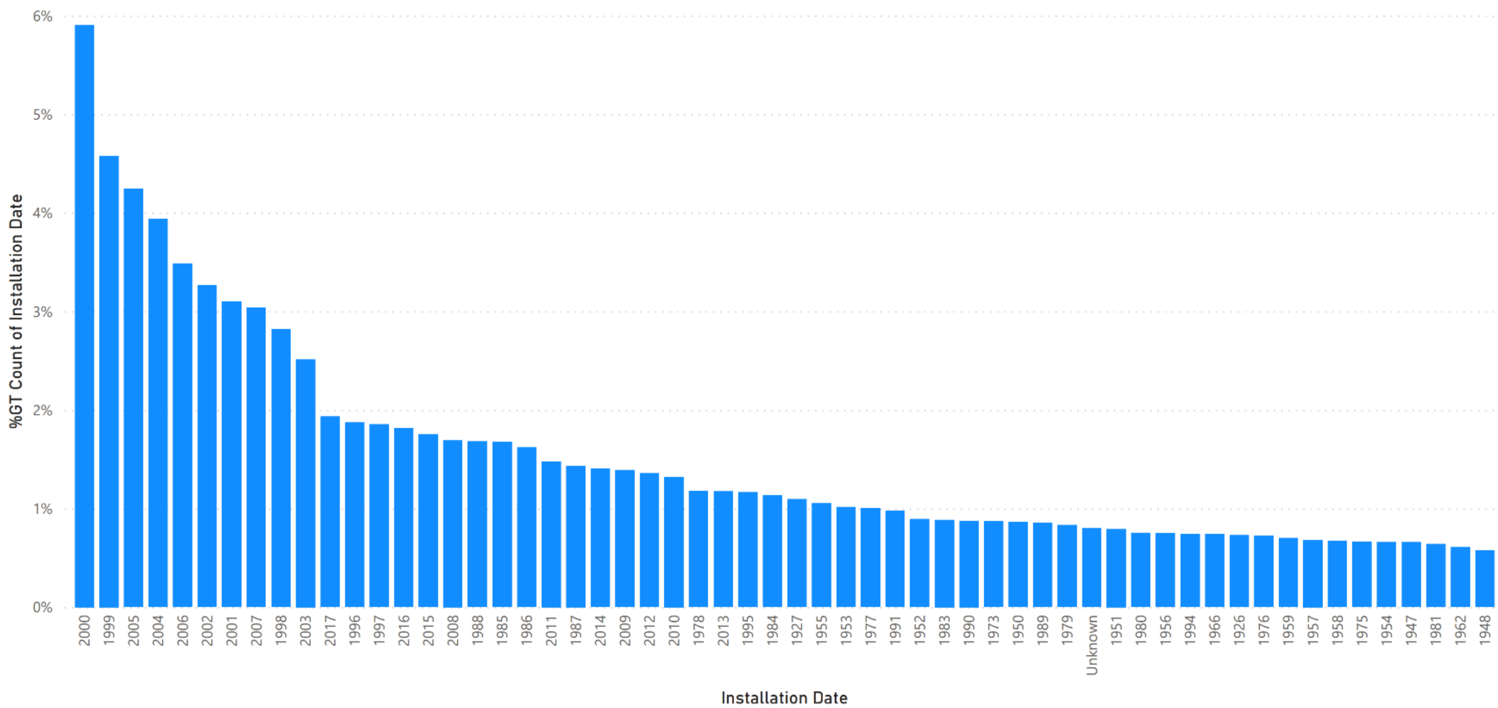


Figure 4-5 The City of Fort Worth Pipelines Installation Date

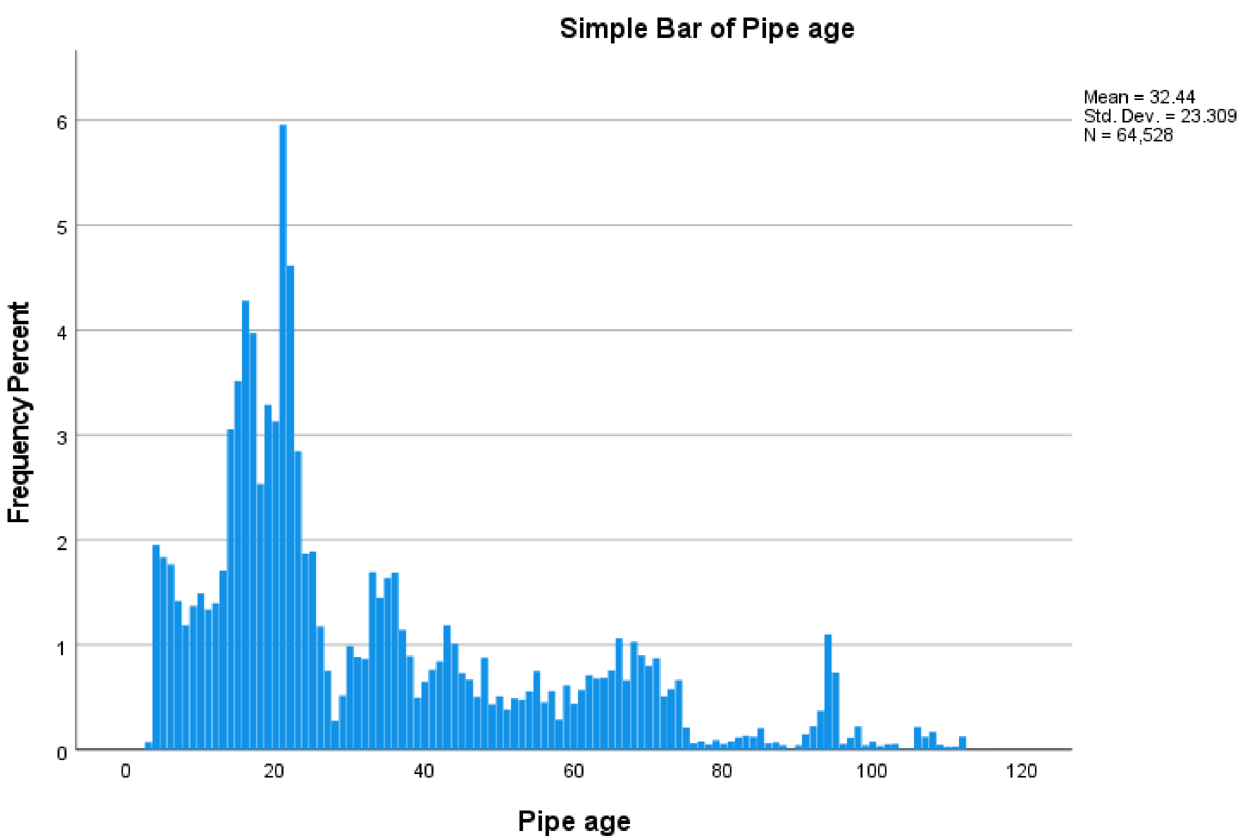


Figure 4-6 The City of Fort Worth Pipeline Age Frequencies

Some pipeline variables are continuous quantitative such as pipe age, pipe diameter, and surrounding elevation difference for the years 2010 and 2015. However, pipe material is nominal categorical, and pipe diameter is discrete quantitative. Missing and duplicate data were eliminated for model development.

The wastewater interceptors were coded into 1 and 0 based on the location; 1 for wastewater interceptors near bodies of water and 0 for wastewater interceptors far from bodies of water. Moreover, the elevation difference between 2010 and 2015 was coded as 1 if the elevation increases and 0 if the elevation decreases.

4.4 Descriptive Statistics

Tables 4-3 and 4-4 show the descriptive statistics based on the regression used.

Table 4-3 Descriptive Statistics for Binary Logistic Regression

	N	Minimum	Maximum	Mean	Std. Deviation
Installation Date	3191	1913	2017	1983.61	19.702
Pipe age	3191	4	108	37.39	19.702
Elevation Difference	3191	-.289885513633174	.289923677390742	-.002233952010206	.131130433686905
Elevation Decrease 0 / Increase 1	3191	0	1	.50	.500
Diameter	3191	24	96	35.87	15.805
Length	3191	1.203040382538900	5009.428335736829000	420.452949897716560	402.190348506222450
Far 0 / Near 1	3191	0	1	.40	.489
Valid N (listwise)	3191	--	--	--	--

Table 4-4 Descriptive Statistics for Multinomial Logistic Regression

	N	Minimum	Maximum	Mean	Std. Deviation
Installation Date	3191	1913	2017	1983.61	19.702
Pipe age	3191	4	108	37.39	19.702
Elevation Difference	3191	-.289885513633174	.289923677390742	-.002233952010206	.131130433686905
Elevation Decrease 0 / Increase 1	3191	0	1	.50	.500
Diameter	3191	24	96	35.87	15.805
Length	3191	1.203040382538900	5009.428335736829000	420.452949897716560	402.190348506222450
Far 0 / Near 1	3191	0	1	.40	.489
Surrounding Conditions	3191	1	4	2.39	1.116
Valid N (listwise)	3191				

4.5 Chapter Summary

This chapter discussed the data acquisition, collection, preparation, descriptive statistics, and data processing before the development of the model phase starts. The database was generated just from The City of Fort Worth GIS database. The logistics regression for developing the model will be followed using SPSS Statistics software.

Chapter 5 Model Development

5.1 Introduction

The development of multinomial logistic regression and binary logistics regression models is demonstrated in this chapter. IBM SPSS Statistics was used to develop the models using 80% of the available data; the rest 20% of the data were used for the model validation phase.

5.2 Multinominal Logistic Regression

5.2.1 Model Clarification

The multinomial logistic model was developed to find the relationship between a nominal dependent variable with multiple independent variables.

The independent variables were pipe age, pipe diameter, pipe material (Nominal categorical; PVC (Polyvinyl chloride), VC (Vitrified clay), Concrete, Steel, DI (Ductile Iron), HDPE (High-density polyethylene), and CI (Cast Iron)). The dependent variable was nominal with four levels, as shown in Table 5-1.

Table 5-1 Dependent Variable Levels

Pipe Surrounding Conditions Rating	Pipe location with reference to bodies of water	Soil elevation difference over the years 2010 to 2015
1	Far	Decrease
2	Near	Decrease
3	Far	Increase
4	Near	Increase

Equation 5.1 represents the general multinomial logistic regression formula when including all the model independents:

$$ln \left(\frac{P(C = i)}{P\ (C = 4)} \right) =$$

$$\alpha_i + \beta_{i1} \times \text{Age} + \beta_{i2} \times \text{Diameter} + \beta_{i3} \times D_{\text{Material=HDPE}}$$
$$+ \beta_{i4} \times D_{\text{Material=CI}} + \beta_{i5} \times D_{\text{Material=DI}} + \beta_{i6} \times D_{\text{Material=PVC}}$$
$$+ \beta_{i7} \times D_{\text{Material=Steel}} + \beta_{i8} \times D_{\text{Material=VC}} + \beta_{i9} \times D_{\text{Material=Concrete}} \qquad \text{Eq. 5.1}$$

Where:

α: intercept

i: pipe surrounding conditions rating

$\beta i1, \beta i2,\ldots, \beta i9$: regression coefficients

D: dummy variable used to assign values to categorical independent variables

5.2.2 Parameters Estimation

SPSS statistics software was used to develop the multinominal logistic regression model using 80% of the data. Pipe surrounding conditions rating 4 was the reference for the model. Figure 5-1 shows the model parameters estimation methodology with the variable's significance tests, which output is indicated in Table 5-2, Table 5-3, and Table 5-4.

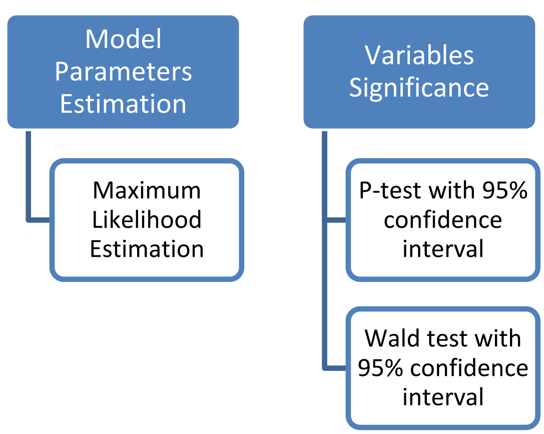


Figure 5-1 Model Parameters

Table 5-2 Surrounding Conditions - Level 1 Parameters

Surrounding Conditions	Coefficient (β)	Std. Error	Wald	P-Value	Expected Value
Intercept	-2.099	0.269	61.039	0.000	--
Pipe age	-0.017	0.004	122.554	0.000	1.109
Diameter	-0.011	0.004	9.354	0.002	0.989
[Material=C.I.]	-0.584	0.569	1.054	0.305	0.558
[Material=CONCRETE]	-0.529	0.192	7.561	0.006	0.589
[Material=D.I.]	-0.750	0.252	8.818	0.003	0.472
[Material=H.D.P.E.]	-0.175	0.514	0.116	0.733	0.839
[Material=P.V.C.]	-0.645	0.228	8.021	0.005	0.525
[Material=STEEL]	-0.252	0.000	--	--	0.777
[Material=V.C.] (Ref.)	0	--	--	--	--

Table 5-3 Surrounding Conditions - Level 2 Parameters

Surrounding Conditions	Coefficient (β)	Std. Error	Wald	P-Value	Expected Value
Intercept	-0.329	0.289	1.291	0.000	--
Pipe age	-0.004	0.004	114.87	0.023	1.088
Diameter	0.001	0.004	0.023	0.880	1.001
[Material=C.I.]	-0.531	0.675	0.619	0.432	0.588
[Material=CONCRETE]	-0.157	0.216	0.525	0.469	0.855
[Material=D.I.]	-0.744	0.301	6.085	0.014	0.475
[Material=H.D.P.E.]	0.093	0.578	0.026	0.872	1.097
[Material=P.V.C.]	-0.225	0.256	0.769	0.381	0.799
[Material=STEEL]	19.829	8111.952	0.000	0.998	409062270.35
[Material=V.C.] (Ref.)	0	--	--	--	--

Table 5-4 Surrounding Conditions - Level 3 Parameters

Surrounding Conditions	Coefficient (β)	Std. Error	Wald	P-Value	Expected Value
Intercept	-1.932	0.272	50.570	0.000	--
Pipe age	-0.012	0.003	211.304	0.001	1.136
Diameter	-0.016	0.004	16.193	0.000	0.985
[Material=C.I.]	-0.216	0.539	0.161	0.689	0.806
[Material=CONCRETE]	-0.420	0.194	4.690	0.030	0.657
[Material=D.I.]	-0.580	0.254	5.219	0.022	0.560
[Material=H.D.P.E.]	-0.265	0.535	0.244	0.621	0.768
[Material=P.V.C.]	-0.402	0.229	3.093	0.079	0.669
[Material=STEEL]	0.016	0.000	--	--	1.016
[Material=V.C.] (Ref.)	0	--	--	--	--

As we can see from the previous results, the significance of variables is different at each level, as in Table 5-5.

Table 5-5 Significance of Variables

Parameter Estimation Level	Significant Variables
Level 1	• Pipe Age • Diameter • Material=CONCRETE • Material=D.I. • Material=P.V.C.
Level 2	• Pipe Age • Material=D.I.
Level 3	• Pipe Age • Diameter • Material=CONCRETE • Material=D.I.

5.2.3 Model Significance

A log-likelihood test was performed to determine the model significance, as in Table 5-6; the significance is less than the cut-off value (0.05).

Table 5-6 Multinomial Logistic Regression Significance

Model	-2 Log-likelihood	Chi-Square	Degree of Freedom	Significance
Null	49,591.67	--	--	--
Full	30,293.735	17,491.2	65	0.000

Multinomial logistic regression for the data set resulted in three equations for pipe surrounding conditions rating 1, 2, and 3.

Multinomial logistic regression equations start with parameters estimation. Below equations were developed based on the coefficient of variables (β):

$$f_1(x) = \ln\left(\frac{P\ (C=1)}{P\ (C=4)}\right)$$

$= -\ 2.099 - 0.017 \times \text{Age} - 0.011 \times \text{Diameter}$

$-\ 0.175 \times D_{\text{Material=HDPE}} - 0.584 \times D_{\text{Material=CI}} - 0.75 \times D_{\text{Material=DI}}$

$+\ 0.645 \times D_{\text{Material=PVC}} - 0.252 \times D_{\text{Material=Steel}}$

$-\ 0.529 \times D_{\text{Material=Concrete}}$ Eq. 5.2

$$f_2(x) = \ln\left(\frac{P\ (C=2)}{P\ (C=4)}\right)$$

$= -\ 0.329 - 0.004 \times \text{Age} + 0.001 \times \text{Diameter}$

$+\ 0.093 \times D_{\text{Material=HDPE}} - 0.531 \times D_{\text{Material=CI}} - 0.744 \times D_{\text{Material=DI}}$

$+\ 0.225 \times D_{\text{Material=PVC}} + 19.829 \times D_{\text{Material=Steel}}$

$-\ 0.157 \times D_{\text{Material=Concrete}}$ Eq. 5.3

$$f_3(x) = \ln\left(\frac{P\ (C=3)}{P\ (C=4)}\right)$$

$= -\ 1.932 - 0.012 \times \text{Age} - 0.016 \times \text{Diameter}$

$-\ 0.265 \times D_{\text{Material=HDPE}} - 0.216 \times D_{\text{Material=CI}} - 0.580 \times D_{\text{Material=DI}}$

$+\ 0.402 \times D_{\text{Material=PVC}} + 0.016 \times D_{\text{Material=Steel}}$

$-\ 0.42 \times D_{\text{Material=Concrete}}$ Eq. 5.4

The probabilities of pipe surrounding conditions rates will be estimated using the below equations:

$$P\ (C=1) = \frac{e^{f_1(x)}}{1 + e^{f_1(x)} + e^{f_2(x)} + e^{f_3(x)}}$$ Eq. 5.5

$$P\ (C = 2)\ =\ \frac{e^{f_2(x)}}{1 +\ e^{f_1(x)}\ + e^{f_2(x)} + e^{f_3(x)}} \qquad \text{Eq. 5.6}$$

$$P\ (C = 3)\ =\ \frac{e^{f_3(x)}}{1 +\ e^{f_1(x)}\ + e^{f_2(x)} + e^{f_3(x)}} \qquad \text{Eq. 5.7}$$

$$P\ (C = 4)\ =\ \frac{e^{f_3(x)}}{1 +\ e^{f_1(x)}\ + e^{f_2(x)} + e^{f_3(x)}} \qquad \text{Eq. 5.8}$$

Table 5-7 shows the result of testing the model, and Figure 5-2 illustrates the correct overall percentages, using 20% of the sample via previous equations.

Table 5-7 Multinomial Classification

Observed	Predicted			
	1	2	3	4
1	100	9	15	13
2	12	59	21	6
3	89	25	43	7
4	97	49	41	52

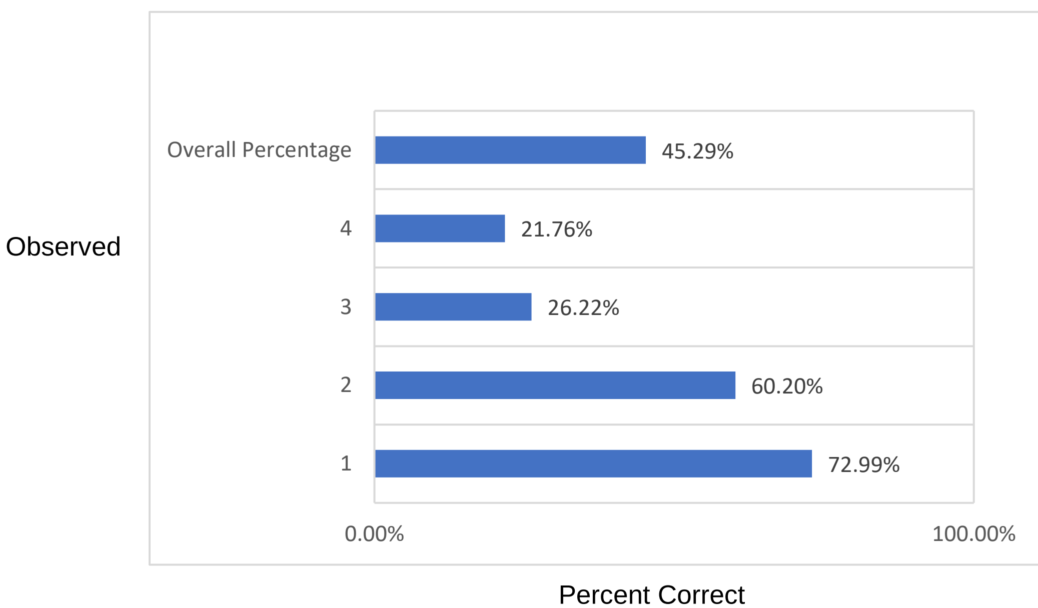


Figure 5-2 MLR Overall Correct Percentages

Based on Table 5-7 and Figure 5-2 above, the overall correct predicted percent for multinomial logistic regression was 45.29%, which is relatively insignificant. Moreover, the correct predicted percent for surrounding condition ratings 4 and 3 are 21.76% and 26.22%, respectively.

Considering that the overall correct predicted percent is relatively low, this dissertation will consider binary logistic regression as another model.

5.3 Binary Logistic Regression

5.3.1 Model Clarification

Binary logistic regression is a statistical modeling method to connect dichotomous or binary variables that take only two variables with predicator variables that may be categorical and numerical values (Laakso, et al. 2018).

The dependent variable was the soil elevation difference between 2010 and 2015, as a binary code; decrease 0 and increase 1.

The binary logistic regression model based on the dependent variable with two values is formulated as follows in equation 5.9:

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

$$\begin{aligned} &\alpha + \beta_1 \times \mathrm{Age} + \beta_2 \times \mathrm{Diameter} + \beta_3 \times \mathrm{D}_{\mathrm{Material=HDPE}} \\ &+ \beta_4 \times \mathrm{D}_{\mathrm{Material=CI}} + \beta_5 \times \mathrm{D}_{\mathrm{Material=DI}} + \beta_6 \times \mathrm{D}_{\mathrm{Material=PVC}} \\ &+ \beta_7 \times \mathrm{D}_{\mathrm{Material=Steel}} + \beta_8 \times \mathrm{D}_{\mathrm{Material=VC}} \\ &+ \beta_9 \times \mathrm{D}_{\mathrm{Material=Concrete}} + \beta_{10} \times \mathrm{D}_{\mathrm{Far/Near\ Bodies\ of\ Water}} \end{aligned} \qquad \text{Eq. 5.9}$$

Where:

α: intercept

$\beta_1, \beta_2, \ldots, \beta_{10}$: regression coefficients

D: dummy variable used to assign values to categorical independent variables

The development of binary logistic regression was based on 80% of the data using the SPSS Statistics software. The model variables can be found in Table 5-8.

Table 5-8 Binary Logistic Regression Model Variables

Dependent Variable	Independent Variables
The soil elevation difference between the years 2010 and 2015. It is coded as 0 if the elevation difference is negative and 1 if the elevation difference is positive.	Pipe age (Continuous quantitative)
	Pipe diameter (Discrete quantitative)
	Pipe material (Nominal categorical) • PVC (Polyvinyl chloride) • VC (Vitrified clay) • Concrete • DI (Ductile Iron) • HDPE (High-density polyethylene) • CI (Cast Iron)
	Pipe location is coded as 0 for pipes far from bodies of water and 1 for the near pipes.

5.3.2 Parameters Estimation

Two statistical tests were used to identify the significance of the variables with a confidence interval of 95%, the Wald test and the P-test.

The method used for logistic regression will determine how the regression model will be constructed. Forward and backward are the available methods. The backward method removes explanatory variables from the full model. Meanwhile, the forward method adds explanatory variables to the basic model.

The backward statistical method was used to develop the binary logistic model. It starts with all the variables mentioned above, then variables with the least effect (highest P-value) will be removed from the model as in Table 5-9.

Table 5-9 Parameters Estimation for Binary Logistic Regression Model

Variables	Coefficient (β)	Standard Error	Wald	P-value	Expected Value
Intercept	-5.105	0.755	33.84	0.000	--
Pipe age	0.004	0.003	2280.199	0.013	1.004
Diameter	-0.003	0.003	63.682	0.024	0.997
Material = VC (Ref.)	0	--	--	--	--
Material (1) = Concrete	0.257	0.454	0.319	0.057	1.292
Material (2) = HDPE	0.150	0.140	1.145	0.028	1.161
Material (3) = CI	0.385	0.189	4.130	0.004	1.469
Material (4) = DI	-0.120	0.356	0.114	0.074	0.887
Material (5) = PVC	0.301	0.166	3.313	0.007	1.352
Material (6) = Steel	-20.982	28420.531	0.000	0.999	0.000
Far 0 / Near 1	0.090	0.083	1.181	0.028	1.094

Based on Table 5-9 results, the significant variables were the input for the backward stepwise. Table 5-10 presents the backward stepwise results.

Table 5-10 Parameters Estimation for Binary Logistic Regression Model

(Backward Stepwise)

Variables	Coefficient (β)	Standard Error	Wald	P-value
Intercept	-6.73	0.098	1978.843	0.000
Pipe age	0.004	0.003	2280.199	0.013
Diameter	-0.001	0.003	63.682	0.024
Material (1) = Concrete	0.227	0.254	0.319	0.047
Material (2) = HDPE	0.120	0.140	1.145	0.028
Material (3) = CI	0.185	0.109	4.130	0.004
Material (4) = DI	-0.120	0.116	0.114	0.054
Material (5) = PVC	0.101	0.136	3.313	0.007
Far 0 / Near 1	0.090	0.053	1.181	0.028

5.3.3 Model Significance

Table 5-11 Significance for Binary Logistic Regression Model

Model	-2 Log-likelihood	Chi-Square	Degree of Freedom	Significance
Null	29,923.7	--	--	--
Full	11,653	9,362.0	13	0.006

Table 5-11 presents the significance of the model based on the log-likelihood test. The significance is 0.006, which is less than 0.05.

5.4 Chapter Summary

This chapter presented the model development process for the wastewater systems. Multinomial logistic regression and binary logistic regression were discussed in brief steps for developing the models.

Chapter 6 Results, Classification, and Discussion

Planning and managing wastewater interceptors are an exceptional task to provide municipalities and governments with the needed resources and get the highest benefit-cost ratio for the plans. Therefore, this chapter considers the validation for this dissertation model.

6.1 Binary Logistic Regression

After developing the logistic regression model, Eq. 6.1 provides the prediction elevation difference over the years 2010-2015.

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

$$-6.73 + 0.004 \times \text{Age} - 0.001 \times \text{Diameter} + 0.12 \times \text{D}_{\text{Material=HDPE}} + 0.185 \times \text{D}_{\text{Material=CI}} - 0.12 \times \text{D}_{\text{Material=DI}} + 0.101 \times \text{D}_{\text{Material=PVC}} + 0.227 \times \text{D}_{\text{Material=Concrete}} + 0.09 \times \text{D}_{\text{Far/Near Bodies of Water}} \quad \text{Eq. 6.1}$$

Where:

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

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

6.1.1 Classification

The next step is to go under the validation phase by identifying the predicted results. The remaining 20% of the data will be used for validation. Table 6-1 presents the classification table for the binary logistic regression.

Table 6-1 Classification for Binary Logistic Regression Model

Observed	Predicted	
	0	1
0	267	52
1	68	251

Based on Table 6-1, the percent of correct predictions is illustrated in Figure 6-1.

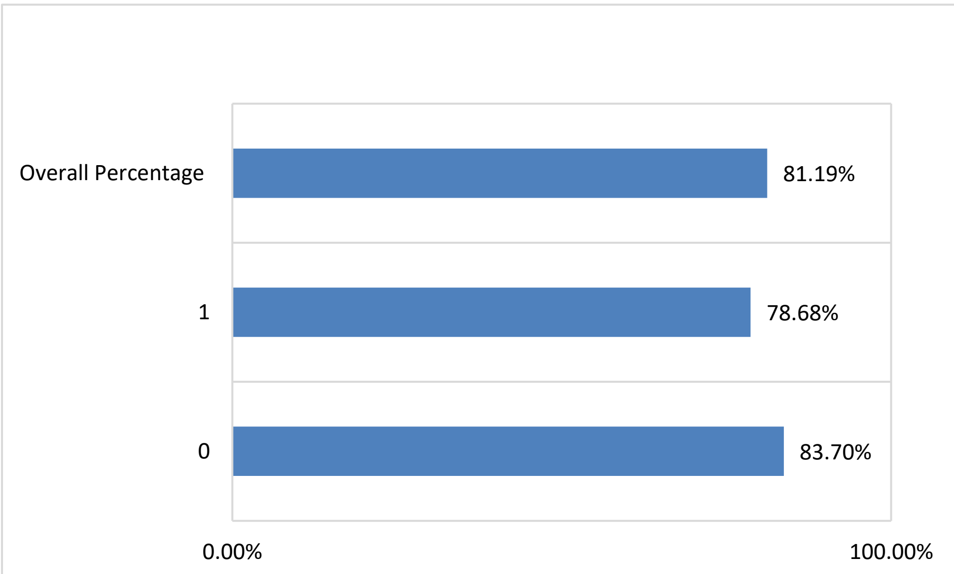


Figure 6-1 Binary Logistic Overall Correct Percentages

6.1.2 True vs. False and Positive vs. Negative

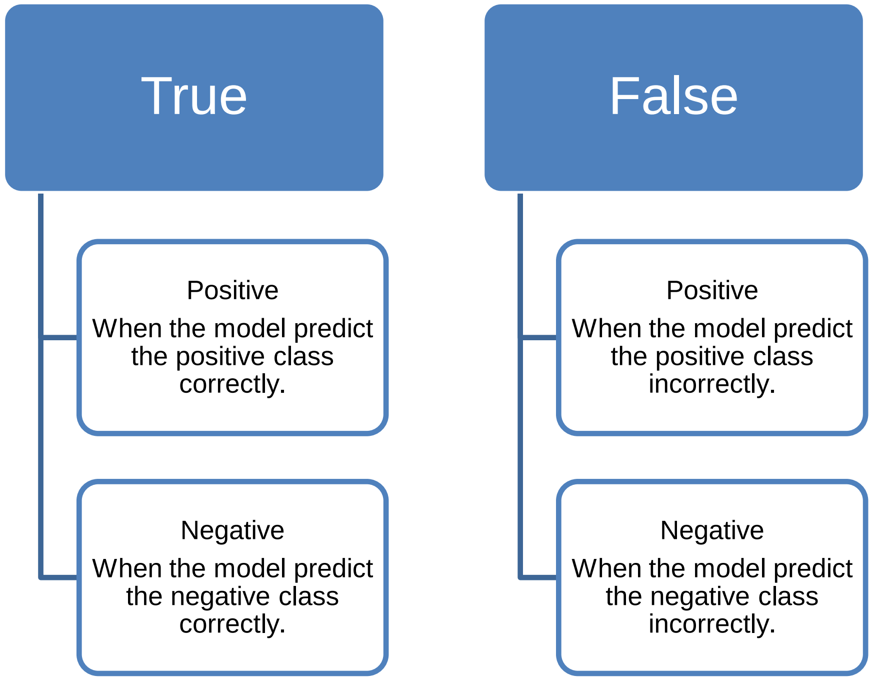


Figure 6-2 True vs. False and Positive vs. Negative

Table 6-2 summarizes the binary logistic regression model using the confusion matrix that shows the four expected outcomes, and we will evaluate our model classification based on these four outcomes.

Table 6-2 Confusion Matrix for Binary Logistic Regression Model

True Positive (TP): • Reality: Elevation Increased. • Model Prediction: Elevation Increased. • Outcome: Correct Prediction.	False Positive (FP): • Reality: Elevation Decreased. • Model Prediction: Elevation Increased. • Outcome: Wrong Prediction.
False Negative (FN): • Reality: Elevation Increased. • Model Prediction: Elevation Decreased • Outcome: Wrong Prediction.	True Negative (TN): • Reality: Elevation Decreased. • Model Prediction: Elevation Decreased. • Outcome: Correct Prediction.

6.1.3 Accuracy

One parameter for evaluating classification models is accuracy; it represents the percentage of correct predictions made by our model. The following equation Eq. 6.4 is the general formula of accuracy:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \qquad \text{Eq. 6.4}$$

Accuracy can also be assessed in terms of positives and negatives in our model, as shown below:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad \text{Eq. 6.5}$$

Where:

TP = True Positives.

TN = True Negatives.

FP = False Positives.

FN = False Negatives.

Based on the above formulas, the degree of accuracy for the binary logistic regression model is 0.812, or 81.2%.

Accuracy alone does not convey the whole evaluating classification to understand our model's performance better. The following section will discuss the ROC curve (receiver operating characteristic curve).

6.1.4 ROC (Receiver Operating Characteristic)

A receiver operating characteristic curve (ROC curve) is a graph that shows how well a classification model performs across all categorization levels. Two parameters are plotted on this curve:

True Positive Rate (**TPR**) is a synonym for sensitivity.

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \qquad \text{Eq. 6.6}$$

False Positive Rate (**FPR**).

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}} \qquad \text{Eq. 6.7}$$

TPR vs. FPR is plotted on a ROC curve at various categorization levels. As the classification threshold is lowered, more objects are classified as positive, increasing both False Positives and True Positives. A logistic regression model can be analysed multiple times with different classification criteria to compute the ROC curve points, but AUC is the fastest sorting-based method. The AUC stands for “Area under the ROC Curve,” which refers to the complete two-dimensional area beneath the entire ROC curve from (0,0) to (1,1).

6.1.5 Sensitivity and Specificity

Other alternatives to check the model performance are sensitivity and specificity. Sensitivity represents how effectively the classifier predicts positive samples, whereas specificity expresses how well classifiers detect negative samples.

Sensitivity is a synonym for True Positive Rate (TPR).

$$\mathrm{Sensitvity} = \mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad \text{Eq. 6.8}$$

Specificity.

$$Specificity = \frac{\mathrm{TN}}{\mathrm{FP+TN}} \qquad \text{Eq. 6.9}$$

$$FPR \ = 1 - Specificity = 1 - \frac{\mathrm{TN}}{\mathrm{FP+TN}} = \frac{\mathrm{FP}}{\mathrm{FP+TN}} \qquad \text{Eq. 6.10}$$

Based on the above formulas, the sensitivity and specificity for our model are as follows:

$$\mathrm{Sensitvity} = \mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP + FN}} = \frac{251}{251 + 68} = 78.68\%$$

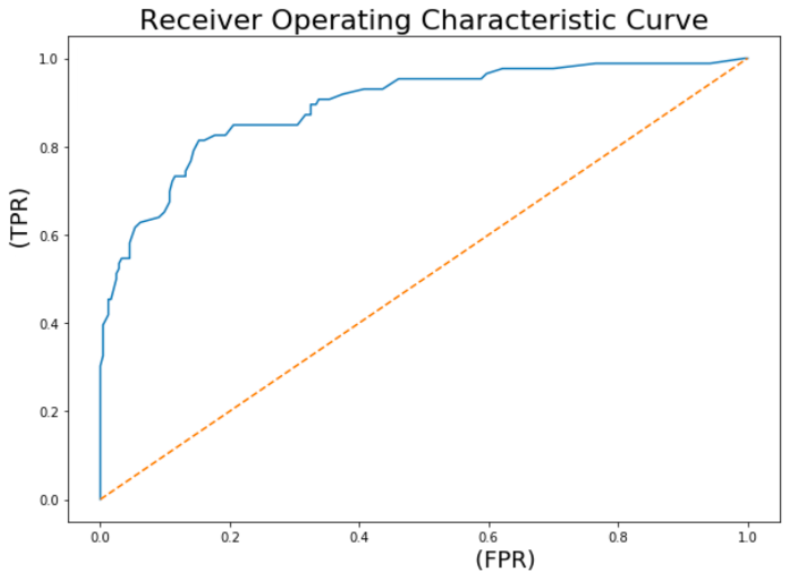
$$Specificity = \frac{\mathrm{TN}}{\mathrm{FP + TN}} = \frac{267}{52 + 267} = 83.70\%$$

Table 6-3 Binary Logistic Regression Model Performance

True Positive (TP): 251	False Positive (FP): 52
False Negative (FN): 68	True Negative (TN): 267

$$\mathrm{TPR} = \mathrm{Sensitvity} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = \frac{251}{251+68} = 78.68\%$$

$$\mathrm{FPR} = 1 - Specificity = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = \frac{52}{52+267} = 16.30\%$$



$$\mathrm{AUC} = 0.879$$

The area under the ROC curve for the binary logistic regression model is 0.879, indicating acceptable results. As a result, the logistic regression equation can forecast the surrounding conditions of wastewater interceptors adjacent to bodies of water.

6.2 Significant Variables

The significant variables for the final model were eight only, as in Figure 6-3

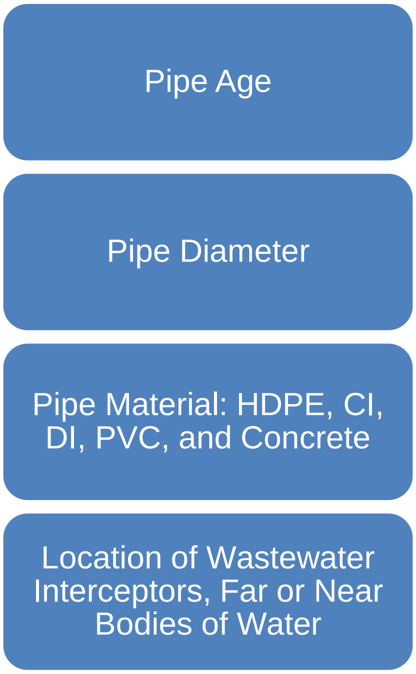


Figure 6-3 Significant Variables

6.3 Insignificant Variables

According to the backward stepwise analysis, the insignificant variables for the binary logistic regression model were Material (VC or Steel). Several aspects must be addressed to assess the influence of pipe lengths, including soil type, water table, pipe material, and pipe diameter.

Different materials used in wastewater interceptors react differently to the environment. For example, abrasion resistance is vital in concrete pipes, and acid resistance is high in clay pipes. Other pipes have superior resistance to acidic and alkaline wastes, but they can distort excessively under strain. Corrosion of steel pipes affects pipe strength, resulting in leaks, breaks, low water pressure, blockages, and other issues.

6.4 Soil Erosion

Annual global soil erosion is substantially higher than annual soil replenishment (Favis-Mortlock, 2008). Soil erosion is the loss of the top layer of soil, which can be caused by various factors, including wind and water.

Streams and rivers are avenues for soil transportation. Watersheds will become prone to floods when vast volumes of soil deposits accumulate in local lakes and reservoirs. This erosion causes valuable agriculture and infrastructures to be destroyed.

Below are some common strategies for effective erosion control:

- Plant Vegetation: Wind can be blocked by trees, bushes, hedgerows, and ground plants. Maintaining continuous ground cover, such as planting cover crops, also aids in binding soil to roots.
- Matting: This ground covering, also known as an erosion control blanket, comprises open-weave, biodegradable materials that insulate the soil while also supporting growing vegetation on bare ground. This erosion control method is generally effective for solar farms and building sites where vast regions are left barren and subject to wind and water erosion.
- Grazing: Rotational grazing involves moving cattle from one pasture plot to the next. Each paddock is given a break and allowed to recover naturally, reducing soil compaction and erosion. Installing fencing and stream crossings to protect pastures from degradation is also practical.

To summarize the recommendations for the City of Fort Worth:

- Future asset management plans must include bodies of water and erosion control methods as essential and influence variables of the plan.
- ArcGIS could be the leading platform for the plan since it can help prioritize time and cost savings.

6.5 Discussions

Wastewater systems collect sewage from different types of users. Generally, infrastructures are designed and constructed to serve for many years. Over its life, the system deteriorates, and the pipe failure likelihood and consequences increase significantly. The asset management plan is the central concept in performing the systems' repair and rehabilitation decisions since inspection and monitoring are time- and budget-consuming tasks. This brings the need for asset management plans developed with statistical tools such as SPSS statistics software based on historical data.

Variables that influence the surrounding conditions for wastewater interceptors were pipe age, pipe diameter, pipe material (HDPE, CI, DI, PVC, and Concrete), and pipes' location with reference bodies of water. Future asset management plans must include these influence variables as an essential and practical part of the plan.

Consequently, surrounding soil elevation for pipelines could be a valuable simple metric compared with a holistic view across the entire wastewater system. A benchmarking approach, every 5-years, could be used to predict the future condition of pipelines based on the condition of similar but older pipelines.

There is no standard approach for evaluating the structural integrity of wastewater pipelines in sewer system asset management (Loganathan 2019). Different researchers have considered the deterioration of the wastewater pipelines as their model. However, the variables used to develop the models were different. Table 6-4 presents a comparison of variables among recent models.

Table 6-4 Comparison of Variables for Recent Models

This Dissertation	Atambo (2021)	Malek Mohammadi (2019)	Loganathan (2021)
Age	Age	Age	Age
Diameter	Diameter	Diameter	Diameter
Material	Depth	Depth	Slope
Surrounding Soil Elevation	Slope	Slope	Length
Location (with reference to bodies of water)	Length	Length	MAPSCOGRID
—	Soil pH	Soil Sulfate	SUBAREA
	Material	Soil pH	PACP
	Soil Type	Watertable	—
	—	Pipe Flow	
		Material	
		Soil Type	
		Soil Hydraulic Group	
		Soil Corrosivity	

This dissertation developed multinomial logistic regression and binary logistic regression, the accuracy for the models was 45.29% and 81.20%, respectively. However, it was compared with different models for different authors as shown in Table 6-5.

Table 6-5 Models' Accuracy Comparison

Model	Model Accuracy	Author
Multinomial Logistic Regression	65.8%	Malek (2019)
	75%	Atambo (2021)
	45.29%	**This Dissertation**
Binary Logistic Regression	84.6%	Malek (2019)
	81.2%	**This Dissertation**
KNN	83.4%	Malek (2019)
Neural Networks	85%	Atambo (2021)

Based on the binary logistic regression model, the influence variables for the wastewater pipelines' surrounding soil were as follows:

- Pipe Age. The coefficient of pipe age is positive in the binary logistic regression equation. With Wald = 2280.199 and P-value = 0.013, the binary logistic regression findings revealed that pipe age has a significant impact on the condition of the surrounding soil condition for wastewater interceptors as it has a positive coefficient, which indicates that an increase in age will probability result in the surrounding condition to be in a risk condition.

- Pipe Diameter. The coefficient of pipe diameter is negative in the binary logistic regression equation. With Wald = 63.682 and P-value = 0.024, pipe diameter was also found to significantly impact soil difference elevation over the years for wastewater interceptors near bodies of water. It has a negative coefficient which means that an increase in the pipe diameter will probably reduce the risk of pipe surrounding conditions change.
- Pipe Material. The Wald and P-values for the significant pipe materials were different. The results of binary logistic regression revealed a moderate significance in High-Density Polyethylene (HDPE), Cast iron (CI), Ductile Iron (DI), Polyvinyl Chloride (PVC), and Concrete materials, as shown in Table 6-6.

Table 6-6 Binary Logistic Regression Variables' Performance

Variables	Coefficient (β)	Wald	P-value	Remarks
Material (1) = Concrete	0.227	0.319	0.047	Positive Coefficient
Material (2) = HDPE	0.120	1.145	0.028	Positive Coefficient
Material (3) = CI	0.185	4.130	0.004	Positive Coefficient
Material (4) = DI	-0.120	0.114	0.054	Negative Coefficient
Material (5) = PVC	0.101	3.313	0.007	Positive Coefficient

- Location of Wastewater Interceptors. In the binary logistic regression model, wastewater interceptors' location with reference to bodies of water as far or near was also determined to be a significant variable with Wald

= 1.181 and P-value = 0.028. The coefficient is positive which indicates that as the pipe is nearest to the bodies of water, the risk of the surrounding pipe soil to be in poor condition and indeed the pipe failure likelihood will increase.

6.6 Practical Applications

The results of this dissertation can help municipalities in managing wastewater interceptors. The model developed in this dissertation may be used to create a wastewater interceptors inspection schedule. A cost-benefit analysis may be conducted to evaluate cost savings the model could save if used in place of yearly inspection programs. The developed model has a degree of accuracy of 81.2%.

Moreover, the significant variables of the model could be an essential input for developing long-term asset management plans. On the other hand, ArcGIS could be the leading platform for the asset management plan since it can help prioritize time and cost savings.

This dissertation was based on a 5-year span data. It is recommended to monitor the wastewater interceptors adjacent to bodies of water in short intervals. Frequent inspections (every 5 years or less) are needed for wastewater pipelines when their locations are less than 10 ft away from the bodies of water. Distance from body of water provides a significant variable in the useful life of the pipeline irespctive of pipe material, such as, concrete, HDPE, CI, or PVC.Moreover, as pipeline age increases, the effect on the surrounding soil elevation also increases.

However, the soil surroundings for DI pipelines were found to be more stable since 69% of these pipelines are installed more than 10 ft away from body of water. This research also showed that pipe diameter variable has a negative coefficient, which means that an increase in the pipe diameter will probably reduce the risk of pipe surrounding conditions change. After inspection and analysis of, wastewater interceptors could be labeled and scored. The high-risk scored interceptors will have priority in the replacement and rehabilitation plan. Indeed, this will limit the cost and time consumed in inspections or in case of unpredicted pipe failure. The model could be used for different data years, which can help in defining the areas where the inspection will take place to enhance asset management planning for municipalities.

6.7 Chapter Summary

This chapter presented the details of the developed model, 80% of the data were used to develop the model. Meanwhile, 20% of the data were used as a sample to validate the models. This dissertation developed multinomial logistic regression and binary logistic regression, the accuracy for the models was 45.29% and 81.20%, respectively. Practical applications for the model were discussed in the chapter.

Chapter 7 Conclusions and Recommendations for Future Research

7.1 Conclusions

Municipalities would benefit from knowing and predicting how the asset management for wastewater interceptors is different with reference to the location of bodies of water. Two logistic regression models were used to predict how bodies of water can affect the soil surrounding wastewater interceptors. The models were created, verified, and tested. Both models were created using 80% of the dataset chosen at random. In the validation of the model, the remaining 20% of the data was used at random.

According to the model's findings, pipe diameter, age, pipe material, and location with reference to bodies of water were the most important parameters. The multinomial and binary logistic regression performances were 45.29% and 81.20%, respectively.

The binary logistic regression results revealed that the surrounding soil elevation difference over the years 2010 to 2015 near water bodies has decreased compared to the interceptors far away from bodies of water. Therefore, the interceptors are at a higher risk of failure. As a result, the binary logistic regression equation is significant, indicating that the area under the ROC curve was 0.879, indicating that the model is reliable.

7.2 Dissertation Limitations

The study's base year data were obtained for 2010 and 2015. It is natural to doubt that the relevance of surrounding conditions changes will be detectable enough. The average soil erosion is within the range of 5.6 to 7.7 tons per acre per year (Denton, 2000). The loss of every 5 tons per acre represents 1/32-in. of topsoil (USDA NRCS, 1996).

This dissertation did not consider other influence factors, such as depth and slope of the wastewater interceptors and soil type. As a result, the lack of information on these factors is the main limitation of this dissertation.

7.3 Recommendations for Future Research

Some of the significant prospective future development areas raised in this dissertation that should be addressed:

- Other independent factors, such as soil type, pipe installation method, and failure history, can improve the model represented in this research.
- Further exploration of deep learning algorithms to develop a model will be an important critical component of future efforts.
- This dissertation is based on data from the City of Fort Worth. To improve the accuracy of models, more inspection data is needed to compare the results of models developed for other cities could be an essential part of future work.
- Future studies should include more data for more years to distribute the findings over more than one five-year span, and the results should be compared to the findings of this study.
- The model developed in this dissertation can be utilized to create a wastewater interceptors inspection schedule. A cost-benefit analysis can be used to determine the cost savings the model could save if used in place of yearly inspection programs.

References

Agresti, A. (2007). An introduction to categorical data analysis. 3rd Ed., John Wiley & Sons, Hoboken, NJ.

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, (2012)

ASCE (2021). Report Card for America's Infrastructure. New York: American Society of Civil Engineers (ASCE). Available at: https://infrastructurereportcard.org/ . (Accessed Feb 21, 2022).

Atambo, D. O. (2021). "Development and Comparison of Prediction Models for Sanitary Sewer Pipes Condition Assessment Using Multinomial Logistic Regression and Artificial Neural Network." Doctoral Dissertation. The University of Texas at Arlington TX, USA.

Calvetti, F., and Di Prisco, C. (2004). '' Experimental and Numerical Analysis of Soil–Pipe Interaction.'' Journal of Geotechnical and Geoenvironmental Engineering December 2004 Volume 130, Issue 12 (1292 - 1299).

Denton, P. (2000). ''Tennessee soil erosion picture fuzzy.'' Southeast Farm Press. Available at: http://southeastfarmpress.com. (Accessed Nov 12, 2021).

EPA (2002). Fact Sheet: Asset Management for Sewer Collection Systems. Available at: http:// www.Ep.gov/nrmrl. (Accessed Oct 16, 2021).

EPA (2002). The Clean Water and Drinking Water Infrastructure Gap Analysis. Available at: http:// www.Ep.gov/nrmrl. (Accessed Oct 5, 2021).

EPA, (2015), Condition Assessment of Underground Pipes. Available at: http:// www.Ep.gov/nrmrl. (Accessed Oct 24, 2021).

Favis-Mortlock, D. (2008). Soil erosion site. Available at: http://soilerosion.net. (Accessed Dec 9, 2021).

Flynn, R., Brando, F., Pinto, M., and Malsch, E. (2018). ‘’Challenges in Non-Destructive Evaluation of Subjects That Are Buried and Underwater.’’ Eighth Congress on Forensic Engineering.

Gall, M.D., Gall, J. P., & Borg, W. R. (2007). Educational research: An introduction (8th ed.). Boston, MA: Pearson Education, Inc.

GAO (2004), ‘’Performance and Accountability Report Fiscal Year 2004’’, GAO-05-62SP Published: Nov 15, 2004. Publicly Released: Nov 15, 2004.

Harris, J. (2017). “Increasing Sustainability in Small Utilities through GIS and Asset Management.” Pipelines Conference, (August 6–9, 2017) Phoenix, Arizona, USA.

Hawari, A., Alkadour, F., Elmasry, M., and Zayed, T. (2020). “A state-of-the-art review on condition assessment models developed for sewer pipelines.” Engineering Applications of Artificial Intelligence, 93.

Hosmer, D. W., Lemeshow S. (2000). Applied logistic regression. Wiley, Wiley Series in Probability and Statistics.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression. John Wiley & Sons. Inc., Columbus, Ohio, USA.

INGENI-UM “International Infrastructure Management Manual” Association of Local Government Engineering NZ Incorporated, 2002.

Kercho, J. and Conlon, L. (2019). “Mains, Trains, and Automobiles: Utilizing Fort Worth’s Risk Assessment Data to Drive Sanitary Sewer Rehabilitation.” Pipelines Conference, (July 21–24, 2019) Nashville, Tennessee, USA.

Khashei, M., and Bijari, M. (2010). “An artificial neural network (p, d, q) model for time series forecasting.” Expert Systems with Applications, 37(1).

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

Loganathan, K. (2021). “Development of a Model to Prioritize Inspection and Condition Assessment of Gravity Sanitary Sewer Systems.” Doctoral Dissertation. The University of Texas at Arlington TX, USA.

Mahaffey, R. (2016). "Challenges and the future of the implementation of bulk water pipeline condition assessment." Journal of IMESA, 2016.

Malek Mohammadi, M. (2019). "Development of Condition Prediction Models for sanitary Sewer Pipes." Doctoral Dissertation. The University of Texas at Arlington TX, USA.

Matthews, J., Piratla, K., and Koo, D. (2016). "Sustainability Evaluation of Pipe Asset Management Strategies."

Meyers, L., S., Gamst, G., and Guarino, A. J. (2006). Applied multivariate research: Design and interpretation, Sage Publications, Thousand Oaks, California, United States.

Najafi, M., and Gokhale, S. (2022). Trenchless technology: Pipeline and Utility Design, Construction, and Renewal (2nd Edition). McGraw-Hill, New York, NY.

Najafi, M. (2016) Pipeline Infrastructure Renewal and Asset Management (1st Edition), McGraw-Hill, New York, NY.

Nardini, P. D. and Zarghamee, M. S., Simpson Gumpertz & Heger Inc. (2013). ''Pipeline Asset Management Using a GIS Solution.'' ASCE Pipelines Conference (June 23-26, 2013), Fort Worth, Texas, USA.

Nardini, P. D., Zarghamee, M. S., Bain, L., Jalbert, C., and Remer, A. (2015). ''Incorporating GIS-Based Structural Evaluation Tools into Pipeline Asset Management.'' Pipelines Conference, (August 23–26, 2015) Baltimore, Maryland, USA.

National Association of Sewer Service Companies (NASSCO) (2015). "Pipeline Assessment Certification Program V 7.0 – Appendix D – PACP Based Risk Management''.

National Association of Sewer Service Companies (NASSCO), Pipeline Assessment Certification Manual, Dallas (2015).

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.

Forsyth, C. Hartzell, S. Chader, and T. Browne (2018). ‘’Underwater Inspection and Imaging Technologies for Pipelines.’’ Conference, (July 15–18, 2018) Toronto, Ontario, Canada.

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

Sever, F. V., Cawley, J., Neukam, Z., Schroeder, J., and Lehmann, M., (2017). ‘’Large Diameter Sewer Assessment: When to Take It to the Next Level.’’ Pipelines Conference, (August 6–9, 2017) Phoenix, Arizona, USA.

Thornhill, R., and Crumb, F. (2014). ‘’Fort Worth Interceptor Condition Assessment Leads to Improved Asset Management and Long-Term Sustainability.’’ Pipelines Conference, (August 3-6, 2014) Portland, Oregon, USA.

Ugarelli, R., Venkatesh, G., Brattebø, H., and Sægrov, S. (2008). ‘’Importance of investment decisions and rehabilitation approaches in an aging wastewater pipeline network.’’ Journal of Water Science & Technology—WST | 58.12 | 2008.

Ugarelli, R., Venkatesh, G., Brattebø, H., Federico, V. D., and Sægrov, S. (2010). ‘’Asset Management for Urban Wastewater Pipeline Networks.’’ Journal of Infrastructure Systems, ASCE, (June 2010).

USDA Natural Resources Conservation Service (NRCS) (1996). Soil Quality Resource Concerns: Soil Erosion. Available at: http://soils.usda.gov. (Accessed Feb 3. 2022).

Venigalla, M. and Baik, B. H. (2007). ‘’GIS-Based Engineering Management Service Functions: Taking GIS beyond Mapping for Municipal Governments.’’ Journal of Computing in Civil Engineering Vol. 21, Issue 5 (September 2007).

Wade, M. G. (2016). ‘’Strategic Integration of Risk-Based Asset Management to Improve Large Diameter Pipeline.’’ ASCE Pipelines Conference - Kansas City, Missouri (July 17–20, 2016).

Yoshizaki, K., and Sakanoue, T. (2004). ''Analytical Study in Soil-Pipeline Interaction Due to Large Ground Deformation.'' 13th World Conference on Earthquake Engineering, Vancouver, B.C., Canada (August 1-6, 2004).