STATISTICAL MODELS OF PIPE SERVICE LIFE TOWARDS ARTIFICIAL INTELLIGENCE BASED DECISION SUPPORT

by

VAHID EGHBALIFARKOOSH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

at
The University of Texas at Arlington
August 2020

Arlington, Texas

Copyright © by Vahid Eghbalifarkoosh 2020

All Rights Reserved

DEDICATION

Dedicated to my beloved mother, father, sister, and nephew, Arsha; helping and motivating me through every step of the way and shoulders to lean on. No words can explain my sincerest appreciation towards them. I thank you and owe you, forever.

ACKNOWLEDEMENTS

I would like to express my sincerest gratitude and appreciation to my doctoral advisor, Dr. Ali Abolmaali, for all of his time, kind guidance, tireless and endless support. This study would not have been successfully completed without his supervision. It has been an honor to have the opportunity to work with him.

Furthermore, I would also like to extend my warmest gratitude and appreciation to my committee members, Dr. Chen, Dr. Chao, and Dr. Zhang for all their time and valuable comments and suggestions.

I also would like to thank my mother, father, sister, and nephew Arsha, whom I love and owe forever, for all their help and support throughout every step of the way in my journey. I am so grateful and fortunate for being blessed by their unconditional love and support.

August 2020

ABSTRACT

STATISTICAL MODELS OF PIPE SERVICE LIFE
TOWARDS ARTIFICIAL INTELLIGENCE BASED
DECISION SUPPORT

Vahid Eghbalifarkoosh, Ph.D.

The University of Texas at Arlington, 2020

Supervising Professor: Dr. Ali Abolmaali

In this dissertation, through utilizing various artificial intelligence-based as well as statistical models, an effort has been made to investigate the deterioration of sewer pipes. Once the deterioration rates of sewer pipes are estimated, by assuming failure criteria, as specified in the dissertation, the associated service lives for the sewer pipes can therefore be estimated. However, it should be noted that for different sewer pipes and based upon the availability of suitable data, and due to various failure modes that could transpire in various sewer pipes, the results will thus be subjected to uncertainties and variations. In other words, depending on different sewer pipes, the adequacy and the availability of suitable data, the decision-makers' priorities and failure criteria, the estimated service lives as well as the associated deterioration curves could be subjected to variation.

Selecting a suitable model plays an important role in reducing the amount of uncertainty associated with estimation of service life of sewer pipes. In order to estimate the service lives of sewer pipes, the first step is to estimate the rate of deterioration which affects the condition rating of sewer pipes. Next, by designating a certain threshold or cut-off value, the service life of sewer pipes could thus be estimated as well. Therefore, depending on the type of selected deterioration modeling, the assignment of threshold (cut-off value) needs to be conducted with adequate engineering judgement. Additionally, based upon the

criteria of decision-makers, the certain threshold for service life may be subject to further change and improvement. Hence, the suitable modeling approach may differ for various projects and sewer pipes; i.e. a model which yields suitable results for one project may not necessarily yield suitable and reliable results in another project. This stems from the assumptions and uncertainties associated with each of the modeling approaches.

Table of Contents

List of Figures

List of Tables

Chapter 1 : Introduction

1.1 Introduction and Project Description

This project aims at identifying the service life of sewer pipes via utilization of statistical, probabilistic, and artificial intelligence models. In order to facilitate the decision-making process with regards to allocating suitable schedules, i.e. intervals and frequency of inspections, for performing maintenance and rehabilitation for sewer pipes, the estimation of service life of sewer pipes is thus of utmost significance. Various factors could affect the deterioration rate of different sewer pipes; these factors may depend upon the environmental conditions that different sewer pipes are exposed to, or the intrinsic characteristics of the materials and processes through which the sewer pipes is constructed. Furthermore, depending upon different sewer pipes, factors such as transportation and installation processes may also contribute to the service life of these assets.

In the case of large-diameter buried gravity sewer pipes, which is the focus of this study, depending on the material from which the sewer pipe is made of, the approach in determining the deterioration rate, and thus service life, of these assets can vary from one to another. This variation stems from different contributing factors such as:

- Various corrosion mechanism in different pipe materials,
- Different. modes of failure for different pipe materials (flexible pipes vs. rigid pipes),
- Distinct change of material properties for different materials (PVC pipes vs. HDPE pipes).

Furthermore, whether the pipes are located under heavy traffic or they are rarely exposed to external loads, can also play an important role in their associated service life. In addition to the stated factors, the properties of the flow, such as flow rate, acidity of the flow, etc. can also influence the deterioration rates of these sewer pipes. With regards to installation process, depending on the diameter of the pipes and the material of the pipes, the precision by which the pipe is being installed as well as the assigned slope of the sewer pipe, can be influential in determining the corresponding service lives of these sewer pipes.

In this study, the applicability of various methodologies such as artificial intelligence, probabilistic models, and different regression techniques will be investigated for large-diameter buried gravity sewer pipes. These assets will be of varying materials, diameters and other properties. Due to significant uncertainties associated with estimating service life of sewer pipes, utilization of different estimation techniques can be beneficial in designating an appropriate service life for these assets.

1.2 Project motivation

By using sewer pipe networks, various types of sewage, including commercial, industrial, and domestic, are collected and sent to treatment facilities. The majority of sewers are designed as gravity sewers. In gravity sewer pipes, the flow of sewer is become possible by using slopes in pipes. About 14,748 treatment facilities assist with treatment of sewage from 240 million users. Sewer pipes are further categorized into public and private sewer pipes; there are about 500,000 miles of private sewer pipes and 800,000 miles of public sewer pipes within the United States.

Moreover, it is projected that by year 2032, 56 million additional users will be connected to the centralized treatment facilities [ASCE, 2017]. Based on ASCE (2017), to attend to the aging problem of both the water and wastewater infrastructures, by year 2025, capital funding gap in. the amount of $150 billion will be required.

Considering the fact that most of the sewer infrastructures were constructed over 100 years ago, and taking into account the annual occurrence of 23,000 up to 75,000 sanitary sewer overflows, which are resulted from various influential parameters such as aging of the pipes, environmental factors, etc., the significance of investigation of sewer pipes with respect to these factors becomes clear [EPA, 2004]. Moreover, prediction of deterioration rate and service life of sewer pipes can also help decision makers in the process of determining the frequency and intervals of the sewer pipe inspections.

1.3 Scope of project

In the study at hand, different approaches will be used to estimate the service life of large-diameter buried sewer pipes. The diameters of sewer pipes in this project vary from 21 inches up to 66 inches. Furthermore, the pipe materials considered in implemented methodologies are as follows:

- FRP (Fiberglass Reinforced Plastic)
- VCP (Vitrified Clay Pipes)
- PVC (PolyVinyl Chloride)
- RCP (Reinforced Concrete Pipes)

Some of the features which will be used within the post-processing data are pipe slope, pipe material, average flow depth, average velocity of sewer flow, length of pipes, etc. The maximum length of individual pipe segment is 1471.2 ft and the average pipe length, considering all pipes, is equal to 357.9 ft.

1.4 Project objectives

The main objective of this project is to predict the service life of sewer pipes. Various approaches will be utilized in the prediction process for the service life of sewer pipes. These methodologies will include the following:

- Deterministic deterioration models
- Statistic deterioration models
- Artificial intelligence approach

Depending upon the abovementioned category, various methods (sub-categories) which are appropriate for obtaining the deterioration rates will be used.

Furthermore, for selected deterioration models, the influence of different independent variables on the service life of sewer pipes will be investigated as well. The need for implementation of different deterioration modeling stems from the fact that based on each individual independent variable, mainly the material of the sewer pipe, the failure mechanisms vary widely. Therefore, the need for an investigation of a method which can be applicable to a set of sewer pipe networks is of utmost significance.

Moreover, a comparison between the various methods will be beneficial in identifying the best approach for different cases.

1.5 Overview of dissertation

Chapter one presents an overview of this study; topics pertaining to literature review such as various inspection methodologies, condition grading of assets, and several deterioration methods are presented in chapter two. Chapter three contains various methodologies utilized in constructing the deterioration models of assets, and each of these models are discussed in details.

The acquisition and analysis of available data set are illustrated in chapter four. In this chapter, available independent variables as well as various condition gradings of assets are investigated. These condition gradings include O&M (operational), structural, and overall conditions of assets, and furthermore, the corresponding binary condition gradings are presented as well. Spearman rank correlations between various independent variables as well as cross table analysis are demonstrated in chapter four too. By using cross table analysis, significances of categorical variables are investigated.

In chapter five, results associated with various models are presented. In this chapter suitability of each model through various tests is investigated as well. Moreover, influence of various independent variables on probabilities of failure with respect to age of assets and therefore, the corresponding service lives are studied too. Additionally, in chapter five, the effect of population growth on probability of failure and service lives are presented as well. Finally, chapter six includes the summary of results and concluding remarks achieved through this study.

1.6 Contribution

In this dissertation, through applying two artificial intelligence-based approaches, namely LightGBM as well as CatBoost, an effort has been made to investigate the deterioration of sewer pipes. Once the rates of deterioration associated with sewer pipes are estimated, subsequently, by designating certain failure criteria, the service lives associated with the sewer pipes are estimated. Furthermore, the effect of population growth on service lives as well as the deterioration rates of sewer pipes has been investigated herein as well. The annual rate of population growth is assumed to be 0.01 and the impact of population growth is incorporated through the volume of sewer flow in sewer pipes.

The results pertaining to service life as well as probability of failure with respect to age of sewer pipes can be utilized as a means to prioritize assets for maintenance, rehabilitation or replacement. However, it should be noted that considering different sewer pipes and based on the availability of both suitable and sufficient data, and by further considering that various failure modes could transpire in different sewer pipes, the results will thus be subjected to uncertainties and variations. In other words, depending on different sewer pipes, the availability and the adequacy of suitable data, the decision-makers' priorities and failure criteria, the estimated service lives as well as the associated deterioration curves could be subjected to changes. Moreover, the aforementioned uncertainties and variations also stem from the assumptions and uncertainties associated with each of the modeling approaches as well.

Chapter 2 : Literature Review

2.1 Overview

Based on the report card published by American Society of Civil Engineers (ASCE) in 2017, considering the noticeable funding gap between what is required and what is indeed expected to be allocated for wastewater infrastructures, it is thus evident that there is a need for efficient allocation of funds for wastewater infrastructures. This implies the significance of asset management programs utilized for achieving desirable service levels through minimized spending of funds.

For instance, with regards to sewer pipelines, condition assessment and service life estimation provides criteria for prioritizing the available assets based on their current condition gradings or estimated service lives. The condition grading of sewer pipes can be attributed based on the data obtained through inspection of sewer pipes and by incorporating standard coding systems [EPA 2009].

Based on an ASCE survey it was realized that for small and large wastewater infrastructure systems, the amount of average density of sewer is equal to 23 feet per capita and 19 feet per capita, respectively. The small and large wastewater infrastructure systems are identified per below categorizations:

- Small wastewater infrastructure systems: Include less than 100,000 people
- Large wastewater infrastructure systems: Include greater than 500,000 people

Furthermore, based on the ASCE survey, the overall average density of sewer is equal to 21 feet per capita. Considering the whole population of the United States at the time of the survey, the total length of the wastewater infrastructure systems can be estimated to be near 1.2 million miles [USEPA and ASCE 1999, USEPA 2002]. Moreover, as stated earlier, sewer pipes are also categorized in either public or private sewer pipes and based on ASCE report card published in 2017, there are about 500,000 miles of private sewer pipes and 800,000 miles of public sewer pipes within the United States [ASCE, 2017].

2.2 Management of Infrastructures

With regards to the standpoint of civil engineering, infrastructures encompass a collection of physical systems / facilities utilized in order to meet necessary public service requirements. Furthermore, management of infrastructures refers to provisions taken in order to maintain the status of these infrastructures at a satisfactory level of performance [Hudson et al. 1997].

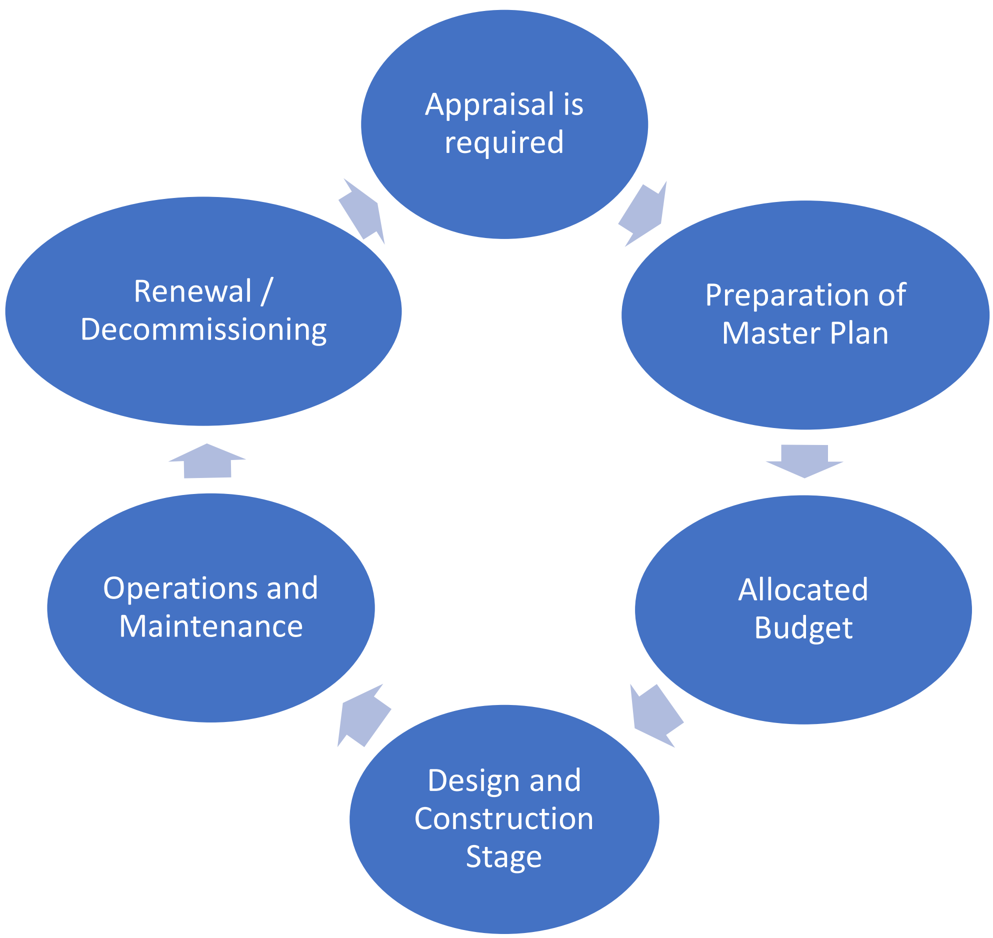

Figure 2.1: Provisions throughout the infrastructure management [adapted from Grigg 2003]

As shown in the above cycle, management of infrastructure can be considered as a cyclic process. Based on the aforementioned figure, the provision throughout the infrastructure management follow a cyclic chain; in this cycle, requirement of appraisal of the infrastructures denote the beginning of each cycle. Following this activity is the preparation of master plan. Master plan is identified based on the results of the first step of the cycle. Next, based upon the availability of allocated budgets, various options within the master plan will be investigated. The next step of the cycle is design and construction stage, in which once the projects are approved, they will be executed. After completion of construction of the project, through scheduled inspections and maintenance activities, it will be assured that the infrastructure is indeed functioning at an acceptable performance level as envisioned. In case through operations and maintenance activities the performance level of the infrastructure cannot be increased to the required level of performance, renewal or decommissioning of the infrastructure will transpire. Renewal of the infrastructure includes rehabilitation or replacement of infrastructure [Grigg 2003].

By utilizing infrastructure management systems, agencies will be able to achieve the following forefronts [Grigg 2003]:

- Through elimination of unexpected failures and regulatory charges, the costs will be substantially reduced
- The performance level of service will be subjected to improvement
- Capital improvement programs as well as operations and maintenance (O&M) will be handled efficiently
- Approval and funding regarding capital improvements will be facilitated
- Publicity as well as customer service will be enhanced.

In an ideal infrastructure management, through implementation of all the necessary provisions, the available resources will be allocated optimally in such a way that the service level of the infrastructures are maximized [Hudson et al. 1997].

Furthermore, it should be noted that the operational level of infrastructure management can be categorized in two different levels as stated below:

- Network level
- Project level

Furthermore, asset management is commonly utilized in infrastructure management. In asset management, services related to management of infrastructures are attempted to be as efficient and cost effective as possible. In the 1980s and 1990s, infrastructure systems were considered assets which have monetary values. Therefore, corporate business principles are utilized in asset management, which contain both financial and management accounting methodologies [Cowe Falls et al. 2001].

2.3 Deterioration of Sewer Pipes

Sewer pipes can be constructed of various materials. In general, the material type of sewers can be either flexible or rigid. For instance, clay pipes as well as concrete pipes (both with and without reinforcements) are considered as rigid sewers [Davies et al. 2001b]. Rigid sewers are designed so as to carry the vertical load on top of them, whereas flexible pipes carry the load based on the adjacent soil support [Abraham et al. 1998].

According to Water Environment Federation (WEF) and ASCE, the rigid pipes utilized in stormwater infrastructures may include the following materials [WEF/ASCE 1992]:

- Asbestos-cement pipes
- Cast iron pipes
- Concrete pipes

Additionally, the flexible pipes utilized in stormwater infrastructures can be of the materials stated below [WEF/ASCE 1992]:

- Corrugated aluminum pipes
- Ductile iron pipes

- Fabricated steel pipes
- Thermoplastic pipes
- Thermoset plastic pipes

Rigid have been conventionally more commonly used than flexible pipes. For instance, it is estimated that a large percentage of sewer pipes (90%) in the United Kingdom are indeed constructed from rigid pipes [Read and Vickridge 1995].

Sewerage Rehabilitation Manual was published by Water Research Center in 1983 and it was considered as the first coherent study which intended to determine the suitable process of sewer appraisal considering the significance as well as condition grading [Davies et al. 2001a]. Based on this manual, considering sewer pipes, a three-stage failure mechanism was introduced as presented below:

- First stage: Initial defect takes place
- Second stage: Due to deterioration soil support is lost
- Third stage: Sewer pipe is collapsed due to occurrence of a random event

Therefore, occurrence of a random event is considered to be the cause of collapse in sewer pipes. In other words, the exact time of failure in sewer pipes cannot be computed; however, considering the observed deteriorations within the sewer pipes, the likelihood of failure can be estimated [Abraham et al. 1998].

Furthermore, WEF/ASCE (2009) also considers the same three-stage failure mechanism as well. At the first stage of this mechanism, leaky joints, improper connections at joints, and excessive loading can be considered as some of the possible defects. Moreover, at this stage, the cracks may appear at springline, crown or invert of the sewer pipe. At the second stage, occurrences such as exfiltration, infiltration can result in the loss of soil support. At this stage, defects such as slight deformation as well as fractures may appear. At the final stage, excessive deformation of the sewer pipe can result in the collapse of the asset. Weather related occurrences as well as nearby excavations are considered as some of the potential events resulting in the collapse of the sewer pipe.

Concrete sanitary sewer pipes are especially subjected to corrosion due to presence of hydrogen sulfide from the sewer flow [WEF/ASCE 2009]. When the flow in the pipe is turbulent, or the slope of the pipes are negative or zero, the probability of occurrence of corrosion and its attributed risk will be increased. Based on the study conducted by Ablin and Kinshella in 2004, due to existence of flat slopes in the pipes as well as higher temperatures, they observed that corrosion had become even more severe [Ablin and Kinshella 2004]. The occurrence of corrosion is explained in two different stages as follows:

Stage 1: The calcium within concrete is removed due to presence of acidic environment. The present acid is caused due to certain bacteria within the sewer pipe.

Stage 2: Once the calcium is removed from the concrete, ettringite is thus created due to presence of sulfuric acid. Calcium hydroxide will then be substituted with the ettringite.

Corrosion itself can be categorized in three different classes as follows:

- Uniform corrosion
- Springline corrosion
- Crown cutting

In uniform corrosion, the portion of the pipe which is located above the sewer level will be impacted. Furthermore, the deterioration rate due to uniform corrosion is considered to be slow. However, the deterioration rate in springline corrosion is high; springline corrosion is a result of changes in the flow level within the sewer pipe. When there is more flow in the pipe, the parts of the pipe which were subject to corrosion during the lower flow levels will be washed away. On the other hand, when the pipe is fully filled with flow, crown cutting will occur. Crown cutting is observed at the joints [Salman 2010].

A thorough investigation of the influential factors regarding the structural deterioration of the rigid sewer pipes is conducted by Davies et al. (2001a). Based on the aforementioned research, various influential factors were classified in three different categories as follows [Davies et al. 2001a]:

- Construction features
- Local external factors
- Other factors (Maintenance, age, sewage properties)

Herein a brief summary of the aforementioned research is presented:

- Infiltration and exfiltration in joints, structural failures or collapses, may be due to poor standard of workmanship [Boden et al. 1975].
- When compared to pipes located in higher depths, pipes placed in shallow depths experience greater number of defects [Lester and Farrar (1979), Anderson and Cullen (1982), O'Reilly et al. (1989), Fenner and Sweeting (1999), Fenner et al. (2000)].
- Studies illustrated conflicting findings with regards to the effect of sewer size on the structural condition of sewer pipes. Findings of one study demonstrated that the smaller diameter sewer pipes were in worse condition comparted to larger diameter sewer pipes [Balmer and Meers 1982]. On the other hand, another study demonstrated that by increasing the sewer pipe diameter, the longitudinal cracks were also increased [O'Reilly et al. 1989].
- In a study conducted by Sikora in 1979, it was determined that angular bedding materials possess greater bedding factors compared to round granular bedding materials [Sikora 1979].
- If the sewer pipe is located below the ground water level, the risk of ground loss will be increased [WRC 2001].
- Greater sewage exfiltrations were observed when there are larger particle sizes within the bedding material [Rauch and Stegner 1994].
- The defect rate is observed to be increased as the sewer pipe age increases [O'Reilly et al. 1989].

- Sewer pipes located in clay soils are observed to have higher defect rates [Balmer and Meers 1981, O'Reilly et al. 1989].

In addition to structural issues, flow capacity (hydraulic condition) of the sewer pipe and various maintenance issues might occur as well [WEF/ASCE 2009]. For instance, infiltration is a major hydraulic related problem in sewer pipes. Infiltration refers to the additional flow forcing its way to the sewer pipe through various defects at joints or other locations within the sewer pipe. A similar event is occurrence of inflow. Inflow, unlike infiltration, which stems from storm occurrences, finds its way to the sewer pipe via ground connections. Both inflow and infiltration drastically impact the hydraulic condition of the sewer pipes [Abraham et al. 1998, WEF/ASCE 2009].

Occurrences of infiltration and inflow (I/I), can result in overflows and also increase the need for sewer pipes. When overflows occur in the sanitary or combined sewer pipes, raw swage will be exposed to the environment, therefore resulting in serious environmental as well as public health issues. In addition to hydraulic capacity, with regards to maintenance issues can transpire as a result of root intrusions, oil and grease, etc. within the sewer pipe [WEF/ASCE 2009].

It should be noted that the due to the fact that structural problems, hydraulic capacity, and various maintenance issues can transpire simultaneously as well, therefore, these aspects should not be considered independently [Wright and Dent 2007, WEF/ASCE 2009]. For instance, if there are structural problems, such as cracks in the pipe, I/I may occur in the sewer pipe, which results in issues in hydraulic capacity aspect. Next due to occurrence of the I/I, support from the surrounding soils may no longer exist, thereby causing a structural problem within the sewer pipe. This is just an example of how occurrence of one category of problems can result in a problem in the other category and it demonstrates that these categories cannot be viewed independently. In this case, the order of observation of problems in the sewer pipe were: structural, hydraulic, structural. Although it should be noted that any other order might also transpire in sewer pipes [Salman 2010].

Considering that in a large network of assets, some assets have not yet been inspected, by utilizing the available information from the inspected assets, and through condition prediction models as well as applying the related data, i.e. defects, etc., the behavior of all pipes can be predicted [Opila 2011].

2.3.1 Deterioration of Rigid Sewer Pipes

Based on laboratory as well as site experiments, Water Research Center based in UK [WRC 1986] investigated the deterioration of rigid pipes. The materials of these rigid pipes were vitrified clay as well as concrete. In this study, the deterioration occurring in rigid sewer pipes were classified in two different categories as follows:

- Structural deterioration
- Hydraulic deterioration

Structural deterioration of rigid sewer pipes is associated with structural defects such as occurrences of either fractures or cracks, etc.; whereas hydraulic deterioration of pipes is associated with hydraulic defects such as root intrusions in the pipes, or deposits gathered in rigid sewer pipes. The conclusions made by Water Research Center based in UK were as follows: In rigid sewer pipes, due to the fact that random occurrences of events such as storms or excavations near sewer pipes have a significant impact on the deterioration process of rigid sewer pipes, therefore, the deterioration is considered to be complicated and probabilistic. Furthermore, it was concluded that the deterioration rate of these sewer pipes is nearly impossible to be measured [Tran 2007].

2.3.2 Structural Deterioration of Rigid Sewer Pipes

Structural defects affecting the shape and the load carrying capacity of rigid sewer pipes are associated with the structural deterioration of these sewer pipes. In other words, defects impacting the structural integrity of the sewer pipes are considered to fall in this category of deterioration. The main structural defects observed in the study conducted by Water Research Center (WRC) based in UK are as follows:

➢ Fractures, cracks, holes, and deformed pipes (distorted in shape)

Furthermore, WRC utilized three phases in order to demonstrate how structural defects are developed. These three phases alongside concept of random damage occurrences are used to further describe the events resulting in the collapse of rigid sewer pipes. These phases are described below:

Phase 1: Poor handling as well as improper construction process may result in minor defects such as leaking in joint or cracks.

Phase 2: Depending on the combination of influential factors, the initial defects mentioned in phase one will be subject to extension with various rates. These influential factors are chemical corrosions, external loading (either static or dynamic loading), erosions, or ground loss. For instance, if the surrounding soil of the sewer pipe is entered in the sewer pipe (through groundwater), ground loss will take place. Therefore, the structural support of the pipe will drastically alter.

Phase 3: Probabilistic damage occurrences including excavations near the sewer pipe and excessive loading conditions could result in the collapse of the rigid sewer pipes. Hence, based on the nature of these events, determining the time at which the sewer pipe will collapse is not possible.

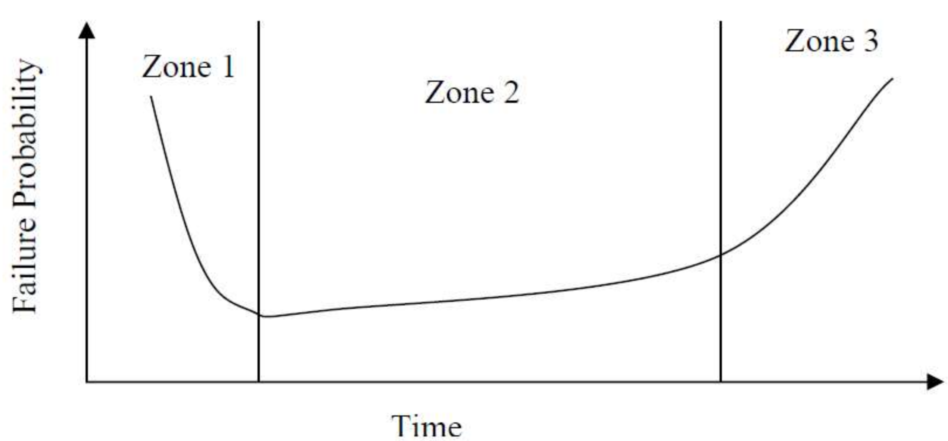

Figure 2.2: Structural deterioration of pipes demonstrated as a bath-tub curve [Tran 2007]

In asset management the aforementioned three phases can be illustrated through a bath-tub curve for sewer pipes as well as water pipes [Davis et al. 2001a, Kleiner and Rajani 2001]. The three zones are schematically demonstrated in the bath-tub curve presented herein. Each of these three zones correspond to the following:

Zone 1: When the pipe is at zone one and when the pipe is at early stages of construction, the probability of failure has a substantial value. However, by completion of construction and when the pipe is ready for use, the probability of failure reduces significantly.

Zone 2: At zone 2, when pipe is in normal operation, due to deterioration process, an increase in the deterioration will be observed in the pipes.

Zone 3: Eventually, when the failure probability is increase to a hazardous level, pipes will be at zone three, indicating that there is a need for the pipes to be either rehabilitated or replaced [Tran 2007].

Additionally, the rigid stormwater pipes can also be presumed to follow the same deterioration model [Micevski et al. 2002]. Micevski et al. (2002) observed the similar deterioration process occurring in stormwater pipes of Newcastle City located in Australia. In the aforementioned study, it was realized that a stochastic process and multi-stage transition from perfect to collapse of the pipes can be considered for structural deterioration of stormwater pipes. In addition to this realization, it was also observed that the intensity of deterioration in the future depends upon the intensity of deterioration in the present. Various studies conducted in the US also demonstrated the same stochastic feature regarding the structural deterioration of sewer pipes [Wirahadikusumah et al. 2001, Kathula 2001, Baik et al. 2006].

The deterioration mechanism in pipes resulting in the failure of pipes depends on the pipe material. On the other hand, the rate of the deterioration of pipes is dependent upon the environmental condition the pipe is exposed to as well as the operational conditions of the pipe [Makar and Kleiner 2000]. In other words, the rate of deterioration for each individual pipe, depends on its specific influential factors.

2.3.2.1 Influential Factors in Structural Deterioration of Pipes

As sated earlier, the rate of structural deterioration for each pipe will be dependent upon the individual characteristic of pipes. For instance, some of these influential factors are as follows:

- Age of pipe
- Material of pipe
- Diameter of pipe
- Ground water level
- Type of soil/backfill
- Depth of pipe
- Location of pipe

In general, the first three properties, i.e. age of pipe, material of pipe and diameter of pipe, due to the fact that are the typically available properties of the pipes, are the most frequently factors that are considered in studies. Although, once these influential factors are available, depth of pipe and location of pipe can also be considered in models as well [Davis et al. 2001b, Tran 2007]. The impact of each of these factors on the structural deterioration of pipes are presented in the following.

2.3.2.1.1 Effect of Material of Pipes on Structural Deterioration

Based on previous studies, for pipes buried in lower depth and considering similar wheel loads, it is observed that when compared to the maximum stress in a flexible pipe, maximum stress in a rigid pipe (for instance vitrified clay pipe or concrete pipe) will be five times greater. Vitrified clay pipes have a much higher resistance to corrosion and chemical agents, on the other hand, vitrified clay pipes can withstand lower stresses compared to concrete pipes [Moser 2001, Kawabata et al. 2003].

2.3.2.1.2 Effect of Diameter of Pipes on Structural Deterioration

Previous studies suggest that due to the fact that pipes with smaller diameters have less moment of inertia, hence, when compared to pipes with larger diameters, they will possess lower resistance to bending moments [Young and O'Reilly 1983]. Typically, the bending moment has the potential to result in failure in pipes which have diameters of 300 mm or less. For smaller diameter pipes, the loads from the traffic as well as the cover requirements may be underestimated. This underestimation can be the main factor resulting in greater deterioration rates in pipes with smaller diameters [Micevski et al. 2002]. Another study suggests that the major component resulting in larger diameter pipes to have lower deterioration rates, is due to the more experienced staff, greater effort and precision that is used in their installation and laying [Davis et al. 2001b].

2.3.2.1.3 Effect of Depth of Pipes on Structural Deterioration

The impact of surface loads (for instance traffic load) is found to be lessened at lower depths of burial in pipes [Davis et al. 2001a]. With regards to pipes buried in shallow depths and considering the traffic load, in rigid pipes with a depth of two meters as cover, the earth pressure on the pipes were four times less than the pipes with one meter depth of cover (while similar wheel loadings occur) [Kawabata et al. 2003].

2.3.2.1.4 Effect of Location of Pipes on Structural Deterioration

Typically, the location of pipes has significant influence on the surface load on buried pipes. Surface loads include both probabilistic and deterministic loads; probabilistic loads are loads resulted from events such as excavation near pipes or repairs, whereas deterministic loads include cyclic loads as well as traffic loads. When discussing location of pipes, it represents the burial position of pipes; for instance, pipes could be buried under gardens, buildings, roads, footpaths, fields, or railways. Furthermore, depending on the location of burial of pipes, exposure to corrosive agents may increase as well. For instance, if pipes are buried near coast lines, pipes will be exposed to corrosive environments [Davis et al. 2001a, Micevski et al. 2002, Tran 2007].

2.3.2.1.5 Effect of Bus Route on Structural Deterioration of Pipes

Structural deteriorations and subsequently failure of buried pipes occur as a result of external loads; therefore, in some cases wheel loads as well as cyclic loads are represented through the "bus route" factor. This is due to the fact that both the frequency of occurrence and the magnitude of the aforementioned loads can be subject to variation, hence the collection of data becomes a more difficult task. Furthermore, these cyclic loads can be categorized as smaller cyclic occurrences transpiring with various frequencies (for instance, they may be seasonal or daily) and also as large one time occurrences [Hahn et al. 2002].

With regards to the smaller cyclic occurrences, they can stem from regular trucks, or various maintenance operations related to other facilities. On the other hand, larger one time occurrences can stem from events which are not related to construction, such as earthquakes, or landslides; or these occurrences may be due to construction related activities such as surface constructions, and in-ground utility constructions. When these occurrences are overlapped with pipes which have significantly deteriorated, they can result in serious impacts on the pipes [Tran 2007].

2.3.2.1.6 Effect of Sewage Type on Structural Deterioration of Pipes

Although pipes may be properly installed and dynamic loads pose low risks, electrochemical, biochemical, and physical reactions can result in deterioration of the material of pipes, for instance can cause the loss of resistance to loads, and therefore result in susceptibility of pipes to structural deterioration. The major categories of deterioration of materials of pipes are as follows:

- Erosion of invert
- External corrosion

- Internal corrosion

Velocity of the sewer flow, presence of solid sewage material, and the material of the pipes can have significant impacts on the erosion of invert in the pipes. Presence of groundwater as well as acidic surrounding environments (such as acidic soils) can result in corrosion on the outside of the pipes (External corrosion). Additionally, the characteristics of the sewage flow will determine the occurrence and severity of internal corrosions in pipes. For instance, hydrogen sulphide is one of the major causes of internal corrosion in concrete sewer pipes [Hahn et al. 2002].

2.3.2.1.7 Effect of Groundwater Level on Structural Deterioration of Pipes

If groundwater is present, this could result in the ground loss and therefore, affecting the support to underground buried pipes [Davis et al. 2001a].

2.3.2.1.8 Effect of Soil Type on Structural Deterioration of Pipes

Based on the sewer rehabilitation manual, fine sands and silts are considered to pose greater risks of resulting in ground loss; on the other hand, when the soil/backfill is clay, it is considered to be of low risk [WRC 1983]. If the bedding of the pipe is of deformed peaty soil, then due to occurrence of differential settlement by presence of external loading, failure would transpire in pipes. Peaty soil is the kind of soil which when soaked in water will experience deformation [Li 2003]. Compared to sand backfill, by utilizing expanded polystyrene (EPS) as backfill material, the reaction force due to interaction between soil and pipe was observed to drastically fall by 50%-60% in magnitude [Yoshizaki and Sakanoue 2003].

Another example of impact of soil/backfill material on pipes is observed by treatment of adjacent soil sections through cement mixing piles which significantly improves the modulus of elasticity of soil by increasing it as much as ten times. This event improvement in modulus of elasticity of oil will therefore result in a decrease of 56% in horizontal deflection as well as a decrease of 57% in vertical deflection of the buried pipe due to deep excavation occurring near pipe [Li et al. 2003].

2.3.2.1.9 Effect of Age on Structural Deterioration of Pipes

Based on the age of pipe, and depending on the both available design procedures and technologies utilized when manufacturing or installing the pipes, and various other parameters, the age of pipe can illustrate significant variations regarding the structural deterioration of pipes. Developments observed due to technological advances are considered to be the main cause of reduction in the defect rate in sewer pipes within the 25 years after the World War II [Davis et al. 2001a]. Pipes are expected to deteriorate with times, however, due to the fact that the rate of deterioration in pipes can vary drastically from one pipe to another, therefore, if a pipe is older, it might not necessarily be in a lower condition compared to a new pipe [Tran 2007].

2.3.3 Hydraulic Deterioration of Sewer Pipes

Hydraulic deterioration in sewer pipes is associated with the decrease in the cross sectional area of sewer pipes as well as increase in the value of coefficient of roughness [Hahn et al. 2002]. For instance, presence of deposits such as encrustations could result in an increase the in the value of coefficient of roughness. Furthermore, tree root intrusions as well as presence of materials such as silts, metals or various other debris resulting in obstruction within the sewer pipe can cause the cross sectional area of sewer pipes to be reduced dramatically [WRC 1986].

Similar to structural deterioration of sewer pipes, various influential factors could impact the hydraulic condition of sewer pipes. These factors include the following [Tran 2007]:

- Age and type of trees
- Depth of pipe
- Location of pipe
- Soil type

Effect of each of aforementioned influential factors on the hydraulic deterioration of sewer pipes are presented below.

2.3.3.1 Effect of Age and Type of Trees on Hydraulic Deterioration

In general, areas which possess older trees typically have root masses. If structural defects including cracks, fractures, or open joints exist in the sewer pipe, these root masses can find their way into the pipe through these structural defects [WRC 1986]. In certain situations, the biological growth of the roots of trees can ultimately enter the concrete pipes, through the wall of these pipes [ASCE 1994]. In an example of tree roots intrusion, in Victoria (located in Australia) and in areas where Eucalyptus and Melaleuca trees were present close to the pipes, and considering low values of temperature and evaporation levels, blockages in sewer pipes had occurred [Pohls 2001].

2.3.3.2 Effect of Depth of Pipes on Hydraulic Deterioration

Typically, when the depth of burial of pipes are at shallow depths, tree roots intrusion has a higher probability of occurrence and these pipes will probably experience roots intrusion more frequently compared to deeper buried pipes [Pohls 2001]. However, for pipes buried in deeper depths, encrustation can occur due to the fact that groundwater can find its way to the sewer pipe through structural defects of the pipe and result in encrustation in the sewer pipe. Encrustation can transpire when groundwater and sea water, considering they both can contain salt, are partially evaporated and therefore deposits are left within the sewer pipe [WRC 1986].

2.3.3.3 Effect of Location of Pipes on Hydraulic Deterioration

The location of sewer pipes can impact both the level and the type of both deposits and debris which accumulate within the sewer pipes. Deposit sources as well as debris sources are substantially dependent upon the portion of impervious surfaces as well as condition of traffics. The aforementioned sources can include biological materials from vegetation and automobiles, materials transferred (through water) from surrounding soils, as well as dry and wet atmospheric depositions [Tai 1991, Sutherland and Tolosa 2000, Tran 2007]. Additionally, soil can play an important role in the accumulation of materials within the sewer pipes. For instance, it has been observed that the majority of the street surface particles stemmed from erosion of surrounding soils [Tai 1991], and in another study it was realized that 76% of the overall street dust mass stemmed from soil materials [Hopke et al. 1980, Tran 2007].

2.3.3.4 Effect of Soil Type on Hydraulic Deterioration of Pipes

If pipes are buried in soil types wherein a suitable growth condition for tree roots are present, the likelihood of the occurrence of blockage due to intrusions from tree roots is greater. For instance, in a previous study it has been observed that deep sands which do not have lime as well as dark grey sand over clay can be suitable growth conditions for tree roots, therefore, the likelihood of tree roots intrusion in sewer pipes located in such environments will be significantly greater [Pohls 2001, Tran 2007].

2.4 Condition Monitoring of Pipes

In the United States, infrastructures such as bridges and pavements are regularly inspected so that any structural defects occurring in these infrastructures is identified. Specifically, regular inspections are required to be carried out in bridges every two years [Madanat et al. 1995]. Therefore, for infrastructures such as bridges, a database containing regularly inspected data or longitudinal data is available. With the help of these inspection data, decision-making procedure can be conducted in a timely manner for performing required maintenances in these assets. Additionally, these inspection data can also be used for modeling the deterioration rates associated with bridge infrastructures. With the help of these deterioration models as well as longitudinal data, condition of bridges can be predicted for future [Tran 2007].

However, in both the United States and Australia, with regards to sewer pipes, the inspection is typically carried out only one time; therefore, no longitudinal database is available for sewer pipes. The associated database available for sewer pipes is considered to be of snapshot type [Kathula 2001, Kleiner 2001, Tran 2007, Wirahadikusumah et al. 2001, Baik et al. 2006].

2.4 Selection of Sewer Pipe Material

The selection regarding the appropriate materials to be used in sewer pipes is dependent upon various factors and expected conditions of service. Some of the influential factors are stated below [WEF/ASCE 1992]:

- Intended use of sewer
- Abrasion conditions of sewer
- Requirements regarding the installation of sewer pipes
- Conditions affecting corrosion of sewer pipes: For this item, both chemical and biological factors related to surrounding soil as well as within the sewer pipe needs to be taken into consideration
- Requirements of flow within sewer pipes: Includes factors such as slope of the pipes, size of the pipe, flow velocity
- Characteristics of pipe material: Includes factors such as fittings and connection requirements, supplementary protective coatings, cross sectional shapes, strength considerations
- Cost analysis and efficiency: Factors such as installation process, maintenance schedules, estimated durability are considered
- Requirements related to handling of the sewer pipe: Includes factors such as weight of the sewer, resistance to impact
- Major physical properties:

  Regarding physical properties, for rigid sewer pipes, crush strength is a significant factor; however, for flexible sewer pipes, stiffness factor of the pipe, or pipe stiffness needs to be taken into consideration.

Furthermore, condition of soil, pipe loading strength and shear loading strength, and pipe flexural strength needs to be considered as well.

2.5 Classification of Sewer Pipe Material

As stated earlier, sewer pipes can be classified in two different categories:

- Rigid pipes
- Flexible pipes

In the first section materials commonly used in rigid pipes will be discussed; flexible pipes will then be discussed as well.

2.5.1 Rigid Pipe Material

Rigid pipes carry a substantial amount of their basic earth load through the structural strength provided by their corresponding rigid wall capacity. The following denote the most commonly used rigid pipe materials [WEF/ASCE 1992]:

2.5.1.1 Asbestos Cement Pipe

In the past Asbestos Cement Pipe (ACP) was utilized for applications in gravity sewer pipes as well as pressure sewer pipes. Asbestos Cement Pipe is constructed from cement and asbestos fibers. The nominal diameters which this pipe material is used for is from 4 inches up to 36 inches. However, in some cases diameters of 46 inches have been used as well. With regards to gravity drain applications, seven strength classification of ACP were available. Each classification is designated by using the minimum crushing strength of ACP, in units of pounds per linear foot of pipe [WEF/ASCE 1992].

The potential advantages of ACP include the following:

- ➢ The availability of long laying lengths
- ➢ The availability of wide range of strength
- ➢ The availability of wide range of fittings

However, the potential disadvantages of ACP include:

- Corrosion will take place in the presence of acids
- If improperly bedded, shear and beam breakage will occur
- Low value of beam strength

When specifying the ACP, pipe diameter as well as class or strength are used. The ASTM specifications related to ACP are as follows:

- ASTM C 428:
  "Standard Specification for Asbestos-Cement Non-Pressure Pipe"
- ASTM C 644:
  "Standard Specification for Asbestos-Cement Non-Pressure Small Diameter Sewer Pipe,"

2.5.1.2 Cast Iron Pipe

Cast Iron Pipe (CIP) has been utilized in gravity as well as pressure drainage networks. However, in recent years, ductile iron pipe has been taking its place. The nominal diameter for this pipe material ranges from 2 inches up to 48 inches. However, the availability of this pipe material is limited, due to the fact that ductile iron may be preferred instead of cast iron.

Cast Iron Pipes are produced with various thicknesses and strengths. Typically, on the inside of these pipes, cement mortar linings alongside asphaltic seal coatings can be utilized. Moreover, exterior asphalt coatings are generally utilized as well. Various coatings can also be used with CIP [WEF/ASCE 1992].

The potential advantages of cast iron pipe or gray iron include the following:

- The availability of long laying lengths
- High capacity of loading bearing
- High capacity for pressure

Moreover, the potential disadvantages of cast iron pipe or gray iron include:

- Corrosion will take place in the presence of acids
- Limited availability of this product
- Corrosive soils will cause chemical attack
- If improperly bedded, shear and beam breakage will occur
- The weight per length ratio is high

When specifying cast iron pipe, pipe diameter as well as lining, class and the joint type are used. The AWWA and ANSI specifications related to CIP are as follows:

- ANSI A 21.6 (AWWA C 106):
  "Cast Iron Pipe Centrifugally Cast in Metal Molds, for Water or Other Liquid"
- ANSI/AWWA C 110:
  "Gray-Iron and Ductile Iron Fittings, 2 through 48-inch, for Water and Other Liquids"
- ANSI/AWWA C-105~A 21.5
  "Polyethylene Encasement for Gray and Ductile Iron Piping for Water and Other Liquids"
- ANSI A 21.15 (AWWA C 115).
  "Flanged Cast-Iron and Ductile-Iron Pipe with Threaded Flanges"
- ANSI A 21.4 (AWWA C-104)
- "Cement Mortar Lining for Cast-Iron and Ductile Iron Pipe and Fittings for Water"

Further details related to cast iron pipes can be obtained from Ductile Iron Pipe Research (1984).

2.5.1.3 Concrete Pipe

Concrete pipes, with and without reinforcements, can be used as gravity storm drainage. However, prestressed concrete pipes and reinforced concrete pressure pipes can be utilized for both gravity and pressure applications. The nominal diameters available for reinforced concrete pipes and non-reinforced concrete pipes are 12 inches through 200 inches and 4 through 36 inches, respectively. Different linings and coatings can be used with these pipes. Based on the required tightness and the operational pressure, various jointing methods can be used in these pipes.

In the process of producing concrete pipes, various procedures such as vibration, centrifugation, packing and tamping are utilized. Based on the required strength for these pipes, through modified strength of concrete, wall thicknesses, reinforcement and through utilizing prestressing process, the required strengths in concrete pipes can be achieved [WEF/ASCE 1992].

The potential advantages of concrete pipes are as follows:

- The availability of wide range of standard lengths: 4 ft- 24 ft
- The availability of wide range of strength (structural as well as pressure)
- Concrete pipes are resistant to abrasion and galvanic corrosion
- The availability of wide range of diameters
- Losses due to friction are low

Moreover, the potential disadvantages of concrete pipes include the following:

- Corrosion will take place in the presence of acids
- Concrete pipes have significant weight

When specifying concrete pipes, pipe diameter as well as D-load strength or class and the joint type are used. The ASTM and ANSI specifications related to concrete pipes are as follows:

- ANSI/ASTM C 14:
  "Concrete Sewer, Storm Drain, and Culvert Pipe"
- ANSI/ASTM C 76:
  "Reinforced Concrete Culvert, Storm Drain, and Sewer Pipe"

- ANSI/ASTM C 655:
  “Reinforced Concrete Arch Culvert, Storm Drain, and Sewer Pipe”
- ANSI/ASTM C 507:
  “Reinforced Concrete Elliptical Culvert, Storm Drain, and Sewer Pipe”
- ANSI/ASTM C250, C789M, and C789:
  “Reinforced Concrete Box Culvert”
- ANSI/ASTM C 361:
  “Reinforced Concrete Low-Head Pressure Pipe”
- ANSI/ASTM C 443:
  “Joints for Circular Concrete Sewer and Culvert Pipe, Using Rubber Gaskets”
- ANSI/ASTM C 877:
  “External Sealing Bands for Non-Circular Concrete Sewer, Storm Drain, and Culvert Pipe”

2.5.1.4 Vitrified Clay Pipe

Vitrified clay pipes (VCP) are produced by using clay and shales. VCP is typically used for gravity storm drainage. At the temperature wherein the clay mineral particles are fused, the pipe becomes vitrified. The nominal diameters available for vitrified clay pipes are from 3 inches up to 36 inches, and in some cases up to 42 inches. Furthermore, various jointing methods are used with these pipes.

Vitrified clay pipes are produced of both standard-strength and extra-strength categories. In general, for diameters of 12 inches and below, standard-strength vitrified clay pipes are not commonly produced. These pipes are produced up to 10 ft in length. The strengths of these pipes are also dependent on their strength classification and diameter [WEF/ASCE 1992].

The potential advantages of VCP include the following:

- The availability of wide range of fittings for VCP

- Losses due to friction are low
- VCP is highly resistant to abrasion
- VCP is highly resistant to chemical corrosion

Moreover, the potential disadvantages of VCP include:

- VCP has significant weight
- The availability of sizes is limited
- If improperly bedded, shear and beam breakage will occur
- The beam strength in VCP is low

When specifying vitrified clay pipes, pipe diameter as well as strength and the joint type are used. The ASTM, ANSI, and NCPI specifications related to vitrified clay pipes are as follows:

- ANSI/ASTM C 700:
  "Standard Specification for Vitrified Clay Pipe, Extra Strength, Standard Strength and Perforated"
- ASTM C 425:
  "Compression joints for Vitrified Clay Pipe and Fittings "
- ANSI/ASTM 301:
  "Pipe, Clay, Sewer," Federal Specification SS-P361d, Standard Methods of Testing Vitrified Clay Pipe"
- NCPI ER4-67:
  "Crushing Strength for Pipe and Fittings for Perforated VCP"

2.5.2 Flexible Pipe Material

Unlike rigid pipes where they carry a substantial amount of their basic earth load through the structural strength provided by the rigid wall capacity of the pipe itself, in flexible pipes, their load carrying capacity is obtained through a summative contribution of both the pipe strength as well as the embedment soil. The load carrying capacity through embedment soil depends on the interaction between pipe and soil and also the amount of deflection of the pipe. The following denote the most commonly used flexible pipe materials [WEF/ASCE 1992]:

2.5.2.1 Ductile Iron Pipe

Ductile Iron Pipe (DIP) is utilized for applications in both gravity as well as pressure drain systems. In the process of producing ductile iron pipe, cerium or magnesium is added to cast iron (gray iron) right before the casting procedure. Ductile iron pipes are produced in lengths up to 20 ft and with nominal diameters of 3 inches up to 54 inches. Both ductile iron fittings as well as cast iron fittings can be utilized for ductile iron pipes. Similar to previously mentioned pipes, there are a number of jointing methods for DIP as well.

Ductile iron pipes are typically used in circumstances where the factors mentioned below are required to be considered:

- High impact capacity
- High loading capacity
- Minimum cover thickness
- Long service life
- Minimum maintenance

Similar to other pipes, ductile iron pipes are available in different classes, strengths, and thicknesses. On the interior of these pipes, various lining choices can be used. For instance, cement mortar lining with asphaltic coating, epoxies and polyethylene can be used as interior linings. Generally, asphaltic coatings as well as polyethylene exterior wrappings can be used as exterior coatings [WEF/ASCE 1992].

The potential advantages of DIP are as follows:

- Losses due to friction are low
- The availability of long laying lengths
- Ductile iron pipes have significant impact strength

- High capacity of loading bearing
- High capacity for pressure
- Ductile iron pipes have significant beam strength

Moreover, the potential disadvantages of ductile iron pipe include the following:

- Ductile iron pipe has significant weight
- Corrosion will take place in the presence of acids
- Corrosive soil will cause chemical attack in the pipe

When specifying ductile iron pipe, pipe diameter as well as lining, class, and the joint type are used. The ASTM, ANSI, and AWWA specifications related to ductile iron pipes are as follows:

- ANSI A 21.5 (AWWA C 10):
  "Polyethylene Encasement for Gray and Ductile Cast-Iron Piping for Water and Other Liquids"
- ASTM A 746:
  "Ductile Iron Gravity Sewer Pipe"
- ANSI/AWWA C 110:
  "Gray-Iron and Ductile Iron Fittings. 3 inch through 48 inch, for Water and Other Liquids"
- ANSI A 21.4 (AWWA C 104):
  "Cement Mortar Lining for Cast-Iron and Ductile-Iron Pipe and Fittings for Water"

2.5.2.2 Fabricated Steel Pipe

This section includes various elements such as corrugated steel pipes, arches, as well as pipe arches will be covered. Galvanized corrugated steel is available in various shapes for different conduits; furthermore, supplementary coatings can be implemented as additional protection as well. The following are some of the different shapes and features for these pipes:

- Pipes in circular shapes with diameters ranging from 12 inches up to 144 inches
- Pipe arches produced from circular profiles with diameters of 15 inches up to 120 inches
- Structural plate structures with diameters ranging from 60 inches up to 312 inches
- Structural plate arches with concrete bases with spans of 6 ft up to 25 ft in length

These pipes are available in standard lengths of 20, 30, and 40 ft and in multiples of 2 ft and 4 ft. Welded seam pipe or lock seam pipe are produced from continuous coils and therefore can be cut to desired lengths. This allows the selection of various lengths of these sections.

By using coupling bands, the sections are joined. These coupling bands can be single piece, two piece or they can be internal expanding types which are used in lining procedures. Sections with larger sizes, such as structural plate conduits, are field bolted. Additionally, helical corrugations which allow for enhanced flow properties, are also available among corrugated pipes. Moreover, corrugated pipes can be used particularly for jacking purposes [WEF/ASCE 1992].

The potential advantages of these pipes are as follows:

- The availability of long laying lengths
- Providing sections with lighters weights
- Providing flexibility in sections
- The availability of wide range of coatings
- Can be useful for application of lining in repair of structures

Moreover, the potential disadvantages of these sections include the following:

- Corrosion will take place in aggressive environments
- Low hydraulic coefficient (this coefficient can be increased by using linings of

bituminous materials)

- Needs special attention for bedding
- The springline of the upper semi-circular arch in pipe or culvert requires satisfactory horizontal support

As stated earlier, in order to improve the hydraulic performance of these sections, bituminous linings can be utilized. Bituminous linings create smooth surfaces by covering the crests of the corrugations. Therefore, these smooth surfaces enhance the hydraulic properties of the flow. Additionally, in order to improve the hydraulic performance of these sections, the invert can also be paved as well. The main issue when using coatings with corrugated steel sections is the bonding between the pipe and the coating.

When coating is used with these sections, the durability of these pipes is expected to increase. However, the coatings will be damaged before the full expected service life of these pipes. If the materials are expected to be frozen and thawed frequently, in these situations, smooth coated corrugated steel pipes should not be utilized. Furthermore, based on the surrounding soil conditions, there may be a need for using external corrosion protection as well. Bituminous coatings are flammable and petroleum waste or solvents can result in damage or destruction of these coatings. During installation of coated pipe, special care should be paid so that the coating remains intact. In order to make sure that the pipe is structurally stable, continuous and sufficient lateral support is required.

When specifying these pipes, the following are used:

- Size of pipe including nominal diameter, span and rise, or length of arc
- Shape of pipe including circular, elliptical, arch, or segment plate arch
- Gage of metal (based on the required strength)
- Assembly by using bolts or bands
- Linings or coatings
- Coupling (single piece and two-piece widths)

2.5.2.3 Corrugated Aluminum Pipe

Various shapes and sizes of corrugated aluminum pipes, arches, and box culverts are available. The following are the general properties for these sections:

- Pipes with circular shapes are available with diameters ranging from 6 inches up to 180 inches
- Arches are available with spans of up to 30 ft in length and with rises of up to 14 ft
- Box culverts can be selected with spans up to 25 ft-5 inches in length and with rises of up to 10 ft-2 inches

The pipes strength can be noted by using range of gages, the type of joints, which can be bolted, welded, or achieved by using mechanical coupling. Furthermore, bedding as well as backfilling methods are also used in specifying the pipe strength. Aluminum pipe sections are lighter in weight and therefore are easier to manage. Aluminum pipe sections are available in up to 40 ft in length. Box culverts and arches with larger sizes are typically filed bolted. On the other hand, smaller sizes of aluminum sections have great versatility to be fabricated in the field and can be cut and welded on site [WEF/ASCE 1992].

The potential advantages of these sections are as follows:

- The availability of long laying lengths
- Providing sections with lighters weights
- Providing flexibility in sections
- Fabrication and connections can be done on site
- Resistant to corrosive environments (e.g. saltwater)

Moreover, the potential disadvantages of these sections include the following:

- Low hydraulic coefficient (this coefficient can be increased by using linings of bituminous materials)
- Similar to steel pipe, requires satisfactory horizontal support

The specifications related to these sections are as follows:

- AASHTO Designation M-196:
  "Standard Specification for Corrugated Aluminum Alloy Culverts and Underdrains"
- AASHTO, Designation M-219:
  "Standard Specification for Aluminum Structural Plate for Pipe, Pipe Arches, and Arches"
- AASHTO, Designation M-197:
  "Standard Specification--For Clad Aluminum Alloy Sheets for Culverts and Underdrains"
- Federal Specification - Pipe, Corrugated (Aluminum Alloy) WW-P-402

2.5.2.4 Thermoplastic Pipes

Various thermoplastic pipes can be used in applications for either sanitary sewer pipes or drainage systems. Some of these thermoplastic pipe materials are as follows:

- Acrylonitrile-Butadiene-Styrene (ABS),
- Polyethylene (PE),
- Polyvinyl Chloride (PVC).

In general, thermoplastic pipes are wide range of plastic materials which through adjusting temperatures and by heating and hardening can be used for manufacturing pipes. Below, some of these pipe materials are discussed.

2.5.2.4.1 Acrylonitrile-Butadiene-Styrene Pipes

Acrylonitrile-Butadiene-Styrene (ABS) pipes have applications in both gravity and pressure drainage systems. The nominal diameters in which gravity acrylonitrile-autadiene-atyrene pipes are available include ranges of 3 inches up to 12 inches. Furthermore, the lengths of these pipes are up to 35 ft. Various jointing systems as well as fittings are available for these pipes.

Through extrusion of acrylonitrile-autadiene-atyrene plastic, ABS pipes are produced. Acrylonitrile-Butadiene-Styrene (ABS) pipes are available in three different ratios; dimension ratios (DR) represent the ratio of mean outside diameter of the pipe to the minimum thickness of wall. The values of dimension ratios are as follows: 23.5, 35, 42. These dimension ratios are selected based on the diameters of the pipes. The pipe stiffness (PS) values for each of the above dimension ratios are respectively as follows (in units of psi): 150, 45, 20.

The potential advantages of ABS pipes are as follows:

- The availability of long laying lengths
- Providing sections with lighters weights
- Providing high values of impact strength
- Facilitates cutting and tapping on site

Moreover, the potential disadvantages of ABS pipes include the following:

- The available sizes of ABS pipes are limited
- ABS pipes can be affected by environmental stress cracking
- Specific organic chemicals can affect ABS pipes
- Long term UV exposure can alter the surface of ABS pipes
- If improperly bedded and haunched, excessive deflection will occur

When specifying acrylonitrile-autadiene-atyrene pipes, pipe diameter as well as dimension ratio, pipe stiffness, and the joint type are used. The ASTM and ANSI specifications related to ABS pipes are as follows:

- ANSI/ASTM D 2751:
  "Acrylonitrile-Butadiene-Styrene (ABS) Sewer Pipe and Fittings"
- ANSI/ASTM D 2235:
  "Solvent Cement for Acrylonitrile-Butadiene-Styrene (ABS) Plastic Pipe and Fittings"

- ANSI/ASTM F 477:
  "Elastomeric Seals (Gaskets) for Jointing Plastic Pipe"
- ANSI/ASTM D 3212:
  "Joints for Drain and Sewer Plastic Pipes Using Flexible Elastomeric Seals"
- ANSI/ASTM F 545:
  "PVC and ABS Injected Solvent Cemented Pipe Joints"

2.5.2.4.2 Acrylonitrile-Butadiene-Styrene Composite Pipes

Acrylonitrile-Butadiene-Styrene (ABS) composite pipes have applications in gravity drainage systems. The nominal diameters in which gravity acrylonitrile-autadiene-atyrene composite pipes are available include ranges of 8 inches up to 15 inches. Moreover, the lengths which these pipes are available are from 6.25 ft up to 12.5 ft. Various ABS systems fittings can be utilized for these pipes. Furthermore, the elasatomeric gasket joints as well as solvent cemented joints are some of the jointing systems that can be used in ABS composite pipes.

Through extrusion of acrylonitrile-autadiene-atyrene plastic, and by using a series of truss annuli, ABS composite pipes are produced; these truss annuli are filled with materials such as light weight Portland cement concrete.

The potential advantages of ABS composite pipes include the following:

- The availability of long laying lengths
- Providing sections with lighters weights
- Facilitates cutting on site

Moreover, the potential disadvantages of ABS composite pipes are as follows:

- The available sizes of ABS composite pipes are limited
- ABS composite pipes can be affected by environmental stress cracking
- Specific organic chemicals can affect ABS composite pipes
- Long term UV exposure can alter the surface of ABS composite pipes

- If improperly bedded rupture may occur

The ASTM and ANSI specifications related to acrylonitrile-autadiene-atyrene composite pipes are as follows:

- ANSI/ASTM D 2680:
  "Acrylonitrile-Butadiene-Styrene (ABS) Composite Sewer Piping"
- ANSI/ASTM D 2235:
  "Solvent Cement for Acrylonitrile-Butadiene-Styrene (ABS) Plastic Pipe and Fittings"
- ANSI/ASTM F 477:
  "Elastomeric Seals (Gaskets) for Jointing Plastic Pipe"
- ANSI/ASTM D 3212:
  "Joints for Drain and Sewer Plastic Pipes Using Flexible Elastomeric Seals"

2.5.2.4.3 Polyethylene Pipes

Polyethylene (PE) pipes have applications in both gravity and pressure drainage systems. The nominal diameters in which gravity polyethylene pipes are available include ranges of 4 inches up to 48 inches. Gravity polyethylene pipes are generally used for relining purposes in sewer pipes. Butt fusion as well as flanged adapters are utilized in jointing procedures. Through extrusion of polyethylene plastic PE pipes are produced

The potential advantages of polyethylene pipes include the following:

- The availability of long laying lengths
- Providing sections with lighters weights
- Facilitates cutting on site
- Provided high values of impact strength

Moreover, the potential disadvantages of polyethylene pipes are as follows:

- The available sizes of polyethylene pipes are limited

- If improperly bedded and haunched, excessive deflection will occur in polyethylene pipes
- Specific organic chemicals can affect polyethylene pipes
- Values of pipe stiffness and tensile strength are low
- Long term UV exposure can alter the surface of polyethylene pipes
- In order to fuse joints, specific tools are needed

When specifying polyethylene pipes, pipe diameter (inner or outer diameter) as well as material type, dimension ratio, and the joint type are utilized.

2.5.2.4.4 Polyvinyl Chloride Pipes

Polyvinyl Chloride (PVC) pipes have applications in both gravity and pressure sanitary sewer and drainage systems. The nominal diameters in which gravity polyvinyl chloride pipes are available include ranges of 4 inches up to 27 inches. Both pressure and gravity fittings are available for these pipes. The maximum length for which polyvinyl chloride pipes are typically available is 20 ft. Elastomeric seal gasket joints is mainly utilized for jointing purposes. However, in particular situations, solvent cement joints can also be used with polyvinyl chloride pipes.

Similar to previously mentioned pipes, through extrusion of particular plastic, polyvinyl chloride pipes are produced. polyvinyl chloride pipes are available in three different ratios; dimension ratios (DR) represent the ratio of mean outside diameter of the pipe to the minimum thickness of wall. The values of dimension ratios are as follows: 35, 41, 51. These dimension ratios are selected based on the diameters of the pipes. The pipe stiffness (PS) values for each of the above dimension ratios are respectively as follows (in units of psi): 46, 28, 80.

The potential advantages of PVC pipes are as follows:

- The availability of long laying lengths
- Providing sections with lighters weights

- Providing high values of impact strength
- Facilitates cutting and tapping on site

Moreover, the potential disadvantages of PVC pipes include the following:

- The available sizes of PVC pipes are limited
- Specific organic chemicals can affect PVC pipes
- Long term UV exposure can alter the surface of PVC pipes
- If improperly bedded and haunched, excessive deflection will occur

When specifying polyvinyl chloride pipes, pipe diameter as well as dimension ratio, pipe stiffness, and the joint type are used. The ASTM and ANSI specifications related to PVC pipes are as follows:

- ANSI/ASTM D 3034:
  "Type PSM Polyvinyl Chloride (PVC) Sewer Pipe and Fittings"
- ANSI/ASTM D 3033:
  "Type PSP Polyvinyl Chloride (PVC) Sewer Pipe and Fittings"
- ANSI/ASTM F 477:
  "Elastomeric Seals (Gaskets) for Jointing Plastic Pipe"
- ANSI/ASTM D 3212:
  "Joints for Drain and Sewer Plastic Pipes Using Flexible Elastomeric Seals"
- ANSI/ASTM F 545:
  "PVC and ABS Injected Solvent Cemented Pipe Joints"
- ANSI/ASTM D 2564:
  "Solvent Cements for Polyvinyl Chloride (PVC) Plastic Pipe and Fittings"
- ASTM F 679:
  "Standard Specification for Polyvinyl Chloride (PVC) Large Diameter Plastic Gravity, Sewer Pipe and Fittings"

2.5.2.5 Thermoset Plastic Pipes

In general, thermoset plastic pipes are classified in two different categories as follows:

- Reinforced thermosetting resin (RTR)
- Reinforced plastic mortar (RPM)

A wide range of plastic materials are included in the thermoset plastic category. Once thermoset plastic materials are cured, for instance through heating procedure, they become significantly insoluble and infusible. Below the aforementioned pipes are described.

2.5.2.5.1 Reinforced Thermosetting Resin Pipes

Reinforced thermosetting resin (RTR) pipes have applications in both gravity and pressure drainage systems. Reinforced thermosetting resin are produced based on ASTM specifications and the nominal diameters available for this product ranges from 1 inch up to 12 inches. However, when produced based on particular manufacturers' specifications, nominal diameters of 12 inches up to 144 inches are also available for reinforced thermosetting resin pipes. For large diameters RTR pipes, fittings need to be produced based on particular requirements; on the other hand, for small diameter reinforced thermosetting resin pipes, RTR fittings are available. Various jointing methods can be used for reinforced thermosetting resin pipes. Different protections on the interior of the RTR pipes can be used; for instance, thermosetting or thermoplastic liners and coatings can be considered as protections inside RTR pipes.

When manufacturing reinforced thermosetting resin pipes, firbrous reinforcements are used. Examples of these fibrous reinforcements are fiberglass which is surrounded by or embedded within the cured thermosetting resin. The methodologies utilized in producing reinforced thermosetting resin pipes include the following:

- Centrifugal casting
- Pressure laminating
- Filament winding

Insituform process is an example of reinforced thermosetting resin pipe material. Insituform process is typically utilized in order to rehabilitate the existing pipes. In this procedure, a liquid thermosetting resin impregnated a polyester fiber tube which has impermeable layer on one side of it. From an access point, for instance a manhole, the tube is placed within the existing pipe; then by using cold water, the tube will be pushed towards the pipe material.

Once the tube is pushed inside the pipe, the temperature of the water inside will be increased, therefore, it will cure the tube. This will leave a pipe inside the existing pipe. The Insituform procedure can be used for a wide range of pipe diameters starting from 4 inches up to several feet. Moreover, this procedure can be used for conduits with various shapes. The thickness of wall is noted as the standard dimension ratio (SDR) and can be chosen based on the structural requirements.

Furthermore, based upon the specific corrosion requirements, the thermoset resin can be chosen to meet this requirement. Once this procedure is completed, due to further smoothness of the inner surface of the pipe, hydraulic performance will be increased; therefore, even though the sectional area for the flow was reduced, due to reduction in the dimensions of the conduit, however, smoothness will compensate for that.

The potential advantages of reinforced thermosetting resin pipes are as follows:

- The availability of long laying lengths
- Providing sections with lighters weights

Moreover, the potential disadvantages of reinforced thermosetting resin pipes include the following:

- Strain corrosion may occur in specific environments
- Specific organic chemicals can affect RTR pipes
- Long term UV exposure can alter the surface of RTR pipes
- If improperly bedded and haunched, excessive deflection will occur in RTR pipes

When specifying reinforced thermosetting resin pipes, pipe diameter as well as method of production, coating and lining, plastic material, pipe stiffness, and the joint type are used. The ASTM and ANSI specifications related to RTR pipes are as follows:

- ASTM D 2996:
  "Filament-Wound Reinforced Thermosetting Resin Pipe"
- ANSI/ASTM D 2997:
  "Centrifugally-Cast Reinforced Thermosetting Resin Pipe"
- ASTM D 2310:
  "Machine-Made Reinforced Thermosetting Resin Pipe"

2.5.2.5.2 Reinforced Plastic Mortar Pipes

Reinforced plastic mortar (RPM) pipes have applications in both gravity and pressure sanitary sewer systems. The nominal diameters available for reinforced plastic mortar pipes range from 8 inches up to 144 inches. Various jointing methods can be used for reinforced plastic mortar pipes. Different protections on the interior of the RPM pipes can be used; for instance, thermosetting or thermoplastic liners and coatings can be considered as protections inside RPM pipes.

When manufacturing reinforced plastic mortar pipes, firbrous reinforcements are used. Examples of these fibrous reinforcements utilized in RPM pipes are aggregates (for instance sand) as well as fiberglass which is surrounded by or embedded within the cured thermosetting resin.

The potential advantages of reinforced plastic mortar pipes are as follows:

- The availability of long laying lengths
- Providing sections with lighters weights

Moreover, the potential disadvantages of reinforced plastic mortar pipes include the following:

- Strain corrosion may occur in specific environments
- Specific organic chemicals can affect RPM pipes
- Long term UV exposure can alter the surface of RPM pipes
- If improperly bedded and haunched, excessive deflection will occur in RPM pipes

When specifying reinforced plastic mortar pipes, pipe diameter as well as beam strength, coating and lining, hoop tensile strength, plastic material, pipe stiffness, stiffness factor, and the joint type are used. The ASTM and ANSI specifications related to RPM pipes are as follows:

- ASTM D 3754:
  "Reinforced Plastic Mortar Sewer and Industrial Pressure Pipe"
- ANSI/ASTM D 3252:
  "Reinforced Plastic Mortar Sewer Pipe"

2.6 Inspection Methodologies in Pipes

Based upon the capabilities of inspection methodologies as well as the information needed by the asset manager, these methodologies can be categorized in three different levels as presented below [Ratliff 2003]:

- Level 1: Field reconnaissance
- Level 2: Internal inspection
- Level 3: External inspection

2.6.1 Field Reconnaissance in Pipes

In field reconnaissance, the associated data for manholes, pits and pipes are gathered and furthermore, the manhole structures will be subject to evaluation for accessibility of both inspectors as well as the required equipment for conducting the inspection. Below are some of the available techniques for field reconnaissance:

- Sonde locators

- Manhole survey
- Global positioning system

At this stage, the existing drawings as well as the current available information are studied. Furthermore, if new information is available at any time, i.e. pipe replacement or repair occurs, the tasks associated with this stage need to be continued [Tran 2007].

2.6.2 Internal Inspection of Pipes

At this stage, the internal inspection of pipes is intended to gather data with regards to the condition of the pipes on the inside. Therefore, by collecting the inspection data at this stage, the required considerations which need to be taken, depending on the condition of pipes, to prevent pipes from undergoing failure or collapse, or experiencing blockage of cross sectional area, will be identified. The following are examples of methodologies for this stage of inspection [Tran 2007]:

- Sonar
- Closed circuit television
- Man-walk through
- Focused electrode leak location
- Sewer scanner and evaluation technology
- Multi-sensor pipe inspection systems
- Laser-based scanning systems

2.6.3 External Inspection of Pipes

This level of inspection of pipes is associated with studying the condition of soil surrounding the pipe which is considered to provide the support for pipes. If voids or loss of support from soil transpire, the pipe can be subjected to collapse [Ratliff 2003]. At this stage of inspection, the following methodologies can be utilized:

- Ground penetrating radar

- Micro deflection
- Infrared thermographs
- Impact echo wave impedance probe

With regards to various inspection methodologies, there have been studies conducted with review related to these methodologies [Wirahadikusumah et al. 1998, Morrison and Thomson 2003, Koo and Ariaratnam 2006].

2.7 Condition Grading of Assets

The first condition grading system which included protocols as well as guidelines pertaining to evaluation of existing condition of each pipe by utilizing the data obtained based on the inspection performed through closed-circuit television was presented by Water Research Center based in UK [Tran 2007, WRC 1986]. Some other condition grading systems were subsequently devised based on the condition grading system proposed by Water Research Center in Europe, Canada, and Australia as well [McDonald and Zhao 2001, Cemagref 2003, WSAA 2002]. Even though the deterioration in pipes (either hydraulic deterioration or structural deterioration) are considered to be continuous, in the aforementioned condition grading systems, in order to record the current condition of pipes, ordinal gradings were attributed to the deteriorations observed in pipes. For instance, in WRC the proposed condition gradings were ordered from 1 to 3 [WRC 1986]. In this grading system, condition grading 1 demonstrated the perfect condition for the pipe, and condition gradings 2 and 3 demonstrated pipes in fair and poor conditions, respectively.

With regards to other infrastructures, similar approaches have been taken in proposing condition grading of assets. For instance, in pavements and bridge infrastructures, the deteriorations of assets were graded from 1 to 8 and 0 to 9, respectively [Madanat et al. 1995]. The condition grades utilized for assessment purposes illustrate a relative ordering rather than illustrating the distance between these grades. Additionally, utilizing these grading systems helps with decreasing the intricacy of the decision-making procedure [Madanat et al. 1997].

In Australia, in 1991 and in order to increase awareness with regards to deterioration of assets, the Australia Conduit Condition Evaluation Manual (ACCEM) was provided by Sydney Water. However, later in year 2002, Water Service Association of Australia provided the Sewer Inspection Reporting Code (SIRC) [WSAA 2002]. In this grading system, grades of 1 through 3 were used for condition evaluation of assets. The Sewer Inspection Reporting Code (SIRC) was used for condition evaluation of rigid sewer pipes, such as concrete pipes as well as vitrified clay pipes. Later, the Conduit Inspection Reporting Code (CIRC) was developed as an update to the Sewer Inspection Reporting Code (SIRC). Unlike SIRC, the Conduit Inspection Reporting Code (CIRC) included flexible pipes as well as rigid pipes and had grading systems from 1 through 5. The tables presented below, demonstrate the associated hydraulic as well as structural conditions for each condition grading [WSAA 2006, Tran 2007].

Table 2.1: Condition grading and corresponding attribution based on WSAA 2002 [Tran 2007]

Condition grading of asset	WSAA 2002	
	Hydraulic Deterioration	Structural Deterioration
1	No apparent need for action	No apparent need to investigate further
2	Consider response on a program basis	Consider overall circumstances on a program basis
3	Appropriate action to be investigated urgently	Urgent need to investigate overall circumstances

Table 2.2: Condition grading and corresponding attribution based on WSAA 2006 [Tran 2007]

Condition grading of asset	WSAA 2006	
	Hydraulic Deterioration	Structural Deterioration
1	No or insignificant loss of hydraulic performance has occurred. Appears to be in good condition	Insignificant deterioration of the sewer has occurred. Appears to be in good condition
2	Minor defects are present causing minor loss of hydraulic performance	Minor deterioration of the sewer has occurred
3	Developed defects are present causing moderate loss of hydraulic performance	Moderate deterioration has occurred but defects do not affect short term structural integrity
4	Significant defects are present causing serious loss of hydraulic performance	Serious deterioration of the sewer has occurred and affected structural integrity
5	Failure of the sewer has occurred or is imminent	Failure of the sewer has occurred or is imminent

2.8 Condition Grading Utilizing PACP

Pipeline Assessment and Certification Program (PACP) is the North American Standard utilized for evaluation of pipelines. National Association of Sewer Service Companies (NASSCO) has provided PACP in partnership with the Water Research Center based in UK. Utilizing the coding proposed by NASCOO, the defects and features of the pipes can be categorized in five different categories. These categories are presented below [EPA 2015]:

- Structural defects
- Operational and maintenance
- Continuous defects
- Construction features
- Miscellaneous features

For each of these defects, a combination of capital letters is utilized to illustrate the defects. For instance, longitudinal fracture in pipes is represented by FL, and circumferential cracks are illustrated by CC. In addition to this letter combination, the numeric values associated with each of these defects illustrate the severity of these defects [EPA 2015].

The previously stated five categories are briefly discussed below [EPA 2015]:

- Structural defects: Structural defect codes comprise of various coding categories. These codes are utilized to identify the deteriorations in pipes which are related to structural degradation of pipes and can stem from numerous causes such as breaks, fractures, and cracks.
- Operational and maintenance defects: Operational and maintenance (O&M) defects are utilized to describe the defects which are originated from lack of maintenance. These defects can include infiltration, roots intrusion, deposits, etc.
- Continuous defects: Continuous defects coding is categorized in two different classes:

  1) Truly continuous defects: Truly continuous defects run along the sewer pipe

2) Repeated continuous defects: Repeated continuous defects can transpire in regular intervals along the sewer pipe

- Construction features: Construction features coding describes the defects which are positioned within or around the sewer pipe. Examples of construction features are intruding seal materials and taps
- Miscellaneous features: Miscellaneous features contains various categories. By using these coding features, miscellaneous features can be demonstrated. Various other letters are combined with the miscellaneous coding letter to describe the specific defect type within miscellaneous class.

The gradings assigned based on Pipeline Assessment and Certification Program (PACP) coding can be described as follows [NASSCO 2015, EPA 2015]:

- Grade 1: Demonstrates minor defect
- Grade 2: Demonstrates minor to moderate defect
- Grade 3: Demonstrates moderate defect
- Grade 4: Demonstrates significant defect
- Grade 5: Demonstrates most significant defect

Furthermore, the relative likelihoods of failure with respect to time based on PACP coding can be described based on the gradings as follows [NASSCO 2015, EPA 2015]:

- Grade 1: Demonstrates unlikely failure in the foreseeable future
- Grade 2: Demonstrates unlikely failure for at least 20 years
- Grade 3: Demonstrates that the deterioration may continue at a timeframe of 10 to 20 years
- Grade 4: Demonstrates risk of failure within 5 to 10 years
- Grade 5: Demonstrates need of immediate attention and illustrated failed segment of pipe or a likely failure in the next 5 years

Moreover, Pipeline Assessment and Certification Program (PACP) also utilizes a quick rating system as well. The quick ratings system demonstrates the frequency by which the two highest grades have been transpired [EPA 2015].

2.9 Deterioration Models

Various types of models may be utilized for different civil infrastructures; for instance, models may be of the following natures [Morcous et al. 2002a]:

- Deterministic,
- Statistical
- Artificial intelligence (also known as soft computing methods)

Moreover, in an upper level, each of these deterioration models maybe viewed as model-driven or data-driven. As an example, deterministic modeling is considered a model-driven approach. This is due to the fact that the detailed construction of these models is dependent upon the expert view. The same reasoning holds true for statistical deterioration models and thus they are also categorized as model-driven approach [Dasu and Johnson, 2003].

On the other hand, data-driven models are deterioration models which are constructed based upon the available data (sample data available from inspections). Deterioration models constructed based on soft-computing or artificial intelligence approach, are considered as data-driven models [Dasu and Johnson, 2003].

Furthermore, each of the abovementioned modeling classifications contains various modeling techniques as presented below [Tran 2007]:

Deterministic Models:

1) Linear Models
2) Exponential Models

Statistical Modeling

1) Markov Chain Models
2) Ordinal Regression
3) Linear Discriminant Analysis

Artificial Intelligence Models (also known as soft computing methods)

1) Case-Based Reasoning
2) Fuzzy Set Theory
3) Neural Networks

The following section contains a brief discussion pertaining to each of these models.

2.9.1 Modeling based upon deterministic approach

Deterministic approach modeling is suitable for occurrences in which the associated relationship between the dependent variable and the independent variables can be expressed by certainty. For instance, power law expressions and time linear relationship fall into deterministic modeling category; these relationships were applied for pavements by Lou et al. in 2001, and in water main by Kleiner and Ranjani in 2001. Furthermore, linear and exponential models are also some of the commonly used approaches for expressing the relationship between the dependent variable and the independent variables in deterioration models. The following section describes application of these methods to obtain deterioration rates.

2.9.1.1 Linear deterioration models

With regards to civil infrastructures, in order to construct a linear model, the following steps can be taken [Madanat et al. 1995]:

- *Step one:* By taking into account the similarities of predictors among various assets, cohorts containing specific facilities are created. Some commonly used predictors are type of material for each asset, size of assets, type of loading exerted on facilities, etc.

- *Step two:* Once the cohorts of facilities are provided, by considering the age of facilities as the independent variable and the condition rating of assets as the dependent variable, the linear deterioration model is thus expressed as follows:

  $$CR_i = a_0 + a_1 \times t + \epsilon_i \tag{2.1}$$

  In which the parameters are as follows:

  CR_i denotes the condition rating of i^{th} facility,

  a_0 is the intercept term of the linear model,

  a_1 denoted the coefficient of age in the model, and

  ϵ_i is the random error associated with the linear model.

Typically, the least square method is the calibration technique which is used in the linear models [Aldrich and Nelson 1990]. Therefore, when using linear models, the rate of deterioration of infrastructures is supposed to be constant throughout the service life of the assets, whereas the actual deterioration of civil infrastructures is a mixture of both damages transpired from random events as well as time dependent processes resulting in deterioration [Morcous et al. 2002b]. Additionally, discrete condition ratings are not suitable to be modeled by utilization of linear deterioration models [Madanat and Ibrahim 1995, Madanat et al. 1997].

2.9.1.2 Exponential deterioration models

Another deterministic approach in modeling the deterioration rate of civil infrastructure is the exponential model. Generally, exponential deterioration models are used for the cases which it is believed the rate of degradation of assets is increasing as the asset ages. For instance, in the case of sewer pipes, assuming that the deterioration rate pertaining to sewer pipes increases by the age of pipe, Wirahadikusumah et al. (2001) used the exponential approach to model the deterioration of sewer pipes in the city of Indianapolis resulting in a model as follows:

$$CR_i = \exp\left(a_0 + a_1 \times t + \epsilon_i\right) \tag{2.2}$$

In which the parameters are similar to that of linear models, although the intercept and the coefficient of age of pipe are within the exponential function. Additionally, these models are calibrated through least square technique.

2.9.2 Limitations of deterministic deterioration models

Even though the deterministic approach in modeling the deterioration rate of infrastructures provides an analytical method which seems to be versatile, the following limitations are still present in this type of modeling:

The only contributing factor which is present in the stated deterioration models is the age of the sewer pipes. Albeit using cohorts of assets, other independent variables are not explicitly present in the relationship describing the deterioration rate. Furthermore, on one hand the range each cohort covers needs to be as narrow as possible in order to result in a homogenous model; and on the other hand, the range of variables in the cohorts need to be sufficiently wide so that they can cover influence of all other predictors contributing to the deterioration of infrastructures [Kleiner et al. 2007].

Based on the abovementioned facts, when using the stated deterministic models for obtaining the deterioration rate of infrastructures, the intricacies involved within the interactions of various independent variables cannot be introduced to the model in a comprehensive approach [Mishalani and Madanat 2002]. Even though creating cohorts of pipes is a step in this direction.

Furthermore, there are numerous uncertainties associated with infrastructures which can vary substantially from one civil infrastructure to another based on their particular applications. As an example, the uncertainties associated with highway bridge structures vary from those present for sewer pipes. For instance, random occurrences resulting in damages can be considered for both of these infrastructures, even though the nature and instances may vary [Morcous et al. 2002b].

Therefore, the aforementioned uncertainties demonstrate the need for a probabilistic approach for identifying the deterioration rate and service life of sewer pipes. However, as stated earlier, deterministic approach in modeling the deterioration rate of infrastructures does not fully cover the probabilistic outcomes predating to service life of these assets; neither do they represent the time-dependent properties influential in deterioration of sewer pipes. Additionally, as it is stated earlier, linear deterioration model as well as exponential deterioration models do not suitably represent the deterioration rates within discrete condition rating system [Madanat and Ibrahim 1995, Madanat et al. 1997].

2.9.3 Statistical deterioration models

With regards to these deterioration models, statistical theory is utilized; in which the effect of random noises within elements are accounted for. Statistical modeling approach has been applied in numerous engineering areas [Johnson and Albert, 1999; Henley and Kumamoto, 1992; and Kuzin and Adams, 2005]. As mentioned earlier, the statistical deteriorating models are categorized as model-driven approach; moreover, in statistical models, relationship between the outcome, i.e. the dependent variable, and the independent variables is assumed to be probabilistic, and therefore, compared to deterministic deterioration models, it provides a more appropriate methodology for considering the uncertainties associated with each of independent variables and their relationship with the dependent variable [Dasu and Johnson, 2003].

2.9.3.1 Linear discriminant analysis

By applying a set of independent variables (predictors of the statistical model), and through utilization of Fisher's linear discriminant analysis (LDA), objects and individuals can be both classified and predicted into classes which are mutually exclusive and exhaustive [Huberty 1994]. Due to the fact that each class is comprised of objects which are similar and are further accompanied by errors when observed or measured, therefore, achieving a linear transformation of the independent variables of the model which maximize the ratio of between class scatter and within class scatter, is the main goal of linear discriminant analysis (LDA) [Laitinen 2007]. Maximizing this ratio is known as Fisher's criterion as well. In linear discriminant analysis (LDA), the classes are taken into consideration and the subspace in which the samples of the same class are the most compact will be found. In the meantime, the samples pertaining to other classes are as far as they could possibly be [Tran 2007].

Both linear discriminant analysis (LDA) and multiple regression method have the similarity in utilizing a linear function of independent variables of the model. However, in multiple regression method, the dependent variable, or the output of the model, needs to be a real number, whereas in linear discriminant analysis, the dependent variable should be of categorized nature. A further restriction when utilizing linear discriminant analysis is the fact that the independent variables of the model need to follow multivariate normal distribution [Tabachnick and Fidell 2001].

Furthermore, linear discriminant analysis (LDA) can be used for cases where the dependent variable is either dichotomous or can be assigned with multiple classes. In cases where multiple response categories are considered for the dependent variable of the model, the linear discriminant analysis is known as multiple discriminant analysis (MDA) [Huberty 1994].

In engineering field as well as researches pertaining to business filed, linear discriminant analysis has been used with various applications [Yang et al. 1999, Shan et al. 2002, Galletti et al. 2003, Tsai 2006, Tran 2007]. Additionally, maximizing the Fisher's criterion is also referred to as the calibration method used for linear discriminant analysis [Johnson and Wichern 2002].

2.9.3.2 Markov chain model

Markov chain model is based on the methodology proposed by Andrei Markov in 1906. Markov chain is indeed a discrete-time stochastic process [Winston 1994] and if a stochastic process has Markov property, it can be labeled as a Markov chain; furthermore, in order for a stochastic process to possess the Markov property, the conditional probability distribution pertaining to the future states of the variable of interest, independent from the past condition states, only depend upon the condition state of the present; in other words, having both the current and the past condition states of the variable of interest, only current condition state, without considering the past condition states, will be influential in determining the future condition states of the variable of interest [Ross 2000]. Considering a discrete parameter stochastic process (denoted by X_t) having a discrete state space, the abovementioned property can be illustrated as follows.

$$P(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = i_{t+1} | X_t = i_t) \quad (2.3)$$

In the above equation, P denotes the conditional probability of the state of the process for future events with respect to the past and current states, and i_t denotes the state of the process at time t. It should be noted that $t \geq 0$ and $i_0, i_1, i_2, \dots i_{t-1}, i_t, \dots$ represent the different condition states associated with the system characteristic [Baik 2003].

With regards to development of deterioration models for infrastructures, and that the present state of the variable of interest is in state i, the probability that the future state of the variable transitions from condition state i to condition state j, is denoted by p_{ij} and based on the aforementioned assumption of Markov chain, is independent of the past condition states and can be computed as follows.

An assumption pertaining to the Markov chain is the stationary assumption; based on the stationary assumption, the probability does not vary with time; in other words, the probability of the condition state to move to state j at time t+1 from state i at time t, will not be dependent upon time t. The stationary assumption is demonstrated in the below equation [Baik 2003].

$$P(X_{t+1} = j | X_t = i) = p_{ij} \quad (2.4)$$

In this equation, p_{ij} is known as the transition probability from state i to state j.

Considering the variable of interest has k possible condition states associated with it, a matrix with a dimension of $k \times k$ will be utilized to construct the transition probability matrix (P), which each of its individual elements contains the transition probabilities. The following equation demonstrates a typical transition matrix attributed to Markov chain models.

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \dots & P_{1K} \\ P_{21} & P_{22} & P_{23} & \dots & P_{2K} \\ P_{31} & P_{32} & P_{33} & \dots & P_{3K} \\ \dots & \dots & \dots & \dots & \dots \\ P_{K1} & P_{K2} & P_{K3} & \dots & P_{KK} \end{bmatrix} \qquad (2.5)$$

In a transition probability matrix, also called transition matrix, the following relationship should be valid.

$$\sum_{j=1}^{k} P_{ij} = 1 \qquad (2.6)$$

In which P_{ij} is the transition probability from condition state i to condition state j, and i can take on values of 1, 2, 3, …, and k.

In order to determine the probability that the condition state i is transitioning to the condition state j, after n transition states (n periods), Chapman-Kolmogorov relationship can be utilized. Therefore, the n-step transition probability matrix $P^{(n)}$, containing the n-step transition probabilities pertaining to each condition state, $P_{ij}^{(n)}$, can be computed by multiplying the one-step transition matrix n times; in other words, the n-step transition probability matrix can be computed by taking the one one-step transition matrix to the power n, as demonstrated below [Baik 2003, Park 2009].

$$P^{(n)} = P^n \qquad (2.7)$$

Furthermore, the one-step transition probability is shown by $P_{ij}^{(n=1)} = P_{ij}$.

Assuming the initial state vector is the probability of the Markov chain to be in state i at time $t=0$, in order to compute the probability of the chain to be in state j after n transition periods, the following equation can be used.

$$Q^{(n)} = Q^{(0)}.P^{(n)} \tag{2.8}$$

In the above equation, $Q^{(0)}$ denotes the probability of the Markov chain to be in state i at time $t=0$ and is equal to:

$$Q^{(0)} = [q_1 \quad q_2 \quad \cdots \quad q_{k-1} \quad q_k] \tag{2.9}$$

In which q_1, q_2, …, q_{k-1}, q_k denote the probabilities of system characteristic to be in condition states 1, 2, …, k-1, and k, respectively.

Additionally, $Q^{(n)}$ is the state vector after n transitions, and contains the probability of the chain to be in state j after n transition periods [Winston 1994].

In order to estimate the transition probabilities in a Markov chain methods such as approaches based non-linear optimization and ordered probit model can be utilized [Baik 2003].

2.9.3.3 Estimation of transition probabilities in a Markov chain

Estimating the transition probabilities in a Markov chain is a crucial task. Two different methodologies are stated herein:

- Approach based on non-linear optimization
- Approach based on ordered probit model

2.9.3.3.1 Approach based on non-linear optimization

The non-linear optimization-based approach is comprised of two major steps: 1) Regression analysis, 2) Non-linear optimization. One of the most popular methodologies for obtaining the values of transition probabilities is using the regression-based expected value method. In this approach, first the regression analysis is conducted and next the sum of absolute distances between the expected values of the model achieved through regression analysis and the Markov chain model, and by utilizing the non-linear optimization methodologies, each of the elements of the transition probabilities are estimated. This procedure is vastly used for predicting the transition probabilities for various infrastructures such as sewers, bridges, and pavements [Wirahadikusumah et al. 2001, Baik 2003, Jiang et al. 1988, Butt et al. 1987, Carnahan et al. 1987].

In some sources, a step by step procedure is used to for non-linear optimization and obtaining the transition probabilities in deterioration models [Park 2009]. In this step by step approach, first, by taking into account the various available properties of the pipes, such as length of pipes, size of pipes, installation depth, water level, condition of soil near pipe, and pipe material, the existing pipes are categorized in different groups. Once the grouping of pipes is completed, then for each group, the relationship between the age of the pipe and the condition rating of the pipe is estimated through regression analysis.

An assumption pertaining to non-linear optimization is that in order to capture the trend of increasing rate of deterioration for later years, a "zoning" concept is thus used. The environment surrounding the infrastructure which contributes to the deterioration of it, and therefore, its impact on the deterioration of the infrastructure might vary with time. This indicates that the transition of the condition states of a system, such as pavements, cannot be estimated with constant components in the transition probability matrix [Butt et al. 1987]. In order to address this issue, the "zoning" concept can be used. Through this concept, the lifetime of an infrastructure is assumed to be comprised of multiple zones, in which each of the zones are indeed periods of times for which the elements of the transition probability matrix are considered to be constant. Therefore, the resulting Markov chain will be homogenous.

Therefore, the transition probability matrix is assumed to be constant in each zoning period and the amount of time considered as a zone is dependent upon inspection frequency and intervals and engineering judgment [Park 2009, Butt et al. 1987, Baik 2003]. For instance, a typical time period of six years can be assumed for pavements, as considered in some sources [Butt et al. 1987] or for bridges [Jiang and Sinha 1989]. However, some sources have used larger time periods, i.e. 25 years, as the period for large combined sewers [Wirahadikusumah et al. 2001].

Finally, by conducting a non-linear optimization process, the transition probabilities corresponding to each condition state of the pipe are determined. This process is illustrated below [Baik 2003].

$$Minimize \sum_{t=t_{in}}^{t_f} \sum_{n=1}^{N} |Y(t) - E(n,P)| \qquad (2.10)$$

In the above equation, the parameters are as follows.

- i and $j = 1, 2, 3, \ldots, k$ (k being the number of condition states)
- $0 \leq P_{ij} \leq 1$
- N denotes the total number of transition periods pertaining to each zone,
- t is the age of the pipe,
- t_{in} denotes the initial age pertaining to each zone
- t_f denotes the final age pertaining to each zone
- n is the number of transition periods (also known as stages)
- $Y(t)$ is the mean estimated condition obtained from the regression analysis,
- $E(n,P)$ denotes the expected value of condition state of pipe for n transitions and is based upon the Markov chain approach

For instance, by considering the condition rating from 1 to 5, where condition rating 1 illustrates that the pipe is in perfect condition and condition rating 5 corresponds to the worst condition of the pipe, the expected value of condition state of pipe will be as shown below.

$$E(n,P) = Q^{(n)}.C^T = Q^{(0)}.P^{(n)}.C^T \qquad (2.11)$$

In the above equation, the parameters are as follows:

$Q^{(0)}$ denotes the initial vector pertaining to condition state of the infrastructure

$Q^{(n)}$ denotes the condition state vector of the infrastructure at the n^{th} stage

$P^{(n)}$ denotes the transition probability matrix after n transitions

C^T is the column vector pertaining to the condition rating system of the pipe.

It should be noted that the transition probabilities are the unknown parameters and the task is mainly to identify these knowns by using non-linear optimization; the obtained transition probabilities will pertain to each of the different zones for which the probabilities of transition from one condition state to another are supposed to remain constant. Furthermore, the above equations can be rewritten as follows.

$$E(n,P) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} . P^{(n)} . \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} \tag{2.12}$$

In the above relationship:

$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ demonstrates the initial condition state of the pipe, $Q^{(0)}$, which is assumed to be the perfect condition where the condition of the pipe is in condition rating 1.

$P^{(n)}$ denotes the transition probability matrix after n transitions.

$C^T = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$ denotes the column vector pertaining to the condition rating system of the pipe.

By utilizing this methodology, a transition probability matrix will be calculated as illustrated below.

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} \\ P_{21} & P_{22} & P_{23} & P_{24} & P_{25} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} \end{bmatrix} \tag{2.13}$$

However, by using several assumptions, the transition probability matrix will become simpler; first, it is assumed that no rehabilitation or maintenance is conducted and therefore, it is realized that the elements of the matrix below the diagonal elements do not exist, thus resulting in a transition probability matrix of the form below.

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} \\ 0 & P_{22} & P_{23} & P_{24} & P_{25} \\ 0 & 0 & P_{33} & P_{34} & P_{35} \\ 0 & 0 & 0 & P_{44} & P_{45} \\ 0 & 0 & 0 & 0 & P_{55} \end{bmatrix} \tag{2.14}$$

Additionally, it can be assumed that the condition rating of a particular pipe will not be reduced more than one condition level in each year. Therefore, considering these assumptions, the transition probability matrix will be simplified as illustrated below.

$$P = \begin{bmatrix} P_{11} & P_{12} & 0 & 0 & 0 \\ 0 & P_{22} & P_{23} & 0 & 0 \\ 0 & 0 & P_{33} & P_{34} & 0 \\ 0 & 0 & 0 & P_{44} & P_{45} \\ 0 & 0 & 0 & 0 & P_{55} \end{bmatrix} \tag{2.15}$$

Moreover, considering the restriction stated earlier as (the value of k is equal to 5, demonstrating the 5 possible condition levels of the pipe):

$$\sum_{j=1}^{5} P_{ij} = 1 \tag{2.16}$$

In which i can be assigned values of 1, 2, 3, 4, and 5. Therefore, the transition probability matrix will further simplify to the following format.

$$P = \begin{bmatrix} P_{11} & 1-P_{11} & 0 & 0 & 0 \\ 0 & P_{22} & 1-P_{22} & 0 & 0 \\ 0 & 0 & P_{33} & 1-P_{33} & 0 \\ 0 & 0 & 0 & P_{44} & 1-P_{44} \\ 0 & 0 & 0 & 0 & P_{55} \end{bmatrix} \tag{2.17}$$

Finally, considering that for the last row of the transition probability matrix only one element is remaining and that the following equation should be satisfied:

$$\sum_{j=1}^{5} P_{5j} = 1 \tag{2.18}$$

Therefore, it is realized that $P_{55} = 1$; in other words, considering that no maintenance is taking into account in this approach, therefore, once the pipe is in the condition rating 5, which is the worst condition, with a probability of 1, it will stay in that condition. The transition probability matrix is thus achieved to be as follows.

$$P = \begin{bmatrix} P_{11} & 1-P_{11} & 0 & 0 & 0 \\ 0 & P_{22} & 1-P_{22} & 0 & 0 \\ 0 & 0 & P_{33} & 1-P_{33} & 0 \\ 0 & 0 & 0 & P_{44} & 1-P_{44} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.19}$$

Now assuming that the transition period considered for each zone is 5 years, i.e. $N = 5$, therefore, considering the first zone, the values corresponding to t_{in} and t_f , i.e. the initial and final timeline of the first zone, will be equal to 1 and 5, respectively. For the second zone however, $t_{in} = 6$ and $t_f = 10$; and this process continues for the other zones as well. Additionally, when computing the mean value of $Y(t)$, in the case of pipelines for instance, if the condition rating obtained based on the regression is greater than the maximum condition rating, in this case 5, the maximum condition rating will replace the value calculated based on the regression analysis for those time periods.

After computation of the transition probability matrix for each zone, the condition state vector pertaining to each transition can be easily obtained. For illustration purposes, for the first zone with a period of 5 years, the condition states for each transition can be obtained as demonstrated below.

$$Transtion\ 1: Q^{(1)} = Q^{(0)}.P_1 \tag{2.20}$$

$$Transtion\ 2: Q^{(2)} = Q^{(1)}.P_1 = Q^{(0)}.P_1^2 \tag{2.21}$$

$$Transtion\ 3: Q^{(3)} = Q^{(2)}.P_1 = Q^{(0)}.P_1^3 \tag{2.22}$$

$$Transtion\ 4: Q^{(4)} = Q^{(3)}.P_1 = Q^{(0)}.P_1^4 \tag{2.23}$$

$$Transtion\ 5: Q^{(5)} = Q^{(4)}.P_1 = Q^{(0)}.P_1^5 \tag{2.24}$$

In the above equation, P_1 denotes the transition probability matrix for the first zone.

Once the state vectors for each of the transitions in the first zone are obtained, by considering the last state vector pertaining to the first zone as the initial state vector for the subsequent zone, i.e. zone 2, the state vectors for each of the transitions in the second zone, assuming the transition probability matrix for the second zone is denoted by P_2, can be obtained as illustrated below.

$$Transtion\ 6\ (transiton\ 1\ in\ zone\ 2): Q^{(6)} = Q^{(5)}.P_2 = Q^{(0)}.P_1^5.P_2 \tag{2.25}$$

$$Transtion\ 7\ (transiton\ 2\ in\ zone\ 2): Q^{(7)} = Q^{(6)}.P_2 = Q^{(0)}.P_1^5.P_2^2 \tag{2.26}$$

$$Transtion\ 8\ (transiton\ 3\ in\ zone\ 2): Q^{(8)} = Q^{(7)}.P_2 = Q^{(0)}.P_1^5.P_2^3 \tag{2.27}$$

$$Transtion\ 9\ (transiton\ 4\ in\ zone\ 2): Q^{(9)} = Q^{(8)}.P_2 = Q^{(0)}.P_1^5.P_2^4 \tag{2.28}$$

$$Transtion\ 10\ (transiton\ 5\ in\ zone\ 2): Q^{(10)} = Q^{(9)}.P_2 = Q^{(0)}.P_1^5.P_2^5 \tag{2.29}$$

Following this process, the state vectors corresponding to all zones of the infrastructure will be calculated; once the state vectors are obtained, then by using the rating column vector, i.e. C^T, the values associated with the condition ratings of each transition for all zones will be obtained based upon the Markov chain model. Therefore, the deterioration model, based on the condition rating, with respect to age of the infrastructure, in this case sewer pipes, can be procured.

2.9.3.3.2 Approach based on ordered probit model

When utilizing ordinary regression, the ordinal scale of the dependent variable, in this case the condition rating system, will not be fully comprehended by the model. This is due to the fact that when using ordinary regression, the actual variation between different levels of condition rating will be assumed to be based on the difference between each level. Therefore, the variation of the condition of an infrastructure will be assumed to be the same when condition rating changes from 1 to 2, when compared to change of condition rating from 3 to 4 [Greene 2003].

The condition rating system pertaining to sewer pipelines is a discrete and ordinal variable. Approximately since the 1970s, two substantial probability methods that are used to model ordinal discrete dependent variables, are the ordered probit model and the ordered logit model [Washington et al. 2003]. The main difference between the ordered probit model and the ordered logit model is the probability distribution that is assumed for the disturbance terms of each of these models. In the ordered logit model, the probability distribution for the disturbance term is considered to be normal, whereas it is assumed that in the ordered probit model, the disturbance term follows a logistic probability distribution [Baik 2003].

Some examples of the application of the ordered probit model is its utilization for bridge decks [Madanat et al. 1995] as well as bridge expansion joints [Lee and Chang 2003]. Additionally, ordered probit model in conjunction with incremental models has also been utilized for wastewater infrastructure assets [Baik 2003]. Below is a detailed description of the incremental model as well as the ordered probit model.

2.9.3.3.3 Incremental modeling for probability transitions

Incremental model has been developed in order for the probabilities procured by the ordered probit model to be used for obtaining a deterioration model based on a Markov chain [Madanat et al. 1995]. The increments are indeed the variations in the condition ratings within a transition period and are utilized as the discrete outcome variables in the ordered probit model. Each of the probabilities associated with discrete outcome variables (which are in fact the increments) are then used as the transition probabilities in the Markov chain approach. For instance, within a transition period, assuming that the condition of an infrastructure, following a discrete condition rating system, transitions from condition state i to condition state j, then the increment associated with this condition change is equal to $j - i$ [Baik 2003].

Therefore, for all condition states, the probabilities pertaining to the corresponding increments are computed. These probabilities indeed correspond to each row in the transition probability matrix. Following this procedure, for each transition, a distinct transition probability matrix will be obtained; thus, unlike the approach based on the non-linear optimization method, in which the elements of the transition probability matrix remained constant throughout each zone resulting in a stationary transition probability matrix, in the case of incremental modeling, the transition probability matrix will not be stationary as it does not remain constant for different transitions [Baik 2003].

2.9.3.3.4 Ordered probit model

In the ordered probit model, in order to achieve ranking of the discrete data, the unobserved variable (the latent variable), denoted by z_{im}, is utilized as the basis of the ranking. For instance, with regards to sewer pipes, the actual degradation of the sewer system is considered to be the latent variable; moreover, it is considered to continuous and in the range of (0, +∞). Considering a specific sewer segment is denoted by m and the condition associated with it is denoted by i, therefore, the latent variable (z_{im}) will be as demonstrated below.

$$z_{im} = \beta_i X_m + \epsilon_{im} \tag{2.30}$$

In the above equation, the parameters are as follows.

- β_i is the estimated vector parameters for condition state i
- X_m denotes the vector of variables for segment m dictating the discrete ordering
- ϵ_{im} denotes the random disturbance term

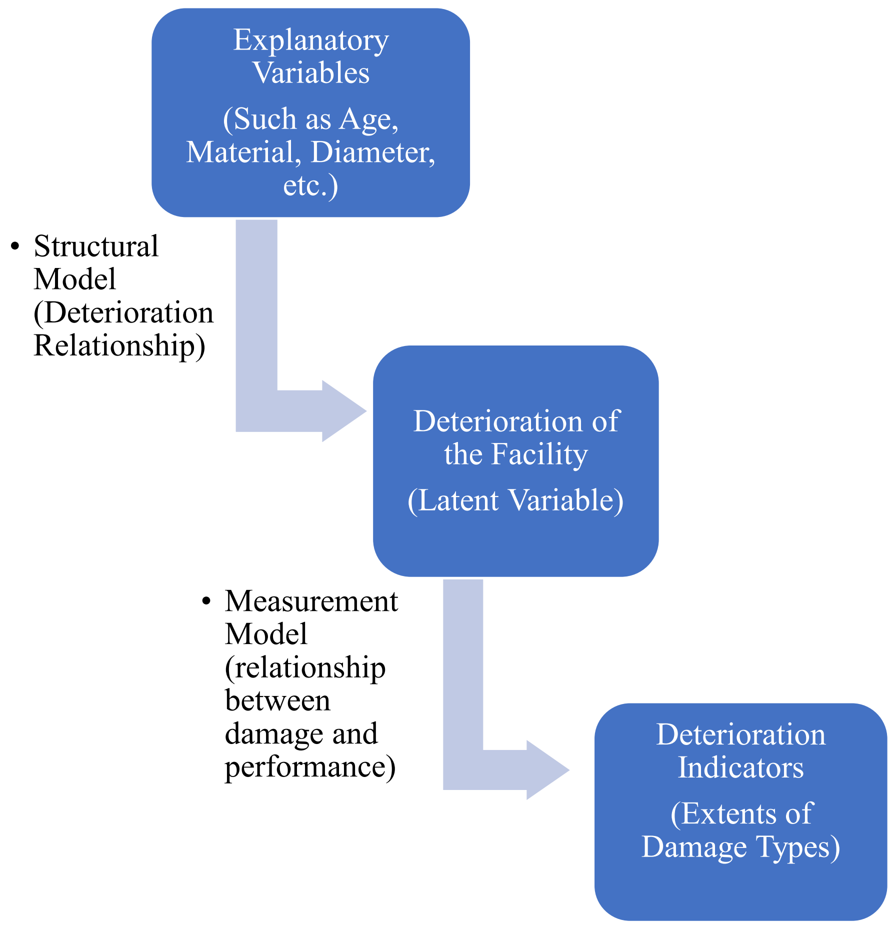

Figure 2.3: Mapping and procedure from latent variables to the indicator variable [Adapted from Ben-Akiva and Ramaswamy 1993]

In order to identify the relationship between the latent variable and the indicator variable, measurement equations are used which can map z_{im} to y_{im}. In other words, the continuous latent variable (i.e. deterioration) is mapped to the discrete indicator variable (i.e. condition increments). This procedure is also portrayed in the above figure.

Additionally, the following equation illustrates the relationship between the indicator variable and the latent variable. Based on this equation, the variation in the condition rating (i.e. increment) is y_{im}, if the latent deterioration, i.e. z_{im}, is between the two defined thresholds [Baik 2003].

$$y_{im} = j - i; when\ \mu_{i(j-i)} \leq z_{im} \leq \mu_{i(j-i+1)}; for\ j - i = 0, \dots, G - 1 \quad (2.31)$$

In the above equation, the parameters are as follows.

- $j - i$ denotes the variation in the condition state of segment m for one transition
- μ denotes the thresholds as follows: $\mu_{i0} = 0$ and $\mu_{i(j-i+1)} = \infty$
- G denotes the greatest value of condition rating

For instance, for sewer pipes, considering that the condition ratings vary from 1 through 5, the increment data will be as follows.

$$y_{im} = 0;\ \ when\ z_{im} \leq \mu_{i1}; \quad (2.32)$$

$$y_{im} = 1;\ \ when\ \mu_{i1} \leq z_{im} \leq \mu_{i2}; \quad (2.33)$$

$$y_{im} = 2;\ \ when\ \mu_{i2} \leq z_{im} \leq \mu_{i3}; \quad (2.34)$$

$$y_{im} = 3;\ \ when\ \mu_{i3} \leq z_{im} \leq \mu_{i4}; \quad (2.35)$$

$$y_{im} = 4;\ \ when\ \mu_{i4} \leq z_{im}; \quad (2.36)$$

Therefore, considering the normal distribution for the disturbance term in the ordered probit model, as stated earlier, the ordered probit model is established as follows.

$$y_{im} = j - i; when\ \mu_{i(j-i)} - \beta_i X_m \leq \epsilon_{im} \leq \mu_{i(j-i+1)} - \beta_i X_m; for\ j - i = 0, \dots, G - 1 \quad (2.37)$$

As explained earlier the disturbance term in the ordered probit model, ϵ_{im}, is assumed to be normally distributed; furthermore, the mean and variance pertaining to this disturbance term are considered to be 0 and 1, respectively. Hence, the transition probability from condition *i* to condition *j*, which in fact is the probability that the variation in the condition change is equal to $j - i$ (i.e. $y_{im} = j - i$), can be written by using cumulative normal distribution $\Phi(.)$; this process is illustrated below.

$$P(y_{im} = j - i) = \Phi(\mu_{i(j-i+1)} - \beta_i X_m) - \Phi(\mu_{i(j-i)} - \beta_i X_m); for\; j - i = 0, \dots, G - 1 \quad (2.38)$$

For the population of M_i, which denotes the overall number of sewer segments being in condition *i*, the likelihood function pertaining to the maximum likelihood estimation (MLE), is shown below.

$$L(y \mid \beta, \mu) = \prod_{m=1}^{M_i} \prod_{j-i=0}^{G-1} \left[\Phi\left(\mu_{i(j-i+1)} - \beta_i X_m\right) - \Phi\left(\mu_{i(j-i)} - \beta_i X_m\right)\right]^{\delta_{im}} \quad (2.39)$$

In the above equation, if the observed increment of condition rating for m^{th} sewer segment is equal to $j - i$, $\delta_{im} = 1$; else, $\delta_{im} = 0$.

Therefore, in an ordered probit model, the log likelihood function will be as follows.

$$Log\ likelihood\ function = LL = \sum_{m=1}^{M_i} \sum_{j-i=0}^{G-1} \delta_{im} ln\left[\Phi\left(\mu_{i(j-i+1)} - \beta_i X_m\right) - \Phi\left(\mu_{i(j-i)} - \beta_i X_m\right)\right] \quad (2.40)$$

Finally, the parameters of the ordered probit model, i.e. β and μ, can be obtained through maximizing the log likelihood function.

2.9.3.3.5 Transition matrix via ordered probit model

When using ordered probit model, in order to obtain the transition probability matrix of the Markov chain, firstly, for each condition state of the dependent variable, the probabilities associated with the increments in the condition changes are to be computed. As an example, in the case of sewer pipes, assuming that no maintenance or rehabilitation of the sewer pipes are conducted and thus the condition of the sewer pipe does not undergo any improvement, and considering the condition ratings 1 through 5, there will be five different increments associated with condition state 1; these increments are 0, 1, 2, 3, and 4. Assuming there are k different condition states pertaining to the dependent variable, in order to build the transition matrix in the Markov chain approach, *k-1* incremental deterioration models will be required (This is due to the fact that the last row of the matrix is considered as an absorbing state) [Baik 2003].

Once the unknown parameters of the ordered probit model are obtained based on the maximum log likelihood function, therefore, for each sewer segment, the transition probabilities can be calculated as demonstrated by the following equations.

$$\hat{P}(y_{im} = 0 \mid X_m, i) = \Phi(\hat{\mu}_{i1} - \hat{\beta}_i X_m) \tag{2.41}$$

$$\hat{P}(y_{im} = 1 \mid X_m, i) = \Phi(\hat{\mu}_{i2} - \hat{\beta}_i X_m) - \Phi(\hat{\mu}_{i1} - \hat{\beta}_i X_m) \tag{2.42}$$

$$\hat{P}(y_{im} = 2 \mid X_m, i) = \Phi(\hat{\mu}_{i3} - \hat{\beta}_i X_m) - \Phi(\hat{\mu}_{i2} - \hat{\beta}_i X_m) \tag{2.43}$$

$$\dots$$

$$\hat{P}(y_{im} = G - 1 \mid X_m, i) = 1 - \Phi(\hat{\mu}_{i(G-1)} - \hat{\beta}_i X_m) \tag{2.44}$$

In the above equations, $\hat{P}(y_{im} \mid X_m, i)$ denotes the transition probability from condition state i to condition state j, for the segment with attribute vector of X_m.

In some cases, instead of considering the individual infrastructure facilities, it might be more desirable to conduct the decision making process on the basis of results pertaining to a set or the entirety of the infrastructure framework. In such cases, in order to compute the transition probability matrix pertaining to the whole infrastructure of a set of individual facilities, the transition probability matrix of each individual facilities will be taken into account. For instance, five methods can be utilized for the purpose of aggregation of the transition probability matrices of individual facilities; these methods are namely as follows: Classification process, statistical differentials process, explicit integration process, sample enumeration process, and average individual process [Ben-Akiva and Lerman 1985]. For instance, considering the average individual process, the transition probability matrix of a set of infrastructure facilities can be obtained by averaging the transition probability matrices pertaining to the individual facilities involved within the set [Baik 2003].

Finally, once by using the ordered probit model, the transition probability matrices are obtained for each transition periods (as explained earlier, the transition probability matrices will be different for each transition period due to its non-stationary feature in the ordered probit approach), then condition vector for each stage can be computed thus resulting in the deterioration curve.

2.9.3.4 Limitations of statistical modeling

Even though the probabilistic nature of the deterioration can be taken into consideration through a statistical modeling, one of the main disadvantages of the statistical modeling is the fact that these models are sensitive to noisy data. Additionally, as the exact cause and effect in the underlying procedure is not identified, therefore, removing the aforementioned noisy data in statistical models is not conveniently possible [Terano et al. 1991, Leung and Tran 2000, Dasu and Johnson 2003, Tran 2007].

2.9.4 Artificial intelligence approach

Deterioration models based upon artificial intelligence approach, fall into the data-driven category. In other words, the construction of the model is based upon the available data (procured from inspections) as opposed to model-driven approaches (i.e. deterministic and statistical models) [Dasu and Johnson, 2003]. This is due to the fact that artificial intelligence approach is essentially designed to operate based on learning through improving and generalization, similar to learning pertaining to human brain [Taylor, 1996; Soulie and Gallinari, 1998; Taylor, 1993].

Due to the inherent learning and improving feature of artificial intelligence technique, numerous applications of this approach within various engineering areas have been observed, wherein the patterns between the inputs and the outputs (in other words, independent variables and the outcomes) are identified through this characteristic of artificial intelligence and eventually the classifications are allocated based on the observed patterns [Seo et al. 2004; Wilmot and Mei, 2005; Moslehi and Shehab-Eldeen, 2000; Singh and Tiong, 2005].

As an example, case-based reasoning (CBR) is an artificial intelligence method which was used by Morcous et al. (2002a) in order to model the deterioration of infrastructures. Case-based reasoning is essentially a modeling technique based on the experiences gained for past cases, in other words it operates based on the process through which human brains makes decisions based on previous experiences [Riesbeck and Schank, 1989; Aamondt and Plaza, 1994]. However, the shortcoming of case-based reasoning is the fact that experienced case library is required in order to obtain the results.

Chapter 3 : Modeling Methodology

3.1 Deterioration models

In the study at hand, the following methodologies are utilized to obtain deterioration models associated with the available data set. The deterioration models obtained using these approaches are utilized to estimate the probabilities of failure with respect to age of assets as well as their associated service lives. Considering the criteria for defining service life of assets can vary based on the decision-makers' priorities and failure criteria, the uncertainties and assumptions associated with these approaches, as well as the suitability of the available data, the deterioration curves as well as the values of service life estimated based on these methods can be subjected to variations and uncertainties for different sewer pipes. Therefore, the modeling and results presented herein are solely for the purpose of illustrating the application of each of these models for the available data set and thus the interpretation of the probabilities of failure and the service lives estimated based on these deterioration models need to be proceeded with caution and are valid as long as the assumptions as well as the criteria for defining service life are taken into consideration.

3.2 Logistic Regression

Logistic regression is a statistical instrument which assists in finding the probabilities of the outcome of interest with the potential of handling both categorial and numerical predictors within the model. Furthermore, within the logistic regression, three major categories are as follows:

Binary logistic regression (dichotomous dependent variable)

Ordinal logistic regression (proportional odds model)

Multinomial logistic regression

With regards to the dependent variable of the model, a *logit* function is applied. The utilization of the *logit* function can be substituted by various appropriate link functions should there be a better presentation of the model by using them. Detailed information regarding each of the three aforementioned logistic regressions are presented below.

3.2.1 Binary Logistic Regression

In order to use the binary logistic regression, the dependent variable is required to be dichotomous or binary; a dichotomous variable can only be assigned two different values. For instance, these assignments could be failure or success (1 and 0). As an example, when applying binary logistic regression model, herein it will be assumed that the structural condition of the pipelines can be either at an acceptable or a failed condition. Additionally, when modeling the deterioration of the pipelines with binary logistic regression method, the dependent variables, also known as predictors of the model, can be assigned either as categorial or numerical variables; thus, allowing the model to be holistic in considering the parameters influencing the deterioration of the pipelines.

In order to use the categorial variables in the model, the corresponding independent variable will be dummy coded into the model. For instance, if a categorial variable consists of *m* potential values, then *m-1* variables would be required to dummy code the model. However, numerical independent variables can be simply used by only one variable representation in the model.

In Logistic regression, the logit function of the dependent variable is associated with the contributing predictors in the model. Below, the equation pertaining to the binary logistic regression is presented [Menard, 2002].

$$Logit\ (Y) = ln\ (\frac{P(Y=1|X_1,X_2,\dots,X_n)}{1-P(Y=1|X_1,X_2,\dots,X_n)})$$

$$= a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_n X_n \qquad (3.1)$$

In which Y is the dependent variable (assumed to be dichotomous), X_1 ... X_n are the independent variables of the model (predictors), *a* is the intercept of the model, and b_1 ... b_n are the coefficients of the binary logistic regression model. Moreover, the *logit* function represents the *log* of the odds ratio.

In order to calculate the probability that the dichotomous dependent variable ends up with the value 1, (for instance corresponding to failure) equation below can be utilized.

$$P(Y = 1|X_1, X_2, \dots, X_n) = \frac{exp(a + b_1\, X_1 + b_2\, X_2 + b_3\, X_3 + \dots + bn\, Xn)}{1 + exp(a + b_1\, X_1 + b_2\, X_2 + b_3\, X_3 + \dots + bn\, Xn)} \tag{3.2}$$

Alternatively, the following equation can be used as well.

$$P(Y = 1|X_1, X_2, \dots, X_n) = \frac{1}{1 + exp(-a - b_1\, X_1 - b_2\, X_2 - b_3\, X_3 - \dots - bn\, Xn)} \tag{3.3}$$

Assuming, for simplicity, that the value of $a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_n X_n$ is equal to *G*, then the aforementioned probability can be written as follows.

$$P(Y = 1|X_1, X_2, \dots, X_n) = \frac{1}{1 + exp(-G)} \tag{3.4}$$

Below is a representation of the logit function for various values of *G*.

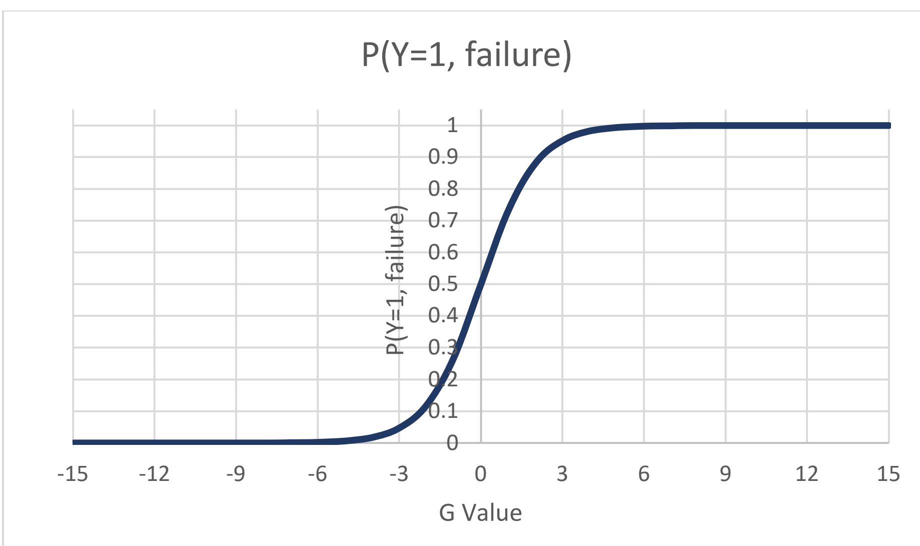

Figure 3.1: Variations in the probability of failure with respect to *G* values

Based on the above figure, it is observed that the probability that the dependent variable equals to 1 is subjected to significant variations for values of *G* around 0; however, for larger and smaller *G* values the probability of the dependent variable to be equal to 1 will only undergo small changes [Salman, 2010].

In order to compute the values for the parameters involved in the binary logistic regression model, Maximum Likelihood Estimation (MLE) will be applied. The parameters of interest in the binary logistic regression models are the intercept and the coefficients pertaining to the independent variables. In case for a particular values of independent variables, more than on observation is made in the resulting outcome, then the likelihood function (L) will be as follows [Agresti 2002]:

$$\text{Likelihood function (L)} = \prod_{i=1}^{N}\left[\frac{\exp\left(a+\sum_{j=1}^{n} b_i X_{ij}\right)}{1+\exp\left(a+\sum_{j=1}^{n} b_i X_{ij}\right)}\right]^{y_i}\left[\frac{1}{1+\mathrm{e}\quad\left(a+\sum_{j=1}^{n} b_i X_{ij}\right)}\right]^{m_i-y_i} \qquad (3.5)$$

In which, N is the total number of observations, m_i denotes the total number of observations pertaining to the fixed set of independent variables, n is the total number of independent variables in the model, y_i is the number of successes for the observation X_i, X_{ij} denotes the j-th value of the independent variable for observation X_i, *a* is the intercept of the model, and b_1 ... b_n are the coefficients of the binary logistic regression model.

3.2.1.1 Determining the significance

Significance can be determined for both the model itself, and the coefficients of the independent variables involved in the binary logistic regression model. Below the details pertaining to determination of each of these significances are stated.

Significance of the model

In a binary logistic regression model, in order to determine the significance of the model, a common methodology is by appraisal through the log likelihood values of the full model in comparison to the model with only the intercept parameter. Thereafter, for obtaining the statistical significance of the model and by multiplying the values of log likelihood by -2, a chi-square distribution is thus obtained. The degrees of freedom for this chi-square distribution is equal to the number of additional terms in the logistic model. In case one of the models is embedded within another model, the aforementioned strategy may be implemented [McCullagh and Nelder 1989].

To compute the difference between the log likelihood values of the base model, the equation presented below can be used [Menard 2002].

$$-2\ Log\ Likelihood\ (Base\ Model) = -2[n_0.Ln(p_0) + n_1.Ln(p_1)] \qquad (3.6)$$

In the above equation, n_0 denotes the number of occurrences where Y is equal to 0, p_0 is the proportion of the occurrences where Y is equal to 0, n_1 denotes the number of occurrences where Y is equal to 1, and p_1 is the proportion of the occurrences where Y is equal to 1.

In the methodology stated above, the null hypothesis is implemented for the binary logistic regression model by considering that the coefficients of the independent variables of the model are equal to zero, i.e.:

$$b_1, b_2, b_3, \dots , b_n = 0 \qquad (3.7)$$

Therefore, the test statistics is obtained by the following equation.

$$X^2 = -2\ Log\ Likelihood\ (Model) - 2\ Log\ Likelihood\ (Base\ Model) \qquad (3.8)$$

And considering the fact that the number of the coefficients of independent variables in the base model is zero, hence, the number of coefficients of independent variables of the model determines the degree of freedom of the model [Salman 2010].

Significance of the coefficients

Once the binary logistic regression model is realized to be a significant model, this subsequently implies that the null hypothesis has been rejected. In other words, this means that the assumption that the coefficients of the independent variables of the models are zero (i.e. $b_1, b_2, b_3, \dots , b_n = 0$) does not hold true, and there is at least one coefficient of the independent variables which is significant and not equal to zero. In general, two different methodology can be used to determine which of the coefficients of the independent variables are indeed significant.

Log Likelihood Method

The first methodology is similar to the one used for determining whether or not the binary logistic regression model is significant. In this method, the model is developed once by eliminating the coefficient of interest and once by including that coefficient. Thereafter, the -2 log likelihoods of these two models (process is similar to the one explained for the significance of the model) are compared through a corresponding chi-square distribution. In effect the difference between the -2 log likelihoods determines the ratio of log likelihoods of the models. The number of eliminated coefficients from the model determines the degrees of freedom related to the chi-square distribution. For numerical independent variables, the degree of freedom of the distribution is equal to 1; if the eliminated coefficient pertains to a categorial variable with only one potential assignment, the degree of freedom of the distribution will still be equal to 1. However, in case the categorical independent variable can take on more than one values, for instance m different values can be assigned to the categorial variable, then the degree of freedom pertaining to the associated chi-square distribution will be m-1 [Salman 2010].

Wald Test

Another strategy to determine whether a specific coefficient of the binary logistic regression model is indeed significant or not, is through implementing the Wald statistic, as illustrated in the following equation.

$$W_i = \left[\frac{b_i}{Standard\ Error\ (b_i)}\right]^2 \tag{3.9}$$

When implementing the Wald statistic, the null hypothesis (H_0) states that the associated coefficient of interest is equal to zero (i.e. $b_i = 0$) and to reject the null hypothesis, the critical value computed based on the corresponding chi-square distribution (with only one degree of freedom) is checked with the value of W_i (Wald statistic) obtained from the above equation [Salman 2010].

In some cases, when the coefficient of the independent variable is a large value and the standard errors are large too, using the Wald statistic might result in failure to reject the null hypothesis, even though it is in fact not true. Therefore, in such cases, the log likelihood comparison yields more acceptable and accurate results [Menard 2002].

3.2.1.2 Influence of predictors in the model

The link function in the binary logistic regression model is a logit function; this function is in fact the natural logarithm of the odds ratio of the dichotomous dependent variable. As stated earlier, the relationship pertaining to the binary logistic regression model is presented as follows.

$$Logit\ (Y) = ln\ (\frac{P(Y=1|X_1,X_2,\dots,X_n)}{1-P(Y=1|X_1,X_2,\dots,X_n)})$$

$$= a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_n X_n \quad (3.10)$$

In this equation, it is observed that considering the dependent variable X_i, one unit increase in this dependent variable will lead to an increase in the amount of b_i in the *Logit (Y)*. It should be noted that with regards to categorial predictors, presence of the specific categorial independent variable with have the same effect. However, the resulting change in the odds ratio due to a unit increase in the variable X_i, is equal to exp(b_i). This observation in the odds ratio is also presented below.

$$\frac{\frac{P(Y=1|X_1,\ X_2,\dots,(X_i+1),\dots,X_n)}{1-P(Y=1|X_1,\ X_2,\dots,(X_i+1),\dots,X_n)}}{\frac{P(Y=1|X_1,\ X_2,\dots,X_i,\dots,X_n\ )}{1-P(Y=1|X_1,\ X_2,\dots,X_i,\dots,X_n\ )}} = \exp(b_i) \quad (3.11)$$

For instance, assuming that the independent variables X_1 and X_2 represent the age and material type of the pipeline, respectively; then an increase in one year of age of pipeline would correspond to exp(b_1) change in the odds ratio. Furthermore, considering that the material type of the pipeline is represented by a categorical variable, values 0 and 1 correspond to Polyvinyl Chloride (PVC) and Vitrified Clay Pipe (VCP). Therefore, the variation resulted in the odds ratio due to change of material of the pipeline would be equal to exp(b_2). It should be noted that as stated earlier, when the Y value is equal to 1 and 0, the pipeline is in the failure and acceptable conditions, respectively [Salman 2010].

3.2.1.3 Verification and Classification Table

Once the binary logistic regression model is developed, the next step is to identify the proportions of the results which are correctly predicted. This task is accomplished through the classification table. The classification table includes the predicted and the observed number of occurrences for each of the dichotomous cases. Once the probability of the dependent variable belonging to each of the two groups, i.e. 0 and 1, is obtained then by utilizing a cut-off value, the dependent variable will be assigned to the appropriate class. In general, the cut-off for dichotomous dependent variables is 0.5; if the resulting probability is greater than 0.5, then in the classification table, it will be assigned to the class corresponding to Y=1; otherwise, if the obtained probability does not exceed the cut-off value, class of Y=0 would be selected for the model outcome. Below is a general representation of the classification table in the binary logistic regression model.

Table 3.1: General form of classification table in binary logistic regression

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

Once the classification table is arranged, by summing up the elements of the diagonal part of the classification table, the percentage of correct predictions based on the binary logistic regression can be calculated as shown below [Salman 2010].

$$\text{Percentage correct predictions} = \frac{100(A_{11} + A_{22})}{A_{11} + A_{12} + A_{21} + A_{22}} \tag{3.12}$$

It is notable that the presented classification table is similar to the confusion matrix used in artificial intelligence methods.

3.2.1.4 Assumptions of the binary logistic regression

The assumptions needed for binary logistic regression model is less restrictive comparted to the assumptions required for the ordinary least squares multiple regression models. These assumptions (for binary logistic regression) are presented below [Meyers et al. 2006]:

- Perfect multicollinearity should not be present between the independent variables used in developing the model
- No specification error should be present in the model (No irrelevant predictor should be used in the model, and all the relevant independent variables should be present in developing the model)
- "A summative response scale, interval or ratio level of measurements" should be associated with the independent variables of the model. Dichotomous variables are also permitted. (For the cases where the categorial independent variables can take on more than two values, they can be dummy coded into dichotomous predictors; thus, satisfying the dichotomy rule. For instance, if a categorial predictor can take on *m* different values, by using *m-1* variables, it can be dummy coded into dichotomous variables.)

3.2.1.5 Application of binary logistic regression model

As it is observed from the characteristics of the binary logistic regression model, it can provide versatility in vast areas of research. Various fields of research take advantage of the binary logistic regression models in order to determine the relationship between a dichotomous dependent variable and known influential independent variables. Marketing, pharmacies, and medicine are some notable examples of areas where this type of modeling is used [Salman 2010]. With regards to the sewer system assets, several studies have also utilized the binary logistic regression models as well. [Ariaratnam et al. 2001, Davies et al. 2001b]

3.2.2 Multinomial Logistic Regression Model

In the binary logistic regression model, the dependent variable was supposed to be dichotomous; however, if the dependent variable has more than two potential values assigned to it, in this case the multinomial logistic regression can be used to model the dependent variable. Assuming that *k* possible values are associated with the dependent variable, by taking one of the possible results of outcome as the reference for calculating the odds ratio, *k-1* equations will be obtained in order to illustrate the logit functions pertaining to the dependent variable. Below is a general form of the multinomial logistic regression.

$$ln\ (\frac{P(Y=i|X_1,\ X_2,\dots,X_n)}{1-P(Y=k|X_1,\ X_2,\dots,X_n)}$$
$$= a_i + \ b_{i1}\,X_1\ +\ b_{i2}\,X_2\ +\ b_{i3}\,X_3\ +\ \dots\ +\ b_{in}\,X_n \qquad (3.13)$$

In this equation, $i=1,2,\dots, k-1$ pertains to each of the possible outcomes for the dependent variable, $X_1 \dots X_n$ are the independent variables of the model (predictors), a_i is the intercept of the model for the i^{th} outcome of the dependent variable, and $b_{i1} \dots b_{in}$ are the coefficients of the multinomial logistic regression model corresponding to the i^{th} outcome of the dependent variable.

Considering the possibility of *k* different categories for the dependent variable, the parameters to be found are *k-1* intercept terms, and *n.(k-1)* coefficients of the independent variables present in the model. These unknown parameters of the multinomial logistic regression model are computed all at the same time for *k-1* logit equations [Agresti 2002]. In order to obtain the probabilities of the dependent variable taking on each of the *k* possible categories, the equations presented below can be used.

$$\pi_i(X) = P(Y = i|X_1, X_2, \dots, X_n) = \frac{\exp(a_i + b_{i1}X_1 + b_{i2}X_2 + b_{i3}X_3 + \dots + b_{in}X_n)}{1 + \sum_{i=1}^{k-1}\exp(a_i + b_{i1}X_1 + b_{i2}X_2 + b_{i3}X_3 + \dots + b_{in}X_n)} \quad (3.14)$$

In which *i=1, 2, 3,…, n-1*

In order to obtain the probability of the dependent variable to be equal to the reference category for which *i=k* (k^{th} category), the following equation is used.

$$\pi_k(X) = P(Y = k|X_1, X_2, \dots, X_n) = \frac{1}{1 + \sum_{i=1}^{k-1}\exp(a_i + b_{i1}X_1 + b_{i2}X_2 + b_{i3}X_3 + \dots + b_{in}X_n)} \quad (3.15)$$

3.2.2.1 Identifying the unknowns

In order to identify the coefficients of independent variables as well as the intercept terms of the multinomial logistic regression model, the maximum likelihood estimation is used (identical to the procedure used for binary logistic regression models). As described earlier, if *k* different categories are possible for the dependent variable, *k-1* equations will be used to define the model, and when maximum likelihood estimation is used, all these equations will be dealt with simultaneously, and the results will be values of parameters maximizing the likelihood function. The j^{th} observation, having independent variables $X_j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jn})$, has the log likelihood function presented as below [Agresti 2002].

$$\log\left[\prod_{i=1}^{k}\pi_i(x_j)^{y_{ij}}\right]$$

$$= \sum_{i=1}^{k-1} y_{ij}\log\pi_i(x_j) + \left[1-\sum_{i=1}^{k-1} y_{ij}\right] log\left[1-\sum_{i=1}^{k-1}\pi_i(x_j)\right] \quad (3.16)$$

In which j=1, 2, 3, …, n. $\boldsymbol{y_j}$ is the multinomial trial for j[th] subject and is expressed as $\boldsymbol{y_j}$ = ($y_{j1}, y_{j2}, y_{j3}, \ldots y_{jk}$). For each element of $\boldsymbol{y_j}$ i.e. y_{ji}, if the outcome falls into the i^{th} category, then a value of 1 will be assigned to y_{ji}, else, it will take on a value of zero. The equation for log likelihood, in case all observations are accounted for, can be expressed as follows [Agresti 2002]:

$$\log likelihood = \log\prod_{j=1}^{n}\left[\prod_{i=1}^{k}\pi_i(x_j)^{y_{ij}}\right] \quad (3.17)$$

3.2.2.2 Identifying the model significance

In order to determine the model significance in a multinomial logistic regression model, similar procedure described in binary logistic regression method, will be used. By using likelihood ratio methodology, one can obtain the significance of a multinomial logistic regression model compared to a model which only comprises of intercept term of the regression. In order to obtain the log likelihood of the model which only contains the intercept term of the multinomial logistic regression (base model), the following equation can be used.

$$\log likelihood\ of\ the\ base\ model = (n_1.\ln(p_1) + n_2.\ln(p_2) + n_3.\ln(p_3) + \cdots + n_k.\ln(p_k) \quad (3.18)$$

In the above equation, values of $n_1, n_2, n_3, \ldots n_k$, denote the number of observations pertaining to different response categories. Moreover, values of $p_1, p_2, p_3, \ldots p_k$, denote the proportion of observations pertaining to each response category.

Considering the fact that the number of coefficients represented in the multinomial logistic regression model determines the degree of freedom associated with the critical chi-square distribution, the significance of the model can be appraised based on the difference between the -2 log likelihood of the base model with intercept term only, and the -2 log likelihood of the multinomial logistic regression model. This is due to the fact that the difference between -2 log likelihoods of the two models can be represented by a chi-square distribution [Salman 2010].

3.2.2.3 Identifying significant coefficients

In order to identify which of the coefficients in a multinomial logistic regression model are indeed significant, a similar procedures to that of explained for binomial logistic regression model can be used. In other words, both Wald statistic method and the likelihood ratio method can be applied. Albeit this similarity between binomial and multinomial logistic regressions, it should be borne in mind that due to the fact that there are more than one equations present in the multinomial logistic regression model (assuming that *k* possible values for the dependent variable, *k-1* equations will illustrate the logit functions pertaining to the dependent variable), therefore, unlike binomial logistic regression, in multinomial logistic regression, a particular variable may be significant in one of the equations but not in the other ones. Hence, utilizing the likelihood ratio might have advantage compared to using Wald statistic method. If the likelihood ratio is used to determine whether a particular variable of the model is significant or not, first the difference between -2 log likelihoods of the model comprising of that particular variable and another model which does not have that particular variable but contains all the other elements of the model is computed. This is in fact the likelihood ratio of the two models. Once the likelihood ratio is obtained, then the result will be compared to the associated critical chi-square distribution; thus, determining whether that variable is significant or not [Salman 2010].

3.2.2.4 Influence of predictors in multinomial logistic regression

As stated earlier, assuming that k possible values are associated with the dependent variable, considering one of the possible response categories as the reference category for calculating the odds ratio, *k-1* equations will be used in order to demonstrate the logit functions pertaining to the dependent variable. Unlike the binary logistic regression models, in multinomial logistic regressions, due to the fact that there are more than one equation describing the model, and considering that each of the logit equations will potentially possess different coefficients of regression, therefore, the influence of different independent variables will be different for each of the possible categories pertaining to the dependent variable. This event can be explained through the probabilities of each response category of the dependent variable.

For instance, as illustrated by the following equation, with regards to the probability of each response category, depending upon the coefficients of the regression and intercept pertaining to each category, the numerator of the following equation will vary; even thought the denominator is the same for all response categories of the dependent variable. Therefore, the influence of predictors should be examined separately for each response category.

$$\pi_i\,(X) = P(Y = i \mid X_1, X_2, \dots, X_n) \\ = \frac{\exp(a_i + b_{i1}\,X_1 \; + b_{i2}\,X_2 \; + b_{i3}\,X_3 + \; \dots \; + \; b_{in}\,X_n)}{1 + \sum_{i=1}^{k-1} \exp(a_i + b_{i1}\,X_1 \; + b_{i2}\,X_2 \; + b_{i3}\,X_3 + \; \dots \; + \; b_{in}\,X_n)} \qquad (3.19)$$

In which n is the number of influential predictors and *i=1, 2, 3,…, n-1.*

3.2.2.5 Verification and Classification Table

The classification table used for a multinomial logistic regression model is similar to the classification table used for binary logistic regression models; however, unlike the binary logistic regression model, for which only two possible values were possible and therefore the classification table had only 4 elements, in a multinomial regression model with k

possible response category for the dependent variable, the classification table will have k^2 elements ($k \times k$ table). Below is a general representation of the classification table used in multinomial logistic regressions.

Table 3.2: General form of classification table in multinomial logistic regression

	Predictions			
Observations	1	2	…	k
1	A_{11}	A_{12}	…	A_{1k}
2	A_{21}	A_{22}	…	A_{2k}
…	…	…	…	…
k	A_{k1}	A_{k2}	…	A_{kk}

In binary logistic regression models, by calculating the probability of the dependent variable belonging to each of the two possible categories, for instance 0 and 1, and by utilizing a cut-off value, the dependent variable was then assigned to the appropriate class. In general, the cut-off used for dichotomous dependent variables was 0.5; if the resulting probability is greater than 0.5, then in the classification table, it will be assigned to the class corresponding to Y=1; otherwise, if the obtained probability does not exceed the cut-off value, class of Y=0 would be selected for the model outcome. In a multinomial logistic regression however, due to greater number of response categories, the assignment of each observed probability of the dependent variable is conducted by taking the greatest value of probability calculated through the logit functions describing the model. Using this procedure for binary logistic regression is analogous to using a cut-off value of 0.5.

Once the classification table is arranged, considering the total number of observations is equal to T, by summing up the elements on the diagonal part of the classification table, the percentage of correct predictions can be computed through using the following equation [Salman 2010].

$$\text{Percentage correct predictions} = \frac{100(A_{11} + A_{22} + \cdots + A_{kk})}{\sum_{i=1}^{k}\sum_{j=1}^{k} A_{ij}} \tag{3.20}$$

$$T\ (Total\ number\ of\ observations) = \sum_{i=1}^{k}\sum_{j=1}^{k} A_{ij} \tag{3.21}$$

3.2.2.6 Assumptions of the regression modeling

In multinomial logistic regression models, similar assumptions as the ones stated for binomial logistic regressions are applied. Overall, multinomial logistic regression can be used to model dependent variables for which multiple (more than two) possible response categories are defined. However, it should be borne in mind that by modeling a dependent variable through multinomial logistic regression, there will be no consideration with regards to the ordinal levels present in the model. In other words, when models are created using multinomial logistic regression, it is assumed that there is no sequence to the corresponding response categories (no ordinal relationship is considered in the model). Therefore, in case there exists an ordinal relationship among response categories of the dependent variable, in order to reflect this relationship, ordinal regression can be used.

The main difference between multinomial logistic regression and ordinal regression is the fact that in multinomial logistic regression, for each of the multiple logit equations involved in defining the model, the coefficients of the independent variables as well as the intercept terms used for each individual equation, can take on different values. However, when using the ordinal regression model, there is an additional requirement which states that the coefficients of the independent variables should be the same for all of the individual logit equations defining the model, further restricting the modeling procedure. This restraint in ordinal regression models is also known as proportional odds assumption. Albeit this restraint in modeling, the intercept terms can be assigned different values in each individual equation. In case the proportional odds assumption cannot be assured in a model, then multinomial logistic regression is used to provide flexibility in modeling [McCullagh 1980].

3.2.2.7 Applications of multinomial logistic regression model

Assuming that multiple alternatives are available for the dependent variable, by applying multinomial logistic regression, the significance of different factors involved in the model can be studied so that decision makers can have a robust tool by which they can select the most efficient alternative. For instance, multinomial logistic regression is utilized in transportation engineering filed for various purposes. Some of the areas in which the modeling is applied by using multinomial logistic regression are as follows. It has been used to study the predictors that influence the selection of airports when several airports are available in a city area [Windle and Dresner 1995]; it is further used to investigate the influence of parking fees on selecting the travel mode [Li et al. 2008]; it is also used to study the factors impacting the purpose of trips [Penn et al. 2008]. Another example is using this regression model to estimate access mode choice of passengers [Lei et al. 2009]. Multinomial logistic regression was also utilized to investigate if a construction company is undergoing declining status [Koksal and Arditi 2004].

3.2.3 Ordinal Regression Model (Proportional odds model)

As explained earlier, through multinomial logistic regression, a dependent variable which consists of different categorial levels can be modeled by using a set of equations defining the relationship between the independent variables (predictors) and the dependent variable. However, in order to use multinomial logistic regression, there is no need for the categories of the dependent variable to have an ordinal relationship; i.e. simply the fact that more than two possible categories are available for the dependent variable, alongside with the other requirements pertaining to utilization of multinomial logistic regression model which was described earlier, suffices using this type of modeling for the dependent variable. On the other hand, when there indeed exists an ordinal relationship between the various categories of the dependent variable, through applying ordinal regression model, the ordinal relationship of various categorial levels of the dependent variable will be accounted for in the model developed via proportional odds assumption. The need for the proportional odds assumption to be taken into account in ordinal regression models, results in a more restrictive strategy for modeling the dependent variable.

The proportional odds model implies that the influence of each independent variable will be similar for all equations defining the model. In other words, unlike multinomial logistic regression, in an ordinal regression model the coefficients of independent variables will be the same in all equations. Each of the equations pertain to different categories of the dependent variable [McCullagh 1980]. This implies that, for instance with regards to the probability calculation, the difference between different categorial levels of the dependent variable will only be created due to the intercept term in each equation (as the intercept are the only parameters that are permitted to vary for different categorial levels of dependent variable).

In general, ordinal logistic regression models can be defined by using the following equation.

$$ln\,(\frac{P(Y \leq i|X_1,\ X_2, \ldots, X_n)}{P(Y > i+1|X_1,\ X_2, \ldots, X_n)} = a_i - b_1\,X_1 - b_2\,X_2 - b_3\,X_3 - \ldots - b_n\,X_n \qquad (3.22)$$

In this equation, $i=1,2,\ldots, k-1$ pertains to each of the ordinal categorial levels of the dependent variable, $X_1 \ldots X_n$ are the independent variables of the model (predictors), a_i is the intercept of the model for the i^{th} categorial level of the dependent variable, and $b_1 \ldots b_n$

are the coefficients of the ordinal logistic regression model and are associated with each independent variable.

In order to compute the cumulative probabilities corresponding to various categorical levels of the dependent variable, the equation presented below can be applied.

$$P(Y \leq i|X_1, X_2, \dots, X_n) = \frac{exp(a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \dots - b_n X_n)}{1 + exp(a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \dots - b_n X_n)} \quad (3.23)$$

By simplifying the numerator and the denominator of the above equation, it can be rewritten as follows.

$$P(Y \leq i|X_1, X_2, \dots, X_n) = \frac{1}{1 + exp(-(a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \dots - b_n X_n))} \quad (3.24)$$

Once the cumulative probability corresponding to each of the categorical levels of the dependent variable is computed, then the probability corresponding to various categorical levels of the dependent variable can be obtained by deducting the cumulative probabilities of the each of the two successive categorical levels. The following equation demonstrates general form of calculation of probability for each categorical level.

$$P(Y = i + 1|X_1, X_2, \dots, X_n) = P(Y \leq i + 1|X_1, X_2, \dots, X_n) - P(Y \leq i|X_1, X_2, \dots, X_n) \quad (3.25)$$

As stated earlier, in an ordinal logistic regression model, the coefficients associated with each of the independent variables remain constant for all equations corresponding to various categorical levels of the dependent variable; and the only parameter which can take on different values for different categorical levels of the dependent variable, is the intercept term. By taking this point into consideration, it is thus concluded that the set of equations pertaining to an ordinal logistic regression model are indeed parallel to one another (in other words, logit equations for each of the categorial levels of the dependent variable are parallel and the difference is due to the intercept term) [Salman 2010].

In order to provide a sample representation of results associated with an ordinal logistic regression model, herein it will be assumed that the ordinal dependent variable is the condition rating of gravity sewer pipes, ranging from 1 to 5. Additionally, assuming that there is only one predictor involved in the model, and that predictor being the age of the sewer pipe, and assuming that the parameters pertaining to this ordinal regression model are as follows: the intercept term for each of the ordinal categories of 1, 2, 3, and 4 are 4, 5, 6, and 7, respectively; furthermore, the coefficient of the independent variable (i.e. age) is 0.1 for all ordinal equations; therefore, by using the general form of equations defining the ordinal logistic regression models, the graphs corresponding to the cumulative probabilities of the model as well as the graph illustrating the logit equations of the ordinal model with respect to the independent variable (i.e. age) are obtained as follows.

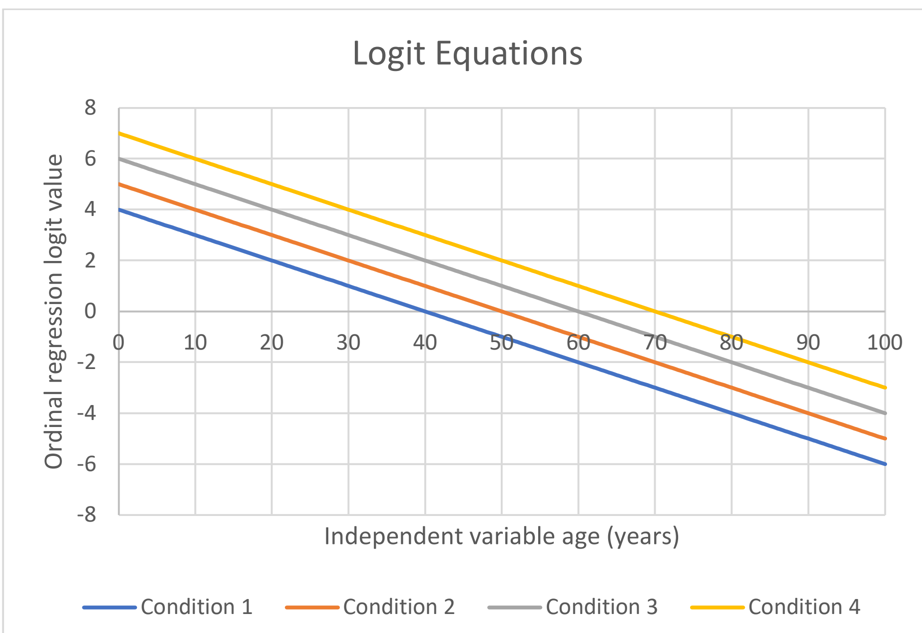

Figure 3.2: Results of the logit equations with respect to age of the pipe for each condition rating

In this example, the abovementioned graphs present the associated values till the age of 100 years for the sewer pipes (for simplicity only one independent variable (age) is considered for this representation). The equations pertaining to each of the condition ratings 1 through 4, considering the parameters stated earlier, are as follows:

$$Condition\ rating\ 1 = ln\ (\frac{P(Y \leq 1|age)}{P(Y > 2|age)} = 4 - \ 0.1 \times age \tag{3.26}$$

$$Condition\ rating\ 2 = ln\ (\frac{P(Y \leq 2|age)}{P(Y > 3|age)} = 5 - \ 0.1 \times age \tag{3.27}$$

$$Condition\ rating\ 3 = ln\ (\frac{P(Y \leq 3|age)}{P(Y > 4|age)} = 6 - \ 0.1 \times age \tag{3.28}$$

$$Condition\ rating\ 4 = ln\ (\frac{P(Y \leq 4|age)}{P(Y > 5|age)} = 7 - \ 0.1 \times age \tag{3.29}$$

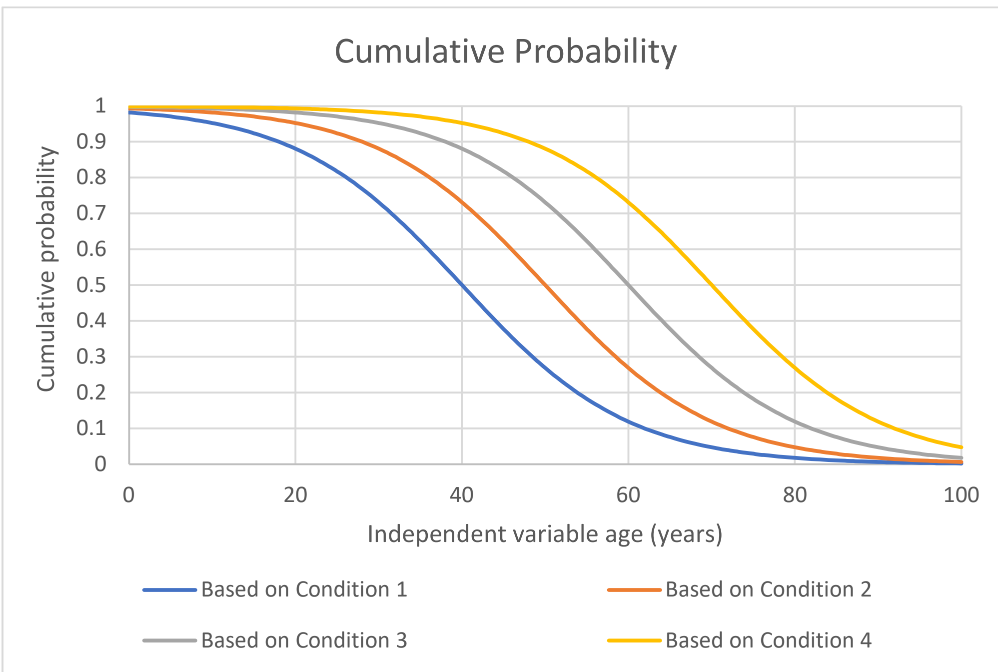

Figure 3.3: Values of the cumulative probability of sewer pipe with respect to age for each condition rating

The equations associated with the cumulative probabilities of the ordinal regression model in this example, are as follows:

$$P(Y \leq 1|age) = \frac{exp(4 - 0.1 \times age)}{1 + exp(4 - 0.1 \times age)} = \frac{1}{1 + exp(-(4 - 0.1 \times age))} \quad (3.30)$$

$$P(Y \leq 2|age) = \frac{exp(5 - 0.1 \times age)}{1 + exp(5 - 0.1 \times age)} = \frac{1}{1 + exp(-(5 - 0.1 \times age))} \quad (3.31)$$

$$P(Y \leq 3|age) = \frac{exp(6 - 0.1 \times age)}{1 + exp(6 - 0.1 \times age)} = \frac{1}{1 + exp(-(6 - 0.1 \times age))} \quad (3.32)$$

$$P(Y \leq 4|age) = \frac{exp(7 - 0.1 \times age)}{1 + exp(7 - 0.1 \times age)} = \frac{1}{1 + exp(-(7 - 0.1 \times age))} \quad (3.33)$$

3.2.3.1 Identifying the unknown parameters

In order to determine the values corresponding to the unknown parameters of the ordinal logistic regression model, similar approach which exhibited for the case of binomial logistic regression and multinomial logistic regression models, will be used. This approach, as stated earlier, is the utilization of maximum likelihood estimation method. For the case of ordinal logistic regression models, the likelihood function is presented in the below equation.

$$\prod_{j=1}^{n}\left[\prod_{i=1}^{k}\pi_i(x_j)^{y_{ji}}\right] = \prod_{j=1}^{n}\left[\prod_{i=1}^{k}\left[P(Y \leq i|X_j) - P(Y \leq i-1|X_j)\right]^{y_{ji}}\right] \quad (3.34)$$

In which j=1, 2, 3, …, n; and it represents the observations of the model, and the various categorial levels of the dependent variable (assuming k ordinal levels exist for the dependent variable) are represented by i=1, 2, 3, …, k. Furthermore, y_{ji} denotes the observation result and if the outcome of j^{th} observation falls into the i^{th} categorial level, then a value of 1 will be assigned to y_{ji}, else, it will take on a value of zero.

Additionally, assuming that the dependent variable has k different categorial levels arranged in an ordinal fashion, similar to what was observed in the multinomial logistic regression model, $k-1$ equations will represent the logit values corresponding to each of the categorial levels of the dependent variable; and by solving these equations simultaneously, all the unknown parameters are obtained. As stated earlier, in the multinomial logistic regression model, in which considering n number of independent variables are involved, the parameters to be found are $k-1$ intercept terms, and $n.(k-1)$ coefficients of the independent variables present in the model. However, unlike multinomial logistic regression model, there are less parameters to be determined in ordinal regression models. This is due to the similar values of the coefficients of the independent variables for different categorial levels of the dependent variables (proportional odds assumption of the model). Therefore, in ordinal regression models with the aforementioned properties, there are $k-1$ intercept terms to be identified as well as n coefficients for the independent variables of the model; hence, a total of $n+k-1$ unknown parameters are to be determined.

Identifying the model significance

In order to determine the significance of the model in an ordinal logistic regression model, similar procedure described in multinomial logistic regression method, can be used. By using the likelihood ratio methodology, the significance of an ordinal logistic regression model can be compared to the base model (a model which only comprises of intercept term of the regression without any independent variables). Based on the difference between the -2 log likelihood of the model comprising of intercept term only, and the -2 log likelihood of the multinomial logistic regression model, the significance of the model can be investigated.

3.2.3.2 Identifying significant coefficients

In order to identify which of the independent variables have significant coefficients in an ordinal logistic regression model, Wald statistic can be utilized. As stated earlier, Wald statistic can be calculated based on the equation provided below.

$$W_i = \left[\frac{b_i}{Standard\ Error\ (b_i)}\right]^2 \tag{3.35}$$

Once implementing the Wald statistic, the null hypothesis (H_0) assumes that the coefficient of an independent variable is equal to zero (i.e. $b_i = 0$) and in order to reject the null hypothesis, the critical value obtained based on the associated chi-square distribution (which only contains one degree of freedom) is checked with the value of W_i (Wald statistic) computed from the aforementioned equation [Salman 2010].

3.2.3.3 Influence of predictors in ordinal logistic regression

Ordinal logistic regression model is quite similar to the previously discussed multinomial logistic regression model; except that in an ordinal logistic regression model, proportional odds assumption exerts more restriction in the modeling of dependent variable. Furthermore, as it was observed in the general form of equations defining the ordinal logistic regression model, unlike multinomial logistic regression, there exists negative signs before the coefficients of each of the independent variables of the model. This indicates that considering a particular independent variable has a positive coefficient, by increasing that particular predictor, the values of the logit functions (set of equations defining the model) will be in turn reduced. Additionally, based on the equations presented for the cumulative probability for each of the ordinal levels of the dependent variable, a similar influence will be observed for the cumulative probabilities of each ordinal level as well. This means that assuming the coefficient of an independent variable is a positive value, increasing the value of the independent variable will result in a reduction in the associated cumulative probability for each of the categorial levels pertaining to the dependent variable.

Moreover, considering that and independent variable is a categorial variable, and that the coefficient of that categorial independent variable is a positive value, then the presentence of that categorial variable will have the same impact as the increase of a numerical independent variable in the values of logit functions and the cumulative probabilities of the ordinal regression model. Similar to the example provided earlier, considering the material type of the pipeline is represented by a categorical variable, and that values 0 and 1 correspond to Polyvinyl Chloride (PVC) and Vitrified Clay Pipe (VCP), it is observed that assuming the coefficient of regression based on ordinal logistic regression approach is positive for this categorial independent variable, hence for Vitrified Clay Pipe the values of logit functions and the cumulative probabilities in an ordinal regression model will be decreased compared to Polyvinyl Chloride pipes.

Furthermore, in the example presented earlier with regards to the age of the sewer pipes as the only independent variable used in the ordinal logistic regression model, it was assumed that the coefficient of the age is 0.1, and it is positive, and additionally, from the graphs representing the values of logit functions and the cumulative probabilities of the ordinal regression model, it was observed that both graphs were portraying decreasing functions for categorial levels of the dependent variables.

3.2.3.4 Verification and Classification Table

The classification table used for a ordinal logistic regression model is similar to the classification table used for multinomial logistic regression models; the number of elements in the classification table for ordinal logistic regression model is similar to the one pertaining to the multinomial regression model. In other words, if there are k possible response categories for the dependent variable of the ordinal logistic regression model, the dimension of the classification table will be $k \times k$ containing k^2 elements (similar to multinomial logistic regression model). The general form of the classification table used in ordinal logistic regressions is identical to the one presented for multinomial logistic regression

Similar to multinomial logistic regression, in an ordinal logistic regression model due to greater number of response categories compared to binary logistic regressions, the assignment of each observed probability of the dependent variable is conducted by taking the greatest value of probability calculated through the logit functions describing the model. As stated earlier, this procedure is analogous to using a cut-off value of 0.5 for binary logistic regression.

Once the classification table is arranged, considering the total number of observations (which is equal to the sum of all elements of the classification table), by summing up the elements on the diagonal part of the classification table, and dividing the result by the total number of observations the percentage of correct predictions can be obtained.

3.2.3.5 Assumptions of ordinal logistic regression

As mentioned earlier, in case there exists an ordinal relationship among response levels of the dependent variable, ordinal regression can be applied in order to reflect the existing ordinal relationship associated with the dependent variable. Furthermore, it was observed that the main difference between multinomial logistic regression and ordinal logistic regression is the fact that in multinomial logistic regression, for each of the set of logit equations used to define the model, the coefficients of the independent variables as well as the intercept terms in each individual equation, may differ from one another. However, when using the ordinal regression model, there is a more restrictive requirement pertaining the coefficient of the independent variables of the model; this additional assumption requires that the coefficients of the independent variables remain unchanged for all of the individual logit equations in ordinal logistic regression. This assumption makes ordinal logistic regression become more restrictive compared to multinomial logistic regression and is known as the proportional odds assumption. However, there's no such restraint with regards to the intercept terms and they can be assigned different values in each individual logit equation.

When using IBM SPSS software, the proportional odds assumption can be verified by using the option provided by the software, known as *Test of Parallel Lines.* When using *Test of Parallel Lines* option, in fact, a comparison is made between -2 log likelihood of a model which maintains the requirement of proportional odds assumption and -2 log likelihood of another model which does not have the restriction of having identical coefficients of independent variable for all individual equations, and the coefficients of independent variables can take on different values in each equation [Norusis 2008]. Based on the null hypothesis (H_0), the coefficients of the independent variables in each of the individual equations do not change and remain constant. Therefore, in order for the ordinal logistic regression model to be valid for the dependent variable, the null hypothesis should not be rejected. In other words, in order for the proportional odds assumption to hold true in modeling the dependent variable, the null hypothesis should be validated.

Hence, if the result of the chi-square test used for the likelihood ratios is not significant, therefore, it means that the null hypothesis is not rejected and that the coefficients of the independent variables in each of the individual equations do not vary and remain constant; thus, the use of ordinal logistic regression is suitable for modeling the dependent variable. Therefore, it is realized that the result of *Test of Parallel Lines* option should not yield a significant outcome. On the other hand, if the result of *Test of Parallel Lines* option is significant, therefore the null hypothesis will be rejected; meaning the coefficient of the predictors will vary in different equations. The outcome of *Test of Parallel Lines* option is dependent upon the independent variables used in the ordinal regression model. In order to avoid the *Test of Parallel Lines* to be significant, the independent variables which are not significant can be removed from the regression model. It should be borne in mind that various link functions can be used in order to achieve a suitable model which appropriately preserves the ordinal relationship of the dependent variable for the desired predictors. Hence, different link functions can be checked to realize which one best models the ordinal relationship of the dependent variable.

3.2.3.6 Alternative link functions for ordinal regression

As stated earlier, SPSS package has various link functions which can be used instead of the logit function, which was the binomial, multinomial and ordinal logistic regressions discussed herein, were all built upon. Ordinal logistic regression is a particular form of generalized linear models, in which logit functions play the role of link functions and build the relationship between the ordinal dependent variable and the predictors of the model [Norusis 2008]. In some cases, the logit function might not necessarily result in a suitable regression result; hence, in such cases, other link functions can be tested to develop the most effective set of equations for constructing the relationship between the independent variables and the ordinal dependent variable. Therefore, by substituting the logit function with the most suitable link function, the regression model pertaining to the ordinal regression will have the following format.

$$Link\ function\ (P \leq i) = a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \ldots - b_n X_n \tag{3.36}$$

Various link functions are provided by SPSS package; and each of these link functions are most suitable for specific cases. Herein, some of the link functions are presented with their best applications.

Complementary log-log is a link function which is best suitable for the cases where the dependent variable has greater probabilities for higher levels of the dependent variable; by using this link function, the format of the regression equations become as follows [Norusis 2008].

$$\ln(-ln\,(1-\gamma)) = a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \ldots - b_n X_n \tag{3.37}$$

The next available link function is *Probit* function which is best suitable when the latent variable follows a normal distribution and using *Probit* function results in the following form of regression equations [Norusis 2008].

$$\phi^{-1}(\gamma) = a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \ldots - b_n X_n \tag{3.38}$$

In case the outcome of the model contains some extreme values, then the most suitable link function might be *Cauchit* function. The following equation illustrates the regression equation with *Cauchit* function as link function [Norusis 2008].

$$\tan\big(\pi(\gamma - 0.5)\big) = a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \ldots - b_n X_n \tag{3.39}$$

Another link function which might best suit the modeling in cases where higher probabilities are attributed to the lower levels of the dependent variable, is *Negative log-log* function. When this link functions is used to describe the model, the equation becomes as follows [Norusis 2008].

$$-\ln(-ln\,(\gamma)) = a_i - b_1 X_1 - b_2 X_2 - b_3 X_3 - \ldots - b_n X_n \tag{3.40}$$

Therefore, as observed herein, several other link functions with their best application scenarios are also available to be used instead of logit function.

3.2.3.7 Applications of ordinal regression model

Ordinal regression is used by various researchers in order to model the deterioration rate of infrastructures. Various authors have recommended that due to the restrictions of ordinary least squares (OLS) regression when used in modeling ordinal data, using ordinal regression will result in more appropriate models. Some examples of application of ordinal regression models in deterioration modeling of infrastructures are as follows: In order to predict the transition probabilities in bridges, ordered probit regression was used along with Markov chain [Madanat et al. 1995]. A similar method was also used to predict the transition probabilities in sewer pipes [Baik et al. 2006]. Furthermore, the condition rating of stormwater pipes was predicted by using proportional odds model and then a comparison was made with probabilistic neural network method [Tran et al. 2009].

3.3 LightGBM as an Efficient Gradient Boosting Decision Tree

One of the widely utilized machine learning techniques is gradient boosting decision tree (GBDT), and this is mainly due to the fact that it provides accuracy, efficiency and interpretability [Friedman 2001, Ke et al. 2017]. With regards to various machine learning tasks, state of the art achievements are accomplished through utilizing gradient boosting decision tree. The aforementioned machine learning tasks include the following [Li 2012, Richardson et al. 2007, Burges 2010, Ke et al. 2017]:

- Multi-class classification
- Click prediction
- Learning to rank

With advent of big data, as the number of instances as well as the number of features are rapidly increasing, the trade-off between efficiency and accuracy has become significantly more notable when using gradient boosting decision tree. With the traditional gradient boosting decision tree is employed, and in order to evaluate the information gain of all split points, all the data instances are required to be scanned. This ultimately means that the number of features as well as the number of instances will determine the computational intricacies of these models. In some cases, when using big data, the aforementioned task will become time consuming [Ke et al. 2017].

One of the possible solutions when using big data is to use less number of features as well as less number of instances. This solution may cause problems such as how to sample data for use in gradient boosting decision tree. To overcome this problem, various techniques have been proposed by researchers which can be used for increasing the speed of the training process in boosting [Friedman et al. 2000, Dubout and Fleuret 2011, Appel et al. 2013, Ke et al. 2017]. These approaches include the following:

- Gradient-based One-Side Sampling (GOSS)
- Exclusive Feature Bundling (EFB)

3.3.1 Gradient Boosting Decision Tree

The decision tree, when using gradient boosting decision tree (GBDT), is obtained through iteration and fitting the negative gradients (residual errors). Therefore, an ensemble model of decision trees results in the gradient boosting decision tree (through sequential training) [Friedman 2001, Ke et al. 2017]. When learning a decision tree, finding the most suitable split point may require considerable time. Pre-sorted algorithm and histogram-based algorithm are two common tools utilized for obtaining the most suitable split point [Mehta et al. 1996, Shafer et al. 1996, Ranka and Singh 1998, Jin and Agrawal 2003, Li et al. 2007]. Through histogram-based algorithm, discrete bins are utilized for continues feature values and throughout training, these bins are used to construct feature histograms, whereas when using pre-sorted algorithm, the most suitable split point is found on the sorted feature values. The algorithm pertaining histogram-based approach is presented below [Ke et al. 2017].

- Input: Training data (*I*), Max depth (*d*)
- Input: *m*: feature dimension
- nodeSet ← {0} ▷ tree nodes in current level
- rowSet ← {{0,1,2,…}} ▷ data indices in tree nodes
- for *i* = 1 to *d* do:
    - for *node* in nodeSet do:
        - usedRows ← rowSet{node}
        - for *k* = *1* to *m* do:
            - *H* ← new Histogram ()
            - ▷ Build histogram
            - for *j* in usedRows do:
                - bin ← I.f[k][j].bin
                - *H*[bin].y ← *H*[bin].y + I.y[j]
                - *H*[bin].n ← *H*[bin].n + I
                -
            - Find the most suitable split on histogram *H*
            - …
    - Update *rowSet* and *nodeSet* based upon the most suitable split points
    - …

Through down sampling the data, the size of the training data can be lessened. For instance, by considering a pre-defined threshold, and by comparing the weights of the data, data filtering can therefore be achieved [Friedman et al. 2000]. Furthermore, by considering stronger features, the number of features can thus be lessened as well. Both projection pursuit as well as principle component analysis can be utilized for this purpose. It should be noticed though, that by assuming features include significant redundancies, accuracy might be impacted, as this presumption may not always hold true. In other words, as each feature may contain specific attributes, therefore, the training accuracy could potentially be influenced due to this assumption [Appel 2013, Jolliffe 2002, Ke et al. 2017, Jimenez and Landgrebe 1999, Zhou 2012].

When utilizing the pre-sorted algorithm, within gradient boosting decision tree the features which have zero values are disregarded, and therefore, this results in a reduction in the cost of training. On the other hand, when utilizing histogram-based algorithm for gradient boosting decision tree, whether the feature values are zero or not, the feature bin values are required for this approach [Chen and Guestrin 2016, Ke et al. 2017]. In the following sections, the two previously stated methods, i.e. gradient-based one-side sampling and exclusive feature bundling are described in detail.

3.3.2 Gradient-Based One-Side Sampling (GOSS)

Gradient-based one-side sampling (GOSS) which was introduced by Ke et al. in 2017 is a sampling approach for gradient boosting decision tree which results in balancing the accuracy of decision trees while lessening the number of data instances.

3.3.2.1 Algorithm for Gradient-Based One-Side Sampling (GOSS)

When utilizing AdaBoost, the importance of data are judged through the weight of the samples. On the other hand, due to the fact that there exist no native sample weights in gradient boosting decision tree, therefore, the sampling approaches utilized in AdaBoost cannot be directly used. However, within gradient boosting decision tree, it is observed that beneficial information with regards to sampling the data can be achieved through the gradient for each data instance. In other words, the data instances which demonstrate minor values of gradients can be disregarded. This is due to the fact that when the gradient value for a data instance is small, it is observed that the data instance is therefore adequately trained and it contains smaller values of training error, which allows these data instances to be disregarded. By disregarding these data instances, the distributions associated with the data will be subjected to change. As a result of altering the distribution of data, the accuracy will therefore be impacted. Gradient-based one-side sampling (GOSS) can be utilized to address this issue.

When applying gradient-based one-side sampling, with regards to the data instances which have smaller gradients, random sampling will be conducted while maintaining the data instances with greater gradients. However, in order to account for the changes in the distribution of data, by using gradient-based one-side sampling, for data instances which have smaller gradients a constant multiplier will be utilized in obtaining the information gain. In other words, when using gradient-based one-side sampling, based upon the absolute gradient values pertaining to each data instance, the data instances are sorted; next, the data instances which have greater values of gradients are selected (top *a* percent of sorted data). With regards to the remainder of the data instances, *b* percent of the data instances are randomly selected.

In the next step, in order to preserve the initial distributions associated with the data as much as possible, and yet to concentrate on the data instances which are not adequately trained, i.e. data instances which have greater values of gradients, in gradient-based one-side sampling the selected data instances which demonstrate minor values of gradients are multiplied by the following ratio: $(1-a)/b$ [Ke et al. 2017].

The following algorithm demonstrates gradient-based one-side sampling approach [Ke et al. 2017]:

- Input: Training data (I), Iterations (d)

- Input: Sampling ratio of data instances with greater gradients (*a*)
- Input: Sampling ratio of data instances with smaller gradients (*b*)
- Input: *loss*: loss function, *L:* weak learner
- models ← {}, fact ← (1-*a*)/*b*
- topN ← a × len (I), randN ← b × len (I)
- for *i* = 1 to *d* do:
    - preds ← models.predict (*I*)
    - g ← *loss* (I, preds), w ← {1,1,…}
    - sorted ← GetSortedIndices(abs(g))
    - topSet ← sorted[1:topN]
    - randSet ← RandomPick (sorted[topN:len(I)], randN)
    - usedSet ← randSet + topSet
    - w[randSet] × = fact ▷ Assign weight *fact* to the data instances with smaller values of gradients
    - newModel ← L(I [usedSet], -g [usedSet], w [usedSet])
    - models.append(newModel)

3.3.3 Exclusive Feature Bundling (EFB)

In order to decrease the number of features, exclusive feature bundling can be used. This method was introduced by Ke et al. in 2017. When data are high-dimensional, they are typically sparse as well. Therefore, in order to decrease the number of features, and by taking advantage of the feature space being sparse, an almost lossless approach can be applied. In particular, when feature space is sparse, this indicates that a lot of the features do not take on non-zero values at the same time; in other words, a lot of the features are mutually exclusive.

Therefore, exclusive feature bundle can be achieved through bundling the exclusive features into a separate feature. Feature histograms constructed by using the feature bundles can be made identical to histograms pertaining to individual features through implementing a specific feature scanning algorithm. Through this approach, the training process of the gradient boosting decision tree can be performed faster and yet the accuracy will not be impacted [Ke et al. 2017].

3.4 CatBoost as unbiased boosting

CatBoost is a novel gradient boosting approach which was introduced by Prokhorenkova et al. in 2018. In CatBoost methodology, ordered boosting, which is a permutation-driven algorithm, as well as an algorithm utilized for handling categorical features are introduced. By utilizing the aforementioned algorithms, the prediction shift which occurs due to a particular type of target leakage, will be handled [Prokhorenkova et al. 2018].

Gradient boosting is a machine-learning approach which can be used in a broad range of problems. For instance, gradient boosting can be utilized in learning problems which contain noisy data, intricate dependencies, and heterogeneous features. Examples of these problems include forecasting weather, web searches, etc.; through gradient boosting methodology, and by utilizing gradient descent within a functional space, an ensemble predictor will therefore be achieved. Gradient boosting approach is based upon constructing strong predictors by iteratively utilizing weaker models, i.e. base predictors [Kearns and Valiant 1994, Prokhorenkova et al. 2018, Roe et al. 2005, Wu et al. 2010, Zhang and Haghani 2015].

When using gradient boosting method, the prediction model, *F*, which is achieved through various boosting steps, is dependent upon the targets of training instances. The paper presented by Prokhorenkova et al. in 2018, illustrates that a shift in distribution of $F(x_k) \mid x_k$ for a training example x_k from distribution of $F(x) \mid x$ for a test sample x will be observed. Therefore, a prediction shift will occur in the learned model. Prokhorenkova et al. recognize this issue as a particular type of target leakage. Furthermore, with regards to preprocessing categorical features, converting categories to their target statistics is an effective approach that can be utilized in gradient boosting. Additionally, a target statistic is a statistical model which may result in target leakage as well as prediction shift [Cestnik et al. 1990, Prokhorenkova et al. 2018, Micci-Barreca 2001].

In order to deal with the aforementioned issues, Prokhorenkova et al. proposed ordering principle. By utilizing ordering principle, ordered boosting, which is a modification of gradient boosting algorithm, will be achieved. Through ordered boosting, occurrence of target leakage will be prevented. Furthermore, Prokhorenkova et al. also introduced a novel algorithm in order to preprocess categorical features as well. Categorical Boosting (CatBoost) is the result achieved based on combination of the aforementioned algorithms [Prokhorenkova et al. 2018].

3.4.1 Categorical features

Categorical features are features which are represented through categories, and furthermore these categories cannot be compared to one another. When utilizing boosted trees, a common approach for considering categorical features is to apply one-hot encoding. When applying one-hot encoding, for each categorical feature, a binary feature will be represented instead [Chapelle 2015, Micci-Barreca 2001]. However, in some cases, for instance when considering high cardinality features, one-hot encoding can result in large number of newly introduced features. In order to deal with this problem, categories can be initially grouped into limited number of clusters and after that, one-hot encoding can be utilized. A common approach for grouping categories is to utilize target statistics (TS) which estimate the expected target values for categories. Furthermore, Micci-Barreca proposed to consider target statistics as new numerical features. Utilizing this approach, categorical features can be managed efficiently and with minimum information loss. For instance, target statistics (TS) can be utilized for click prediction task; in this application, regions, ads, publishers, and users may be considered the main categorical features [Bottou and Cun 2004, Prokhorenkova et al. 2018, Langford et al. 2009, Ling et al. 2017].

3.4.2 Prediction shift in gradient boosting

Based on the study conducted by Prokhorenkova et al. in 2018, the prediction shift in gradient boosting occurs as a result of a particular type of target leakage; furthermore, in this study it is stated that in order to deal with the prediction shift, ordered boosting can be utilized. Base predictor h^t can be approximated by the following equation:

$$h^t = \arg\min \frac{1}{n} \sum_{k=1}^{n} (-g^t(x_k, y_k) - h(x_k))^2 \qquad (3.41)$$

In the above equation, the parameters are as follows:

$D = \{(x_k, y_k)\}_{k = 1..n}$ denotes the dataset,

$x_k = (x_k^1, \dots, x_k^m)$ denotes random vector of m features,

y_k denotes the target and can be binary or numerical value,

h^t is a base predictor and is selected from a family of functions H to minimize the expected loss and can be described as follows:

$$h^t = \arg\min L\ (F^{t-1} + h) \qquad (3.42)$$

Where,

$L(., .)$ denotes a smooth loss function

F is the function which is minimizes the loss function. Furthermore, within a gradient boosting methodology, a sequence of F^t are constructed for $t = 0, 1, \dots;$ F^t is found based on F^{t-1} and through an additive approach:

$$F^t = F^{t-1} + \alpha\, h^t \qquad (3.43)$$

In which, α is the step size [Friedman 2001, Prokhorenkova et al. 2018].

Additionally, in order to deal with the minimization, the Newton method utilizing second order approximation of $L\ (F^{t-1} + h^t)$ at F^{t-1} or through negative gradient step is implemented. These methodologies are considered as types of functional gradient descent [Friedman et al. 2000, Mason et al. 2000, Prokhorenkova et al. 2018]. Moreover, h^t , the gradient step, is selected so that $-g^t(x, y)$ is approximated by $h^t(x)$ in which the following equation represents $g^t(x, y)$:

$$g^t(x, y) := \frac{\partial L(y, s)}{\partial s}|_{F^{t-1}(x)} \tag{3.44}$$

As described by Prokhorenkova et al., the following shifts can transpire within gradient boosting approach:

- Shift occurring in conditional distribution of gradient $g^t(x_k, y_k) \mid x_k$ from that distribution on a test example $g^t(x, y) \mid x$ (By considering randomness in $D \setminus \{x_k\}$)
- Furthermore, the base predictor h^t will be biased
- Eventually, the trained model F^t will be impacted [Prokhorenkova et al. 2018].

Further details and analysis related to prediction shift can be obtained from Prokhorenkova et al. (2018).

3.4.3 Ordered Boosting

Prokhorenkova et al. proposed boosting algorithm through which previously stated prediction problem will not occur. For each boosting step, a new dataset D_t will be sampled independently and therefore, by utilizing the existing model for the new training example, unshifted residuals will be achieved. Considering that a model is learned by using I trees, and in order to have an unshifted residual $r^{I-1}(x_k, y_k)$, F^{I-1} trained without example x_k will be required. Due to the fact that unbiased residuals are required for every training example, therefore, examples cannot be used for training F^{I-1}. For instance, by considering that a random permutation σ of the training examples is used and n distinct supporting models $M_1, M_2, \ldots, M_n$ are maintained, wherein M_i is learned through the first i examples within the permutation, therefore, for each step, in order to find the residual for the j^{th} sample, M_{j-1} model will be utilized. This algorithm pertaining to this approach is known as ordered boosting. The boosting approaches utilized in CatBoost be either Ordered or Plain. The Ordered boosting is obtained through efficient adjustments made in the algorithm illustrated below [Prokhorenkova et al. 2018]. In the following section, the algorithm pertaining to CatBoost will be presented.

- Input: $\{(x_k, y_k)\}_{k=1}^{n}$, I
- σ ← random permutation of [1,n]
- M_i ← 0 for $i = 1, 2, \ldots, n$
- for t ← 1 to I do:
  - for i ← 1 to n do:
    - r_i ← y_i - $M_{\sigma(i)-1}(x_i)$
  - for i ← 1 to n do:
    - r_i ← y_i - $M_{\sigma(i)-1}(x_i)$
    - ΔM ← LearnModel $((x_j, r_j) : \sigma(j) \leq i)$;
    - M_i ← $M_i + \Delta M$
- Return M_n

3.4.4 CatBoost Algorithm

In CatBoost approach, initially $s+1$ random permutations of the training dataset are constructed. The permutations $\sigma_1, \ldots, \sigma_s$, which are independent as well, are utilized for finding splits which determine tree structures (internal nodes). Furthermore, σ_0 is utilized for selecting the leaf values b_j pertaining to the acquired trees. The algorithm for CatBoost is presented below [Prokhorenkova et al. 2018]:

- Input: $\{(x_i, y_i)\}_{i=1}^{n}$, I, α, L, s, *Mode*
- σ_r ← random permutation of [1,n] for $r = 0.. s$;
- $M_0(i)$ ← 0 for $i = 1, 2, \ldots, n$
- if *Mode* = *Plain* then:
  - $M_r(i)$ ← 0 for $r = 1..s$, $i : \sigma_r(i) \leq 2^{j+1}$

- if *Mode* = *Ordered* then:
    - for $j \leftarrow 1$ to $[log_2\, n]$ do:
        - $M_{r,j}(i) \leftarrow 0$ for $r = 1..s$, $i = 1..\, 2^{j+1}$
- for $t \leftarrow 1$ to I do:
    - $T_t, \{M_r\}_{r=1}^{s} \leftarrow BuildTree\,(\{M_r\}_{r=1}^{s}, \{(x_i, y_i)\}_{i=1}^{n}, \alpha, L, \{\sigma_i\}_{i=1}^{s}, Mode)$;
    - $Leaf_0\,(i) \leftarrow GetLeaf\,(x_i, T_t, \sigma_0)$ *for* $i = 1..n$;
    - $grad_0 \leftarrow CalcGradient\,(L, M_0, y)$;
    - foreach *leaf* j in T_t do:
        - $b^t_j \leftarrow$ -avg ($grad_0\,(i)$ for $i : leaf_0\,(i) = j$)
    - $M_0\,(i) \leftarrow M_0\,(i) + \alpha\, b^t_{leaf0\,(i)}$ for $i = 1..n$
- return $F(x) = \sum_{t=1}^{I} \sum_j \alpha\, b_j^t \mathbb{1}_{\{GetLeaf(x,T_t,ApplyMode)=j\}}$

In the CatBoost algorithm, the *BuildTree* function is as presented in the following algorithm:

- Input: M, $\{(x_i, y_i)\}_{i=1}^{n}$, α, L, s, $\{\sigma_i\}_{i=1}^{s}$, *Mode*
- $grad \leftarrow CalcGradient\,(L, M, y)$;
- $r \leftarrow random\,(1, s)$
- if *Mode* = *Plain* then:
    - $G \leftarrow (grad_r\,(i)$ for $i = 1..n$);
- if *Mode* = *Ordered* then:
    - $G \leftarrow (grad_{r\,,[\log_2(\sigma_r(i)-1]}\,(i)$ for $i = 1..n$);
- $T \leftarrow$ empty tree;
- foreach *step of top-down procedure* do:
    - foreach *candidate split c* do:
        - $T_c \leftarrow$ add splic c to T
        - $leaf_r\,(i) \leftarrow GetLeaf\,(x_i, T_c, \sigma_r)$ for $i = 1..n$;
        - if *Mode* = *Plain* then:
            - $\Delta\,(i) \leftarrow$ avg ($grad_r\,(p)$ for $p : leaf_r\,(p) = leaf_r\,(i)$) for $i = 1..n$;
        - if *Mode* = *Ordered* then:

- $\Delta(i) \leftarrow$ avg ($grad_{r,[\log_2(\sigma_r(p)-1]}$ (i) for $p : leaf_r(p) = leaf_r(i)$, $\sigma_r(p) < \sigma_r(i)$) for $i = 1..n$;
  - $loss(T_c) \leftarrow \cos(\Delta, G)$
- $T \leftarrow \arg\min_{T_c}(loss(T_c))$

- $leaf_{r'}(i) \leftarrow GetLeaf(x_i, T, \sigma_{r'})$ for $r' = 1..s, i = 1..n;$
- if *Mode* = *Plain* then:
  - $M_{r'}(i) \leftarrow M_{r'}(i) - \alpha$ avg($grad_{r'}(p)$ for $p : leaf_{r'}(p) = leaf_{r'}(i)$) for $r' = 1..s$, $i = 1..n$;
- if *Mode* = *Ordered* then:
  - for $j \leftarrow 1$ to $[\log_2 n]$ do:
    - $M_{r',j}(i) \leftarrow M_{r',j}(i) - \alpha$ avg($grad_{r',j}(p)$ for $p : leaf_{r'}(p) = leaf_{r'}(i)$, $\sigma_{r'}(p) \leq 2^j$) for $r' = 1..s$, $i : \sigma_{r'}(i) \leq 2^{j+1}$;
- return T, M

3.4.5 Reduced Complexity and Feature Combinations in CatBoost Algorithm

As observed in the algorithm illustrating CatBoost approach, in order to decrease the complexity of computations of the model, and when the mode of the CatBoost model is the ordered mode, solely the values pertaining to $M'_{r,j}(i) \coloneqq M_{r,2^j}(i)$ for values of $j = 1, \ldots, [\log_2 n]$ and for values of i when $\sigma_r(i) \leq 2^{j+1}$ are stored and updated. Implementing this approach results in a reduction of complexity of computations in the CatBoost model. Additionally, a significant property of CatBoost is that combinations of categorical features are utilized as additional categorical features; through this approach, CatBoost will be able to deal with high order dependencies [Prokhorenkova et al. 2018].

Chapter 4 : Acquisition and Analysis of Data

4.1 Overview

In this chapter descriptive data analysis associated with sewer pipes is performed. In total 410 pipe segments were studied in this study. The analysis of sewer pipe data assists with understanding of the current condition of data with respect to various independent variables. The insights obtained through descriptive data analysis further helps with determining which modeling techniques can be utilized for deterioration modeling of assets.

In this chapter of dissertation, the analysis pertaining to the following independent variables are presented:

- Diameter of sewer pipes
- Sewer pipe material
- Age of sewer pipe
- Pipe slope
- Length of sewer pipe
- Average flow in pipe (%full)
- Average velocity of sewer flow
- Average flow depth

Additionally, the analysis pertaining to condition grading of sewer pipes are presented as well.

4.2 Diameter of sewer pipe

In this section the diameters associated with sewer pipes are presented. As observed from the below figures, diameters of sewer pipes considered in this study range from 21 inches up to 66 inches. Furthermore, it is observed than more than 50% of the sewer pipes have diameters between 20 inches and 30 inches and only 1.46% of the pipes have diameters greater than 55 inches. Based on the results obtained from descriptive analysis, it is observed that the mean value for diameter of sewer pipes is 32.21 inches and standard deviation of sewer pipe diameter is 9.533 inches.

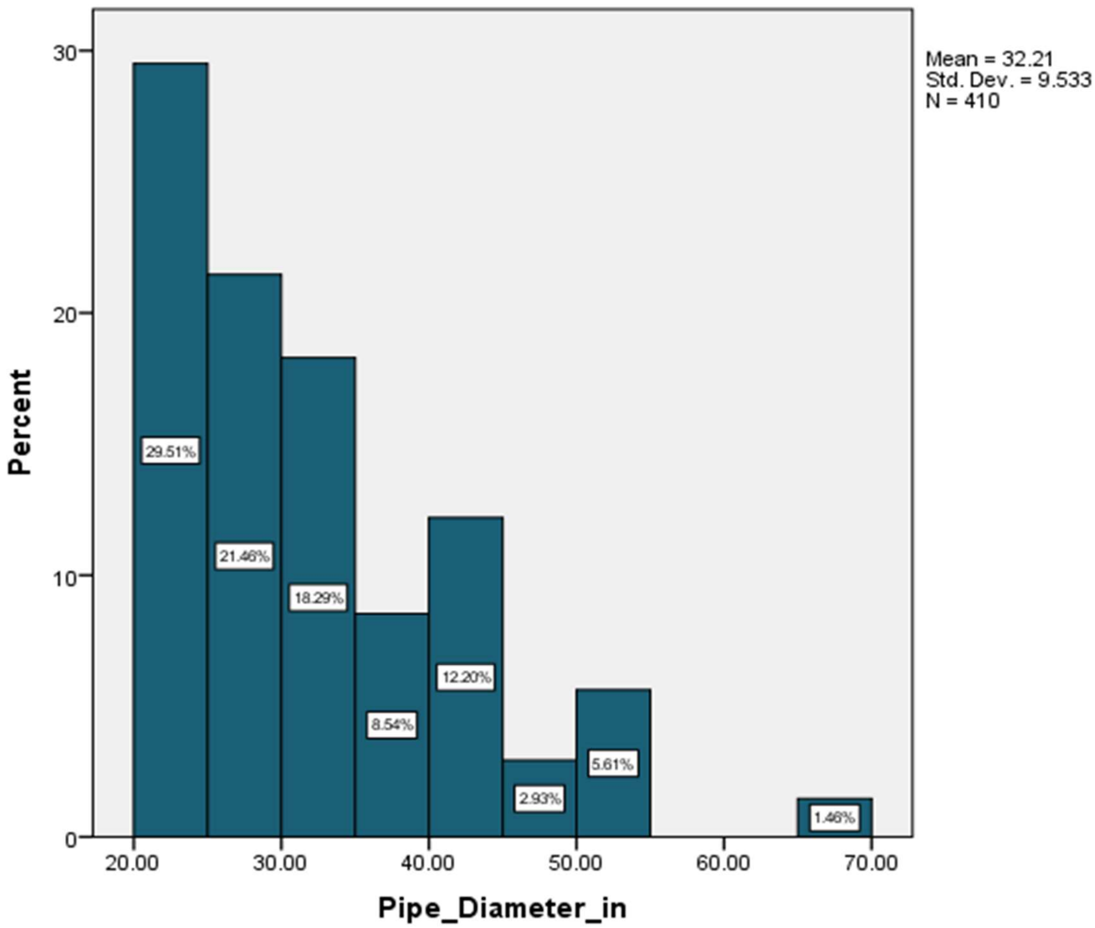

Figure 4.1: Distribution associated with pipe diameter

The following figure illustrates the diameters of sewer pipes for different pipe materials. Based on this figure, it is observed that sewer pipe with the following pipe material have diameters less than 40 inches:

- PVC
- CCFRPM
- VCP

Furthermore, it is observed that pipes with diameters larger than 40 inches are either RCP or FRP.

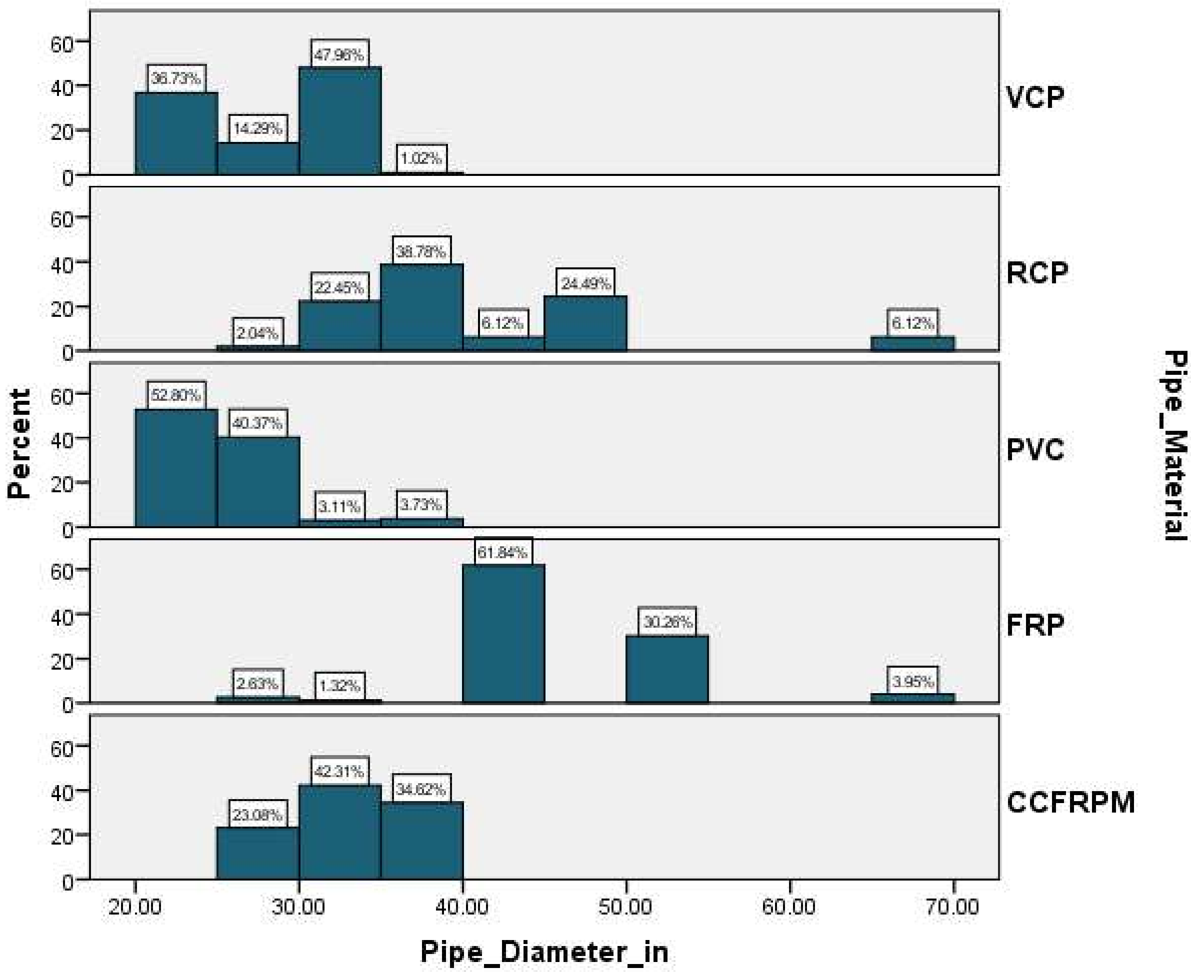

Figure 4.2: Distribution of pipe diameter for different pipe materials

4.3 Age of sewer pipe

The analysis of the age of sewer pipes is illustrated in the following figures. Based on the below figure it is observed that the mean value of age of sewer pipes is 25.06 years and the standard deviation of the age of sewer pipes is 13.29 years. The two highest frequency of the age of pipes are associated with 35 year to 40 year range with 21.71% of the sewer pipes and 15 year to 20 year range with 20.24% of the assets. Furthermore, it is observed that more than 99% of the sewer pipes are less than 50 years old. The minimum and maximum values of the age of the pipes are 4 and 64 years, respectively.

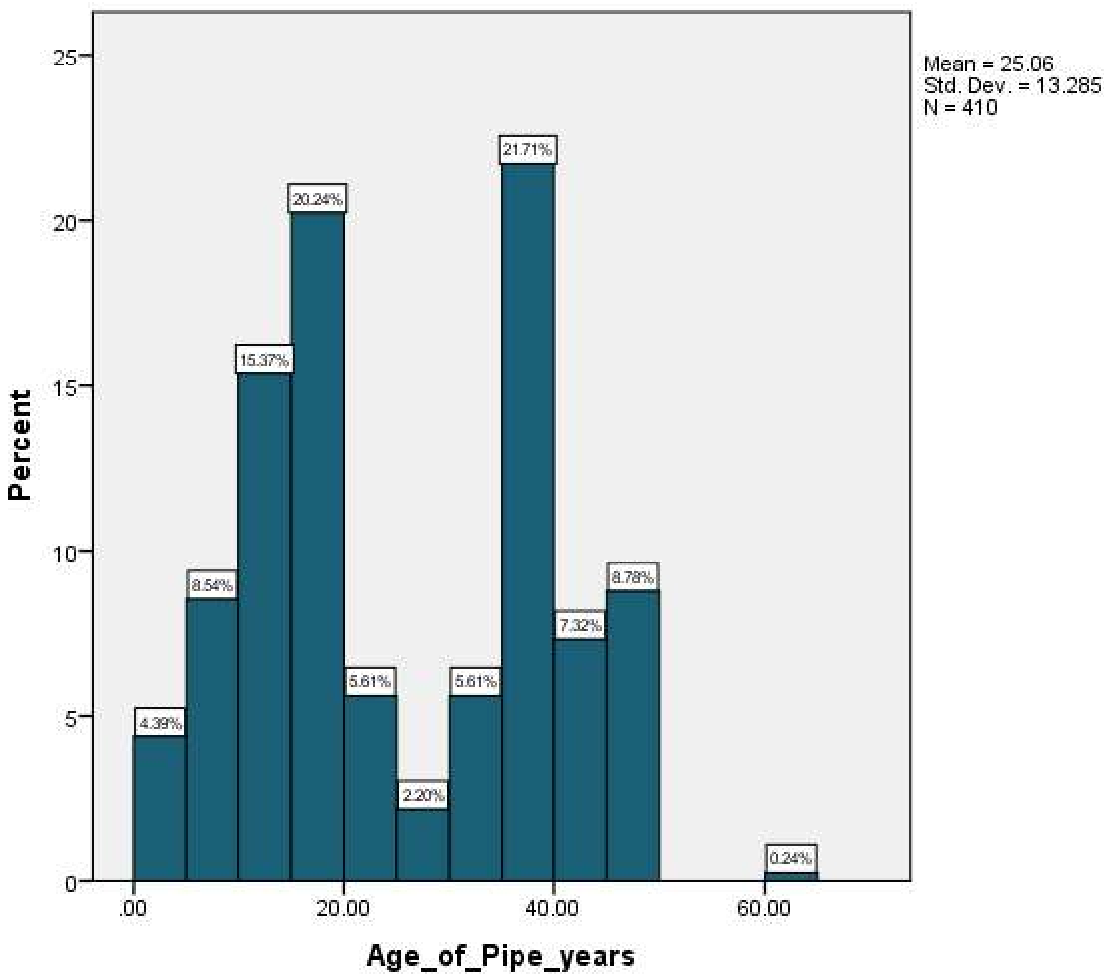

Figure 4.3: Distribution associated with age of pipe

The following figures illustrates the distribution of ages of pipes for different pipe materials. Based on observations made in this figure, the oldest pipes are from vitrified clay. Furthermore, FRP pipes are within 15 to 25 years old and CCFRPM and RCP, and PVC pipes are less than 15, 45, and 40 years old, respectively.

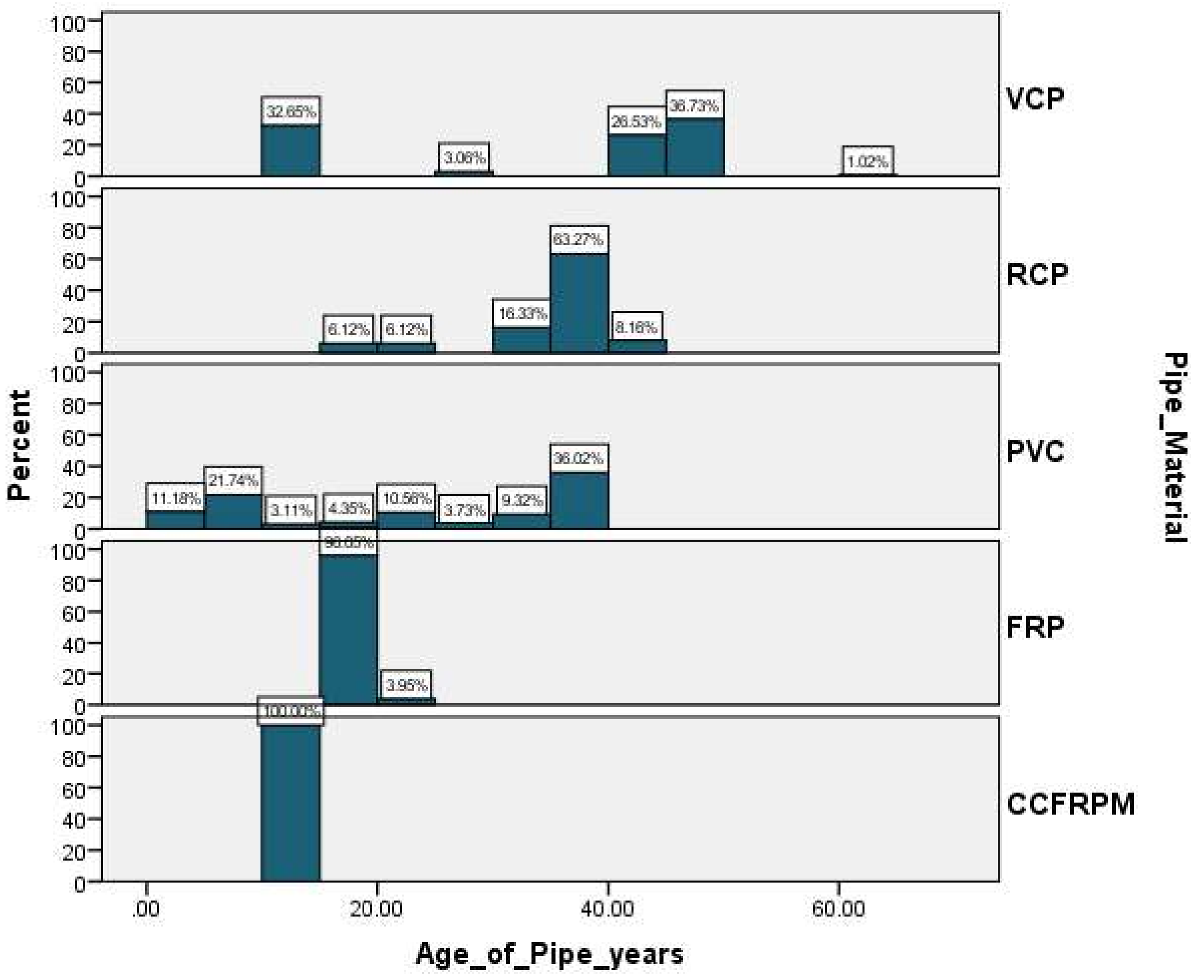

Figure 4.4: Distribution of age of pipe for different pipe materials

4.4 Slope of sewer pipes

The analysis of the slope of sewer pipes is presented in the following figure. Based on this figure, it is observed that more than 95% of the sewer pipes have slopes less than 1%. The mean value and the standard deviation of the slope of the sewer pipes are 0.44% is 1.072%, respectively.

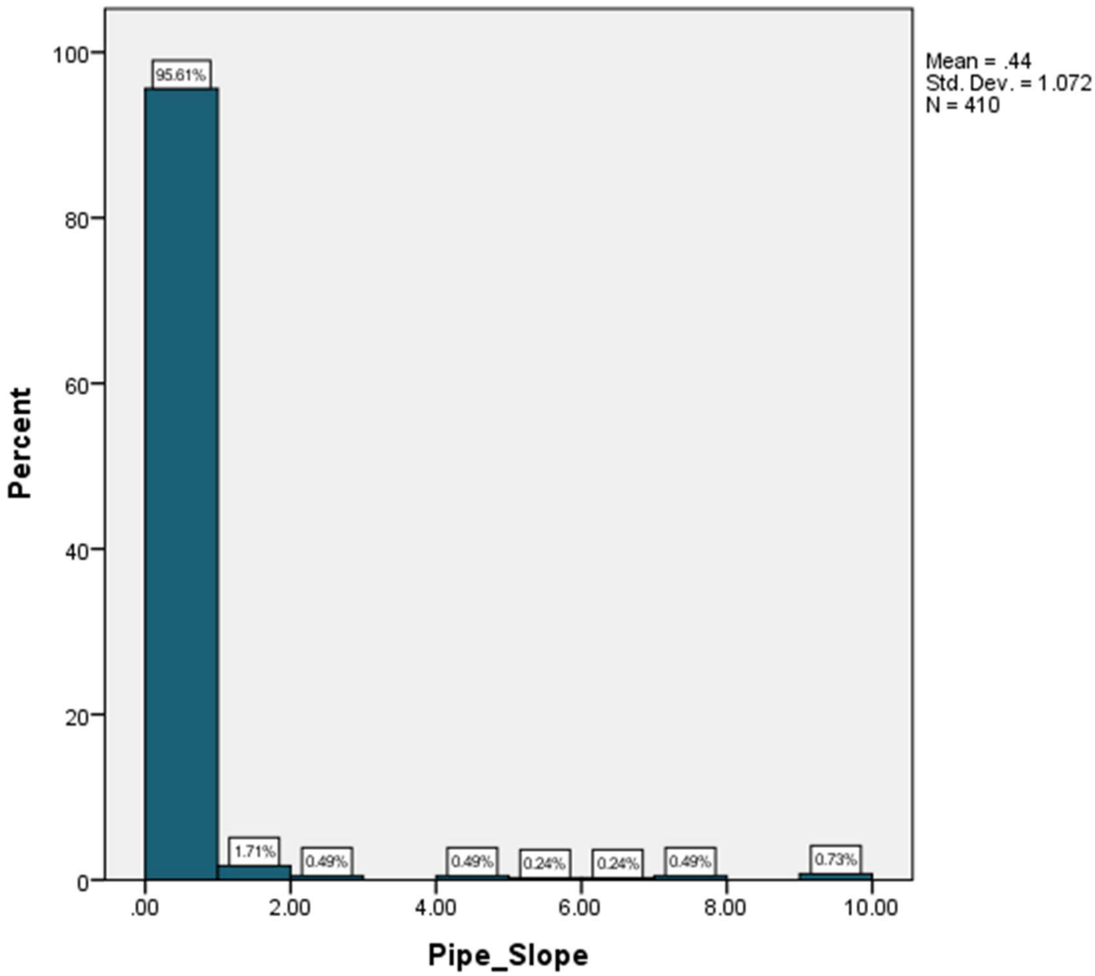

Figure 4.5: Distribution associated with pipe slope

Based on the following figure, it is realized that the highest slope is associated with PVC pipes with the slope value of 9.7%. Moreover, vitrified clay pipes also illustrated the second highest slopes among various pipe materials with the slope value of 7.44%.

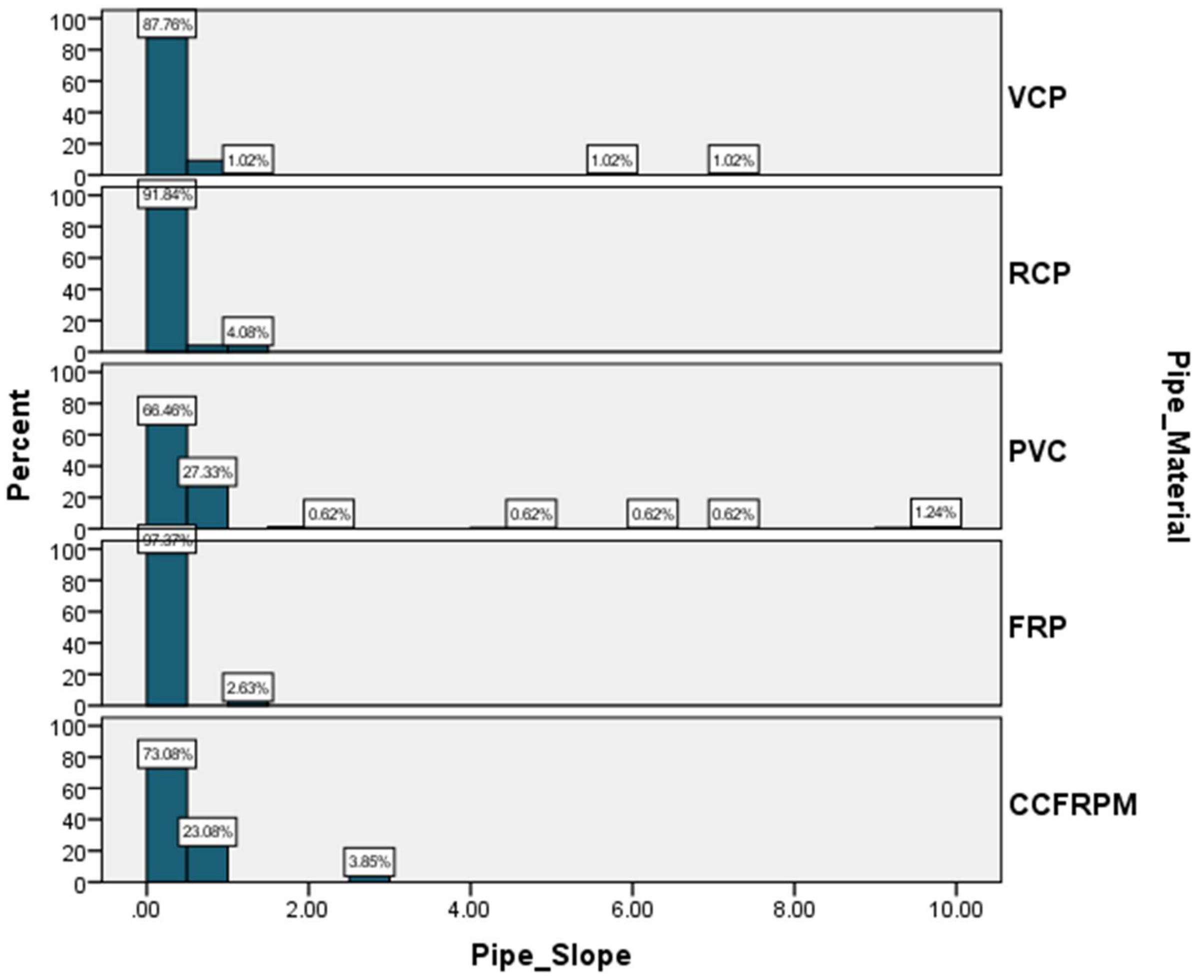

Figure 4.6: Distribution of pipe slope for different pipe materials

4.5 Length of sewer pipes

The distribution of length of sewer pipes is provided in the following figure. Based on this figure, it is realized that the mean value of length of sewer pipes is 358.77 ft, with standard deviation of 270.50 ft. Furthermore, the minimum and maximum values of the length of sewer pipe considered in this study are 7.6 ft and 1471.2 ft, respectively. It is further observed that the majority of the sewer pipes, i.e. 19.27%, have lengths between 100 ft and 200 ft.

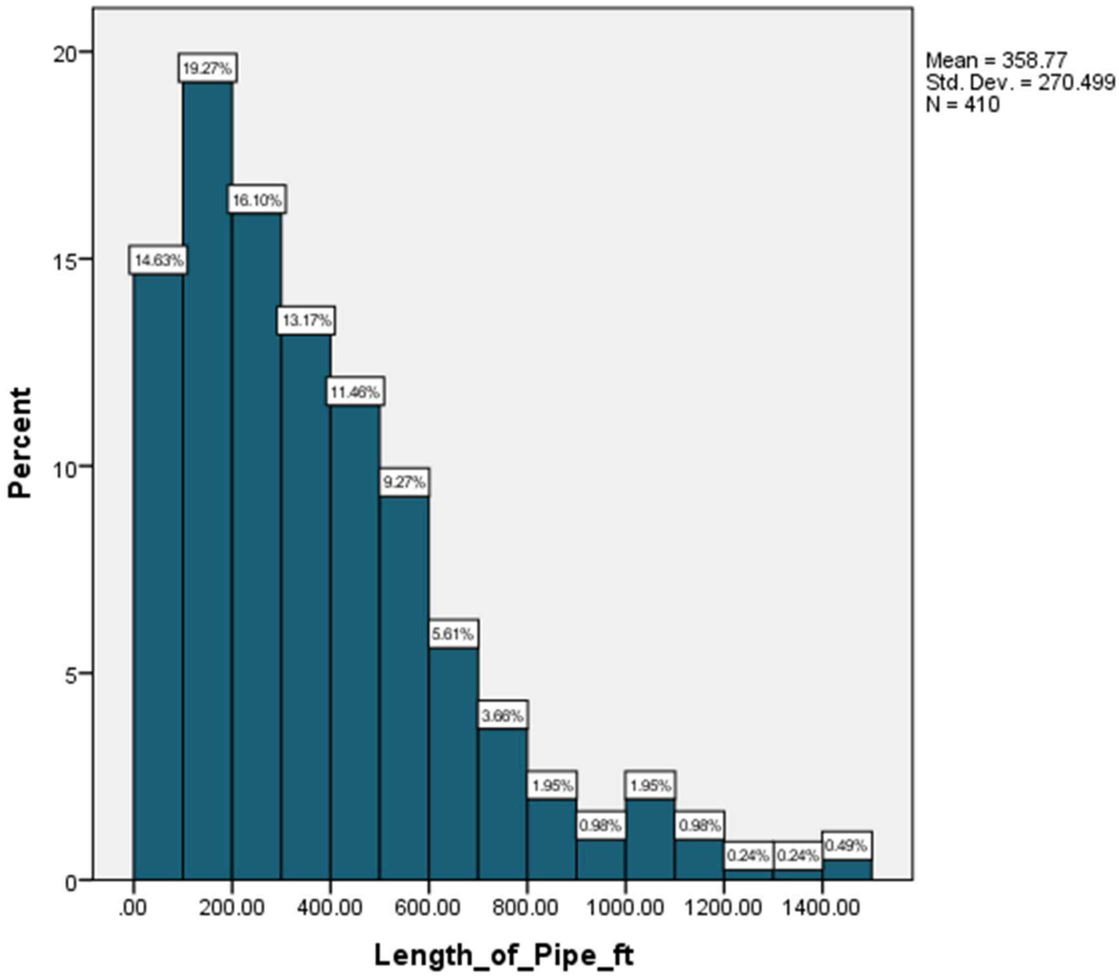

Figure 4.7: Distribution associated with length of pipe

The following figure illustrates distribution of lengths of sewer pipes for different pipe materials. Based on this figure, it is observed that PVC, RCP, and VCP have the three longest sewer pipes with 1471.2, 1451.9, and 1377.7 ft lengths for their pipe segments, respectively.

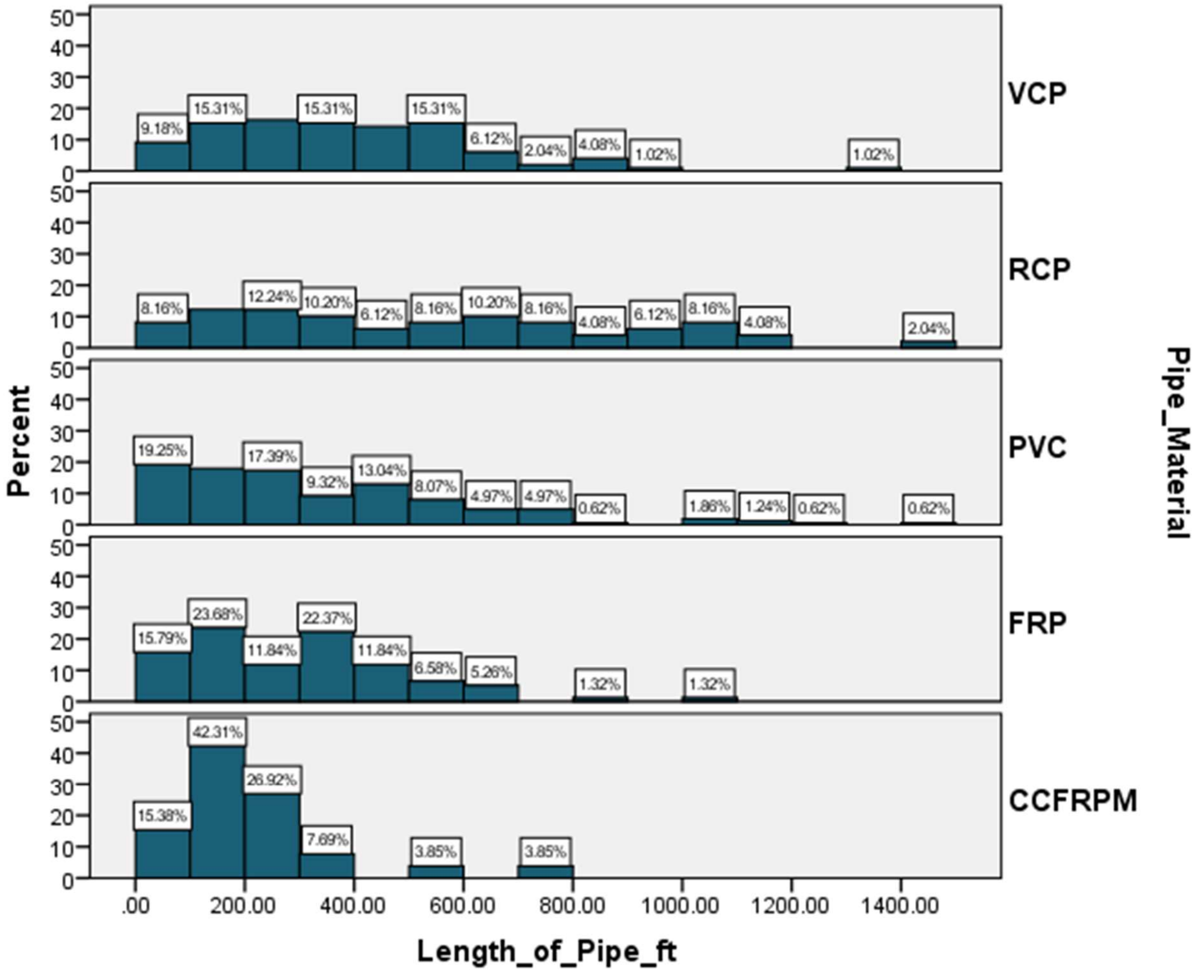

Figure 4.8: Distribution associated with length of pipe for different pipe materials

4.6 Average flow in sewer pipes (%full)

Considering the distribution of average flow in sewer pipes (normalized values of percentage of flow in pipes), it is observed that the mean value and the standard deviation of the average flow in sewer pipes are 24% and 10.8%, respectively. Furthermore, it is observed that the highest frequency of average flow in pipe is equal to 56.34% and this value is associated with average flow in pipe within the range of 20% up to 30%. Furthermore, it is realized that more than 87% and 95% of the sewer pipes have average flow values of less than 30% and 40%, respectively.

When considering the average flow in sewer pipes associated with each of the pipe materials, it is realized that the value of average flow in pipe for the majority of sewer pipes is less than 30%; moreover, it is observed that the highest average flows in sewer pipes occur in some of the vitrified clay pipes.

4.7 Average velocity of flow in sewer pipes

The distribution pertaining to average flow in sewer pipes, measured in units of feet per second, is presented in the following figure. Based on this figure, the highest frequency of average velocity of sewer flow in pipes is 25.37% and is between 3 ft/s and 3.5 ft/s. Furthermore, the mean value and the standard deviation of average flow velocity in pipes are 2.74 ft/s and 0.953 ft/s, respectively. Additionally, the values of average flow velocities for more than 80% of the sewer pipes are less than 3.5 ft/s.

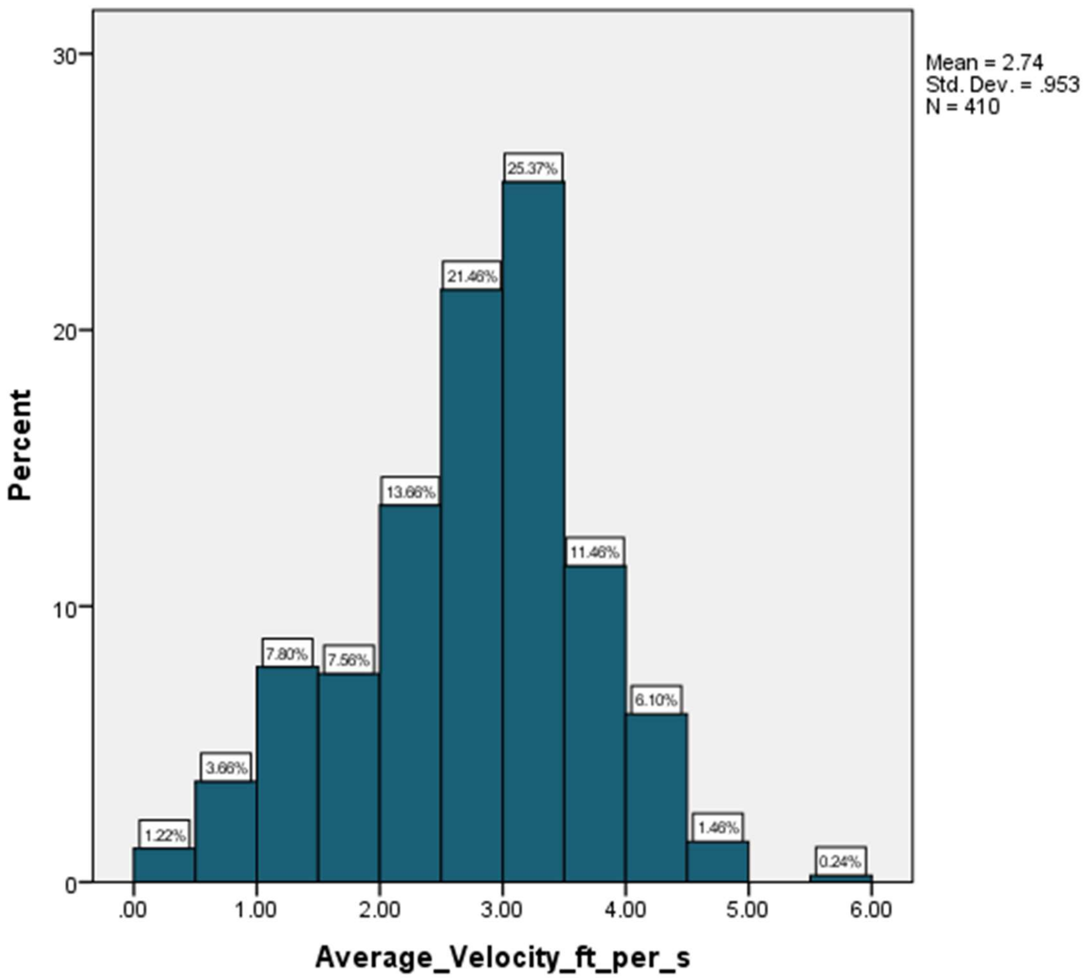

Figure 4.9: Distribution associated with average flow velocity

The figure below illustrates the average velocity of sewer flow for various pipe materials. It is observed that the maximum velocity of sewer flow (with the value of 5.83 ft/s) is associated with reinforced concrete pipes. It is further observed that based on the following flow velocity distributions, the average velocity of sewer flow in vitrified clay pipes is less than the values observed in other pipe materials.

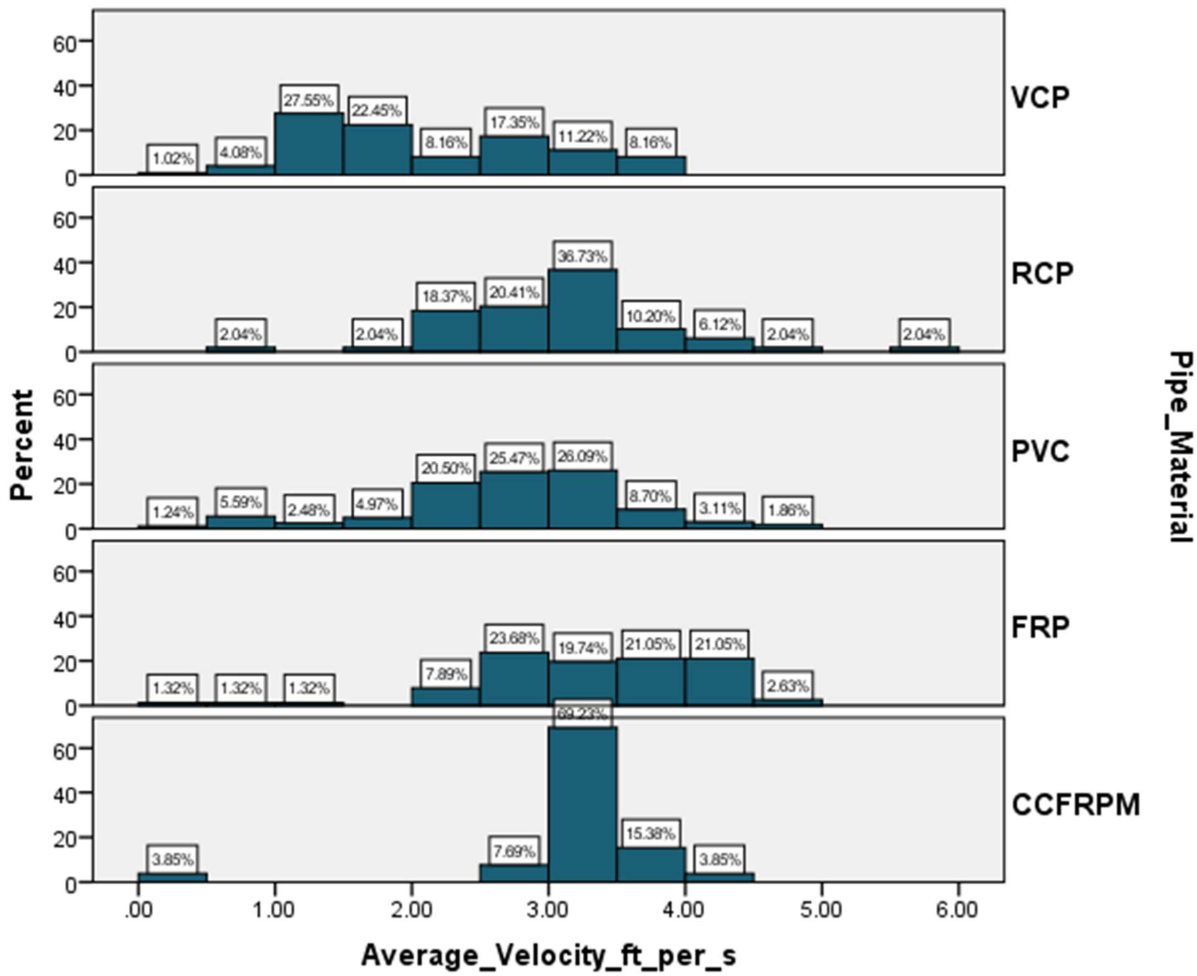

Figure 4.10: Distribution of average flow velocity for different pipe materials

4.8 Average flow depth in sewer pipes

The average flow depth distribution in sewer pipes is presented in the following figure. Based on this figure, the minimum and maximum values of average flow depths are 1.57 inches and 29.89 inches, respectively. Furthermore, based on the following figure, the values of mean and standard deviations of the distribution of average flow depth are 7.82 inches and 4.04 inches, respectively. The highest frequency of average flow depth is observed to be 35.12% and corresponds to the range within 5 inches and 7.5 inches of flow depth. Additionally, the following figure illustrates that more than 56% of sewer pipes have average flow depths of less than 7.5 inches, and about 79.5% of sewer pipes have average flow depths of less than 10 inches.

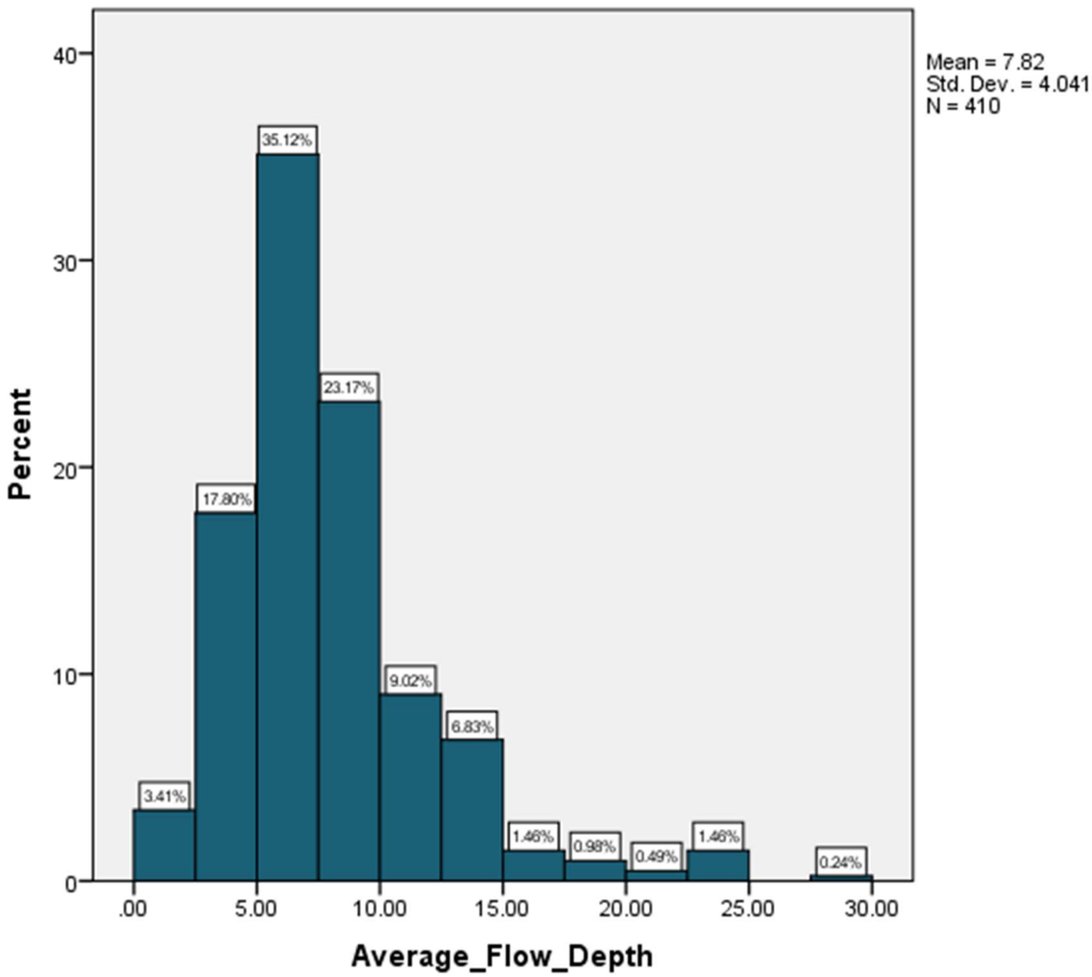

Figure 4.11: Distribution associated with average flow depth

Values of average flow depths for different pipe materials are presented in the following figure. Based on this figure, it is observed that the highest flow depth occurs in vitrified clay pipes. Furthermore, for CCFRPM and PVC pipes, the values of average flow depths are observed to be lower compared to other pipe materials. In the case of PVC pipes, more than 95% of the sewer pipes have average flow depths of less than 7.5 inches (when all pipe materials are considered, the corresponding percentage is 56.33%).

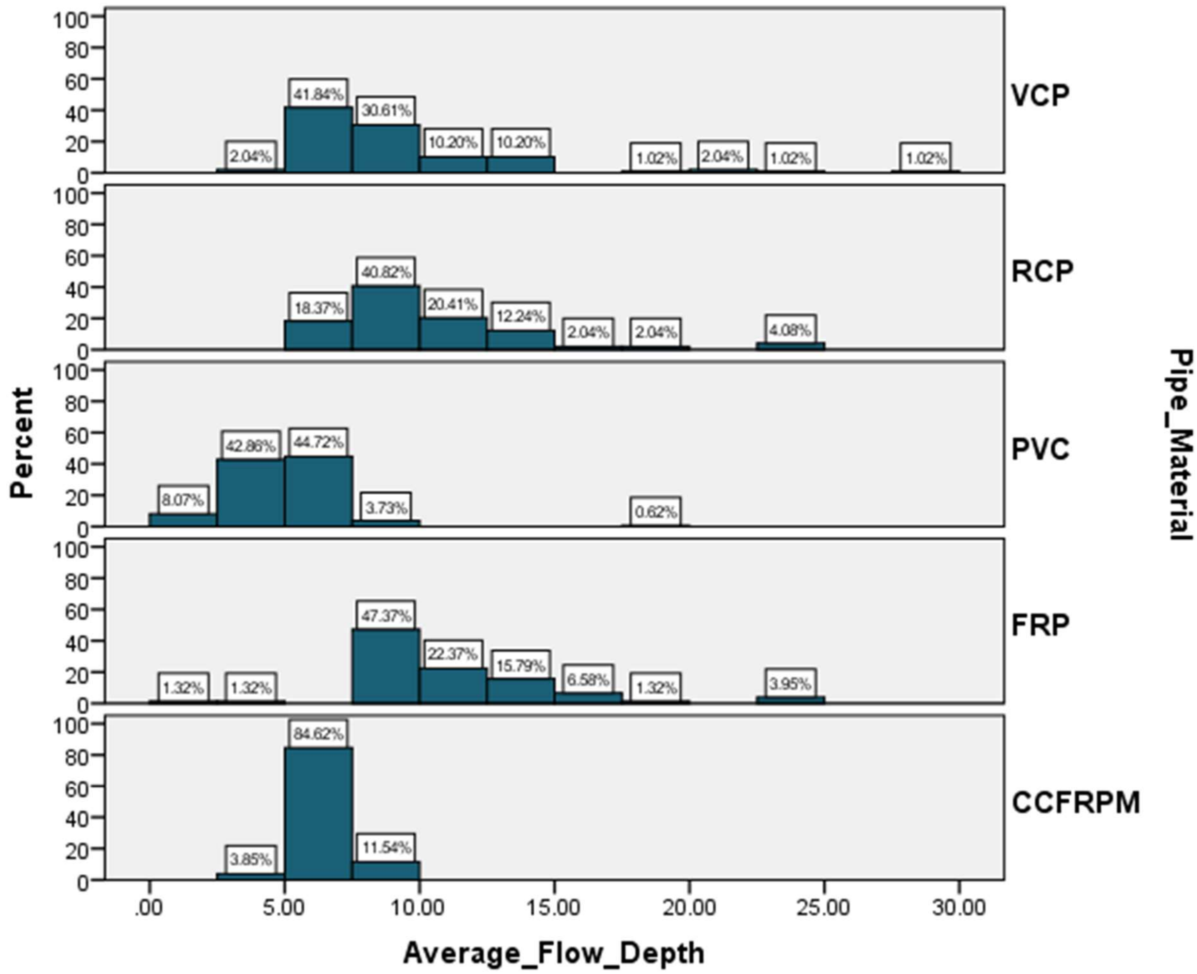

Figure 4.12: Distribution of average flow depth for different pipe materials

4.9 Operational condition grading of sewer pipes

The operational and maintenance (O&M) condition gradings of sewer pipes are presented in this section. Operational condition grading discussed in this dissertation refers to O&M condition grading. The operational condition grading is obtained by considering the highest operational condition grading within each of the individual sewer pipes; this value represents the worst O&M defect of the corresponding sewer pipe. Based on the following figure, it is observed that the majority of the sewer pipes have operational condition grading of 1 (59.27% of sewer pipes), and furthermore, 92.93% of the sewer pipes have operational condition gradings within ranges of 1 to 3. In other words, this figure illustrates that the majority of the sewer pipes are in good condition with respect to operational condition grading.

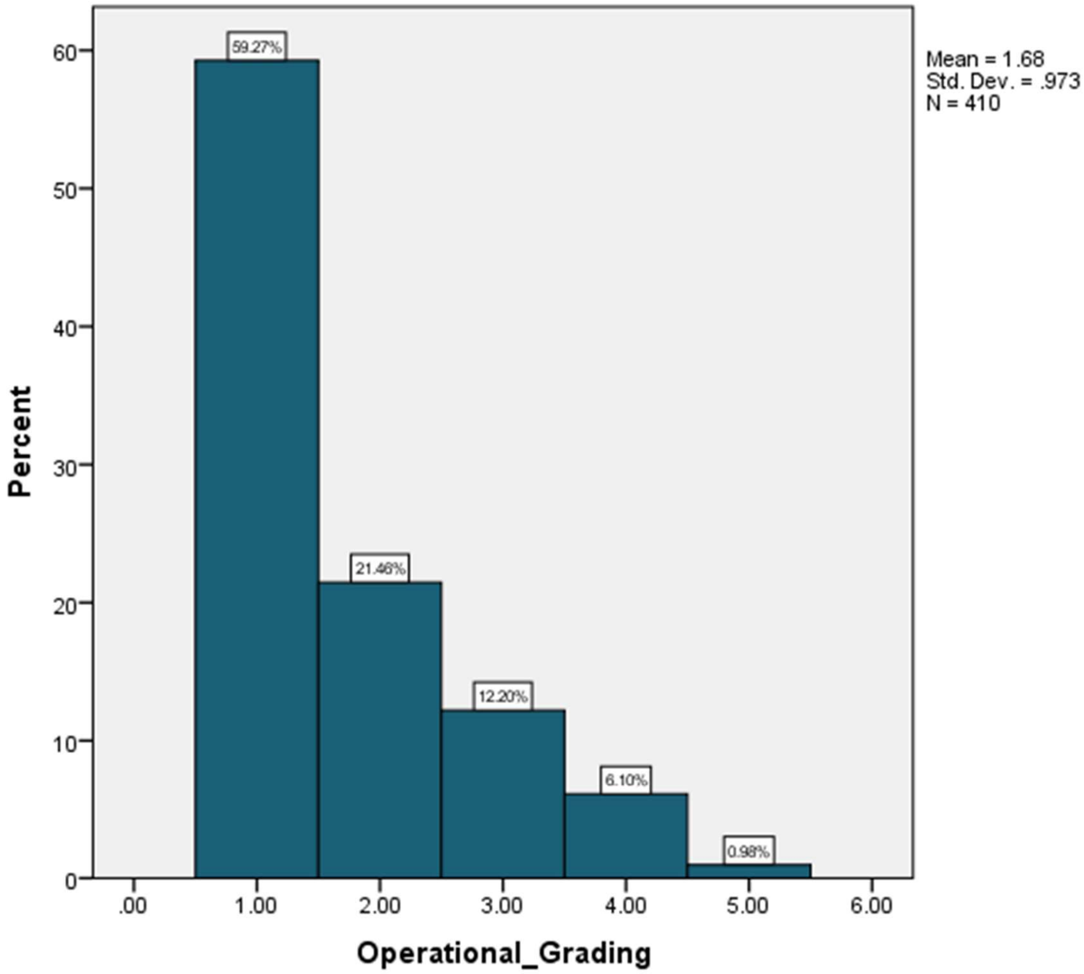

Figure 4.13: Distribution associated with operational (O&M) condition grading

Distributions of operational condition gradings for various pipe materials are presented in the following figure. Based on this figure, it is observed that reinforced concrete pipes and vitrified clay pipes have the highest percentages of the sewer pipes in operational condition gradings 4 and 5. It is further observed that all CCFRPM pipes considered in this study are in operational condition grading 1, and 90.79% of FRP pipes are in operational condition grading 1.

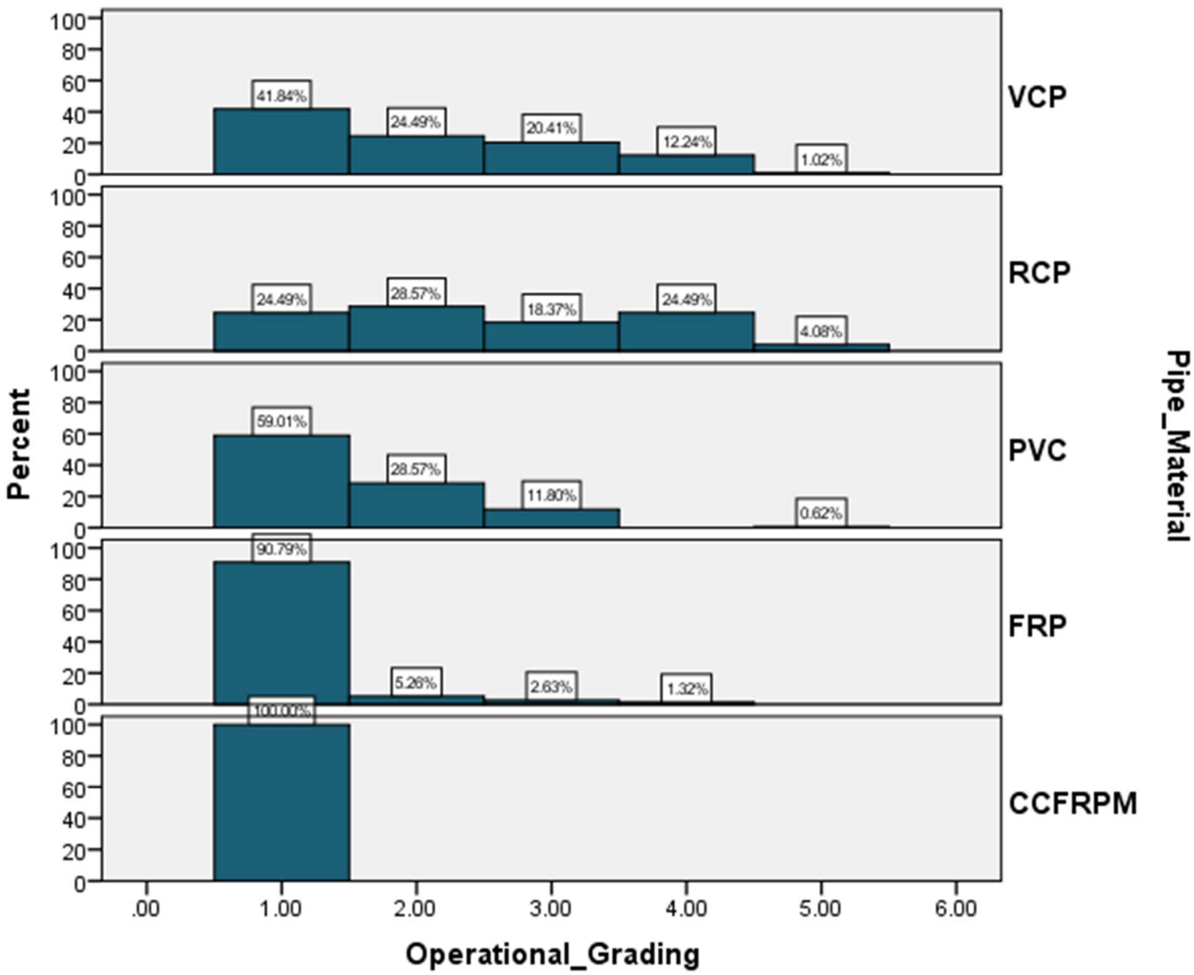

Figure 4.14: Distribution of operational (O&M) condition grading for different pipe materials

4.10 Structural condition grading of sewer pipes

The structural condition grading considered in this study is the highest structural condition grading associated with each sewer pipe, which demonstrates the worst structural defect within the sewer pipe. The following figure illustrates the distribution of structural condition gradings of sewer pipes. Based on this figure, the majority of sewer pipes (74.63% of sewer pipes) have structural condition grading of 1. Furthermore, 90.49% of the sewer pipes have structural condition gradings of 1 through 3. However, only 0.49% of the sewer pipes are in structural condition gradings 2.

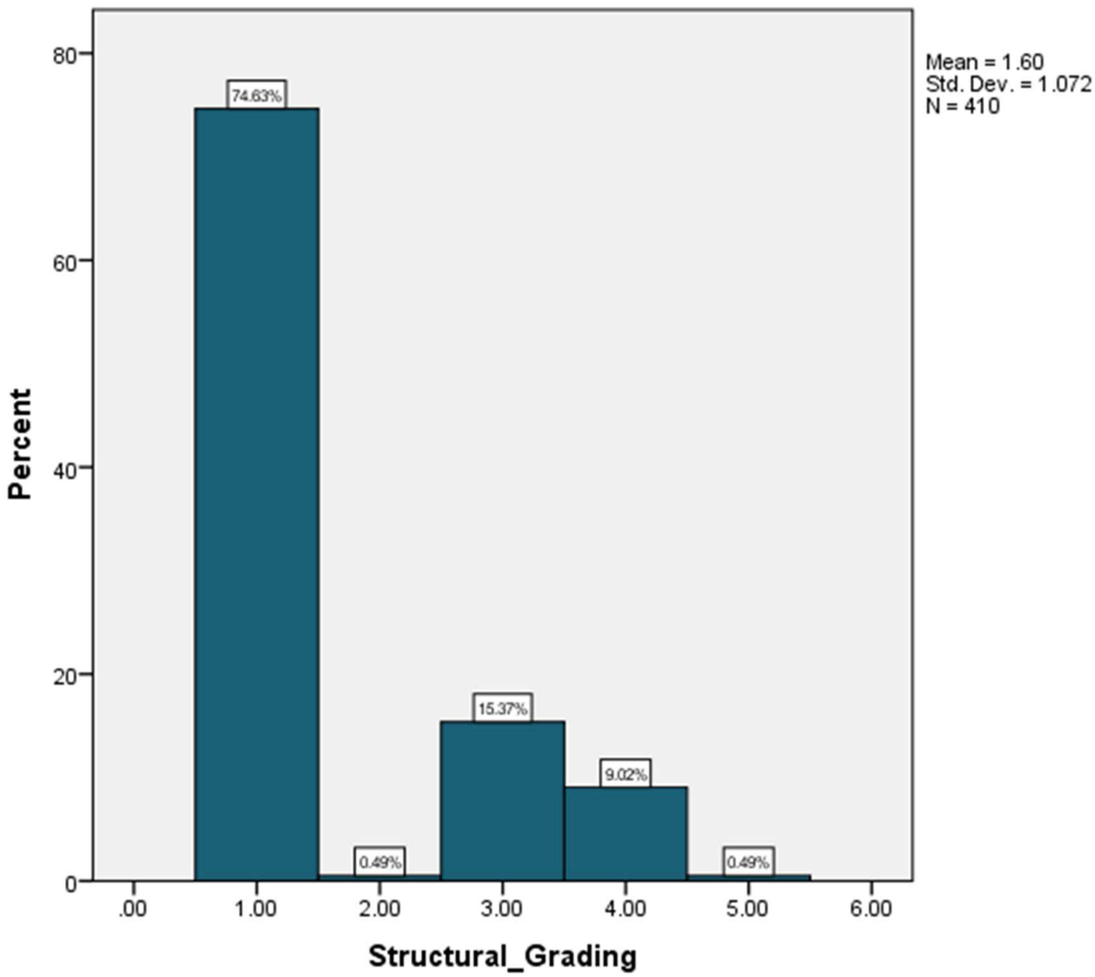

Figure 4.15: Distribution associated with structural condition grading

The structural condition grading for different pipe materials is presented in the below figure. Similar to the observation made in operational condition grading of sewer pipes, 100% of CCFRPM pipes are in structural condition grading 1. Moreover, it is observed that all the pipes in structural condition grading 5 are vitrified clay pipes. Additionally, it is realized that vitrified clay pipes and reinforced concrete pipes constitute the highest and the second highest percentages of pipes in structural condition grading 4 by 22.45% and

10.20% of each of the associated pipe materials being in this condition grading, respectively.

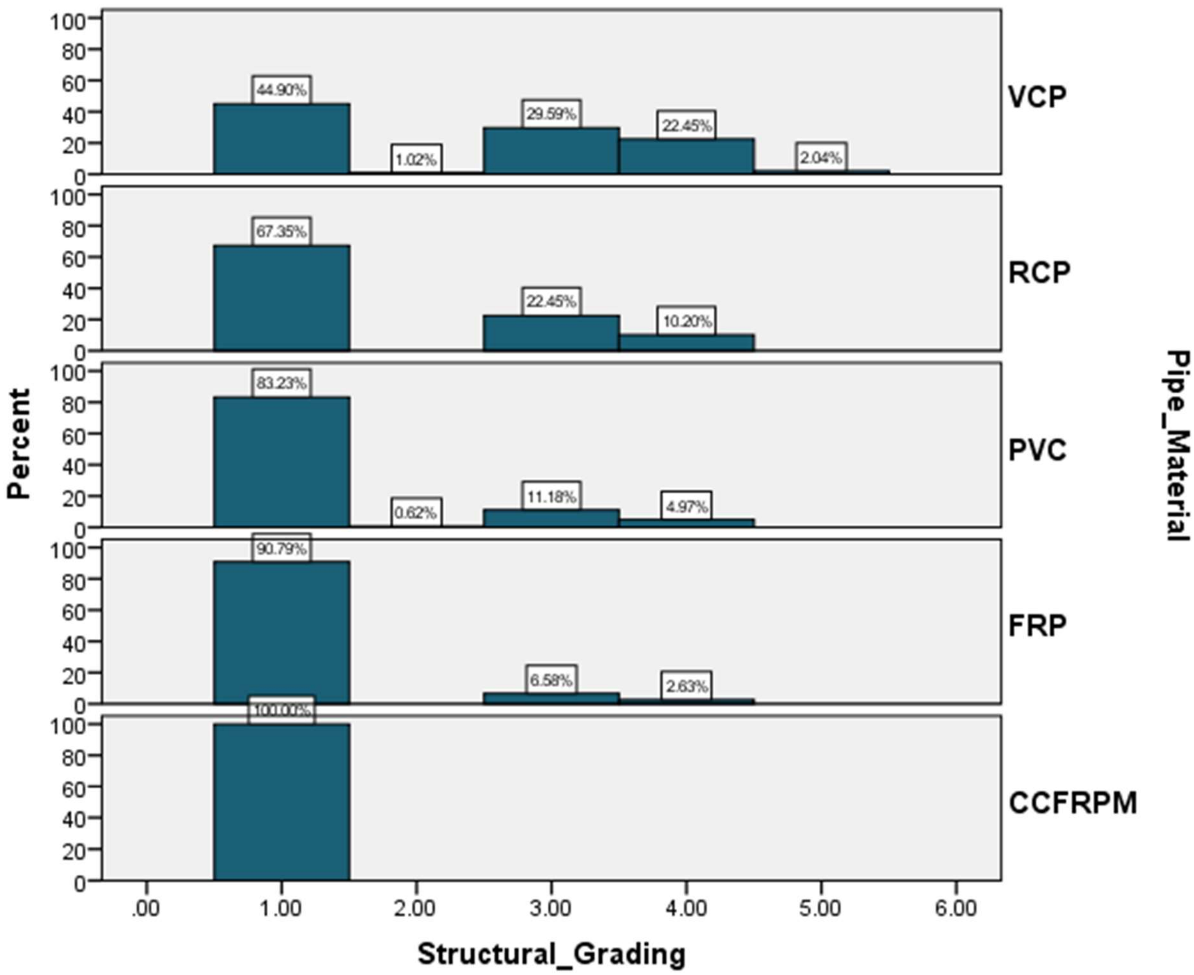

Figure 4.16: Distribution of structural condition grading for different pipe materials

4.11 Overall condition grading of sewer pipes

The overall condition grading of sewer pipes is considered to be the maximum value of operational (O&M) and structural condition gradings of sewer pipes. Therefore, based on the condition gradings provided earlier in operational and structural categories, the following figure illustrates the distribution of the overall condition grading of sewer pipes. Similar to previous observations, herein it is also realized that the highest frequency of pipes are in overall condition grading 1 (51.71% of pipes). The second highest frequency of pipes are in overall condition grading 3 with 19.02% of sewer pipes in this category. It is further observed that 84.14% of the sewer pipes have overall condition gradings of 1, 2 and 3.

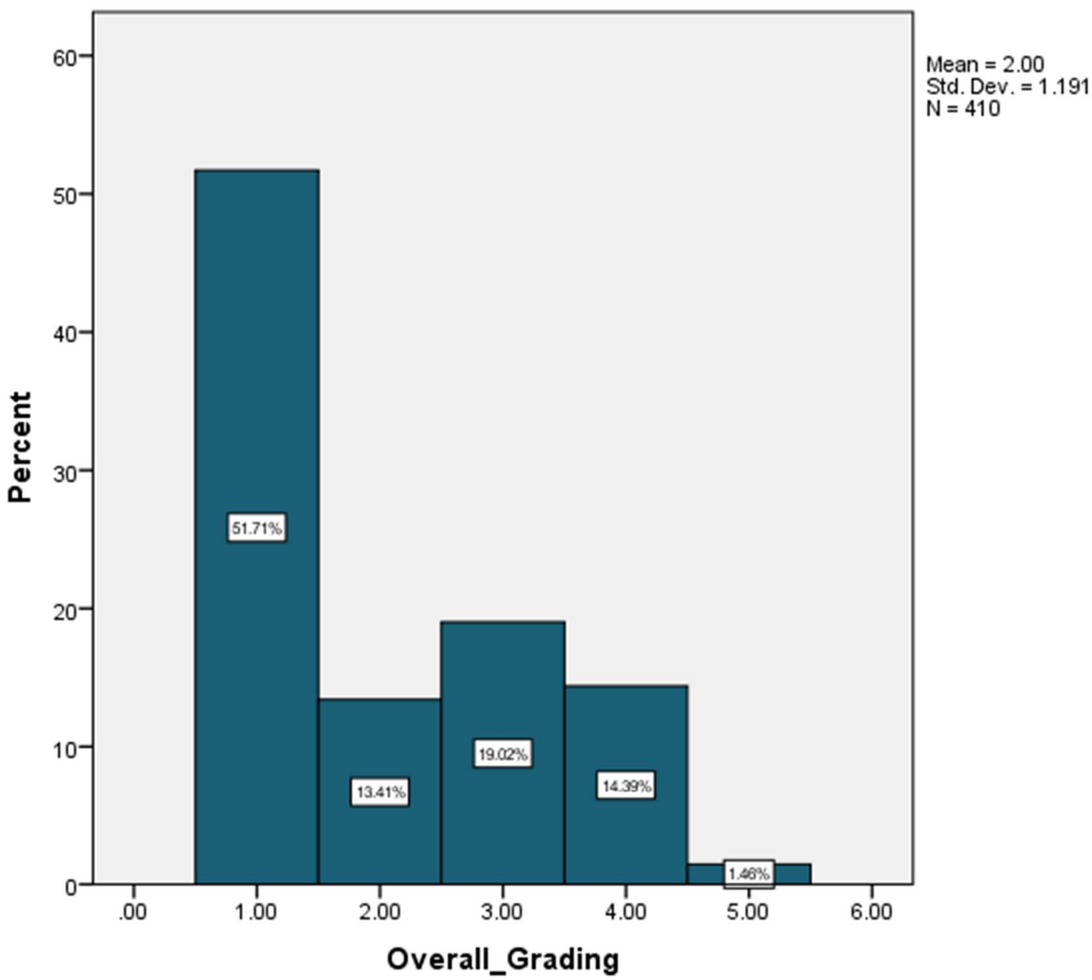

Figure 4.17: Distribution associated with overall condition grading

The overall condition grading of sewer pipes pertaining to different pipe materials is illustrated in the following figure. Similar to observations made in operational and structural condition gradings, herein it is also observed that all of the CCFRPM pipes are in overall condition grading 1. It is further observed that vitrified clay pipes and reinforced concrete pipes have the highest percentage of the overall condition gradings 4 and 5. Moreover, 96.05% of FRP pipes are within overall condition gradings 1 through 3 and there are no overall condition gradings 5 associated with FRP pipes.

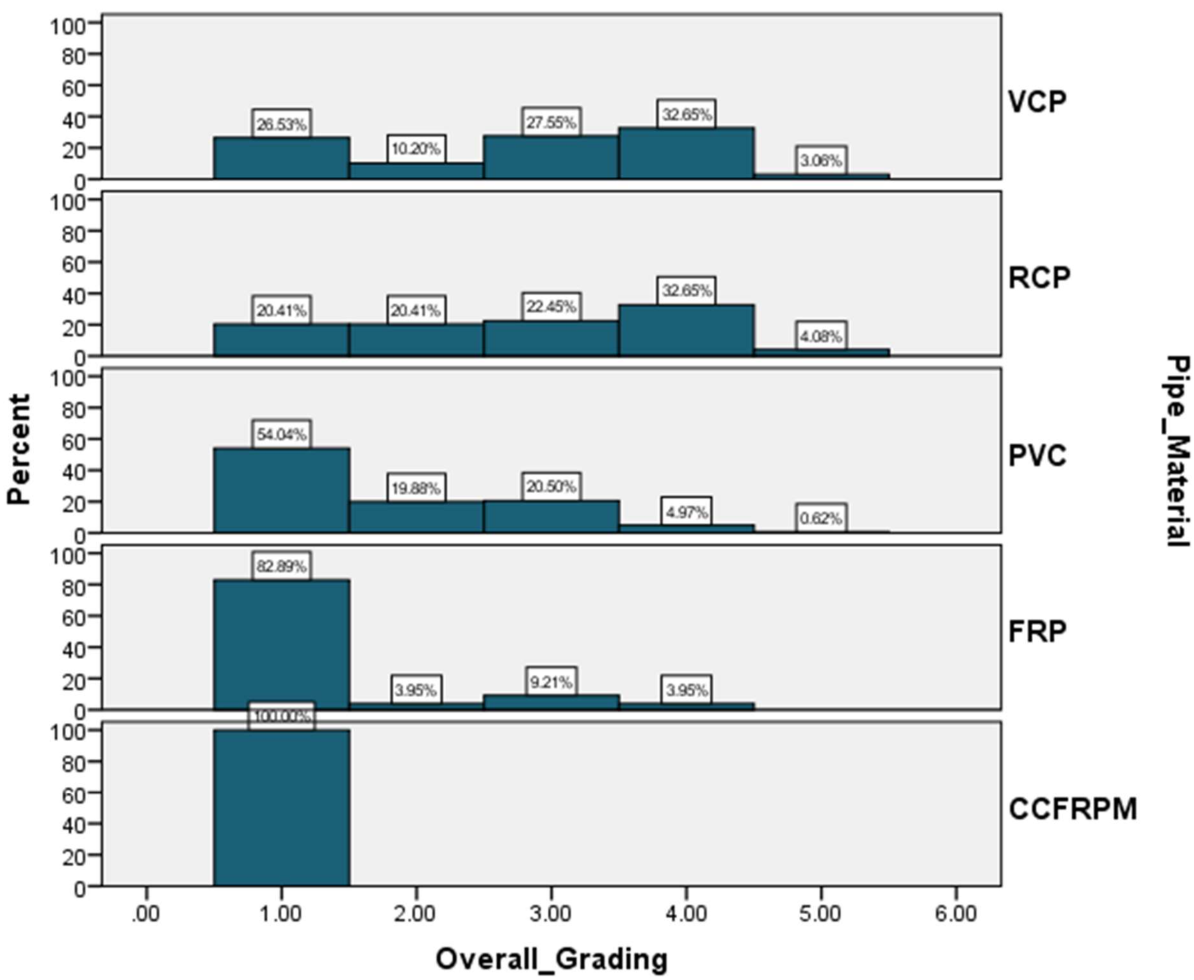

Figure 4.18: Distribution of overall condition grading for different pipe materials

4.12 Pipe materials of sewer pipes

The pipe materials and their associated percentage are presented in the following figure. There are five different pipe materials in the sewer pipe data set: PVC, VCP, RCP, CCFRPM, and FRP

The majority of the sewer pipes considered in this study are PVC pipes which constitute 39.27% of the sewer pipes. The second highest frequency of pipe materials pertains to vitrified clay pipes with frequency of 23.90%. Furthermore, it is observed that the lowest frequency of pipe material is associated with CCFRPM pipe with only 6.34% frequency.

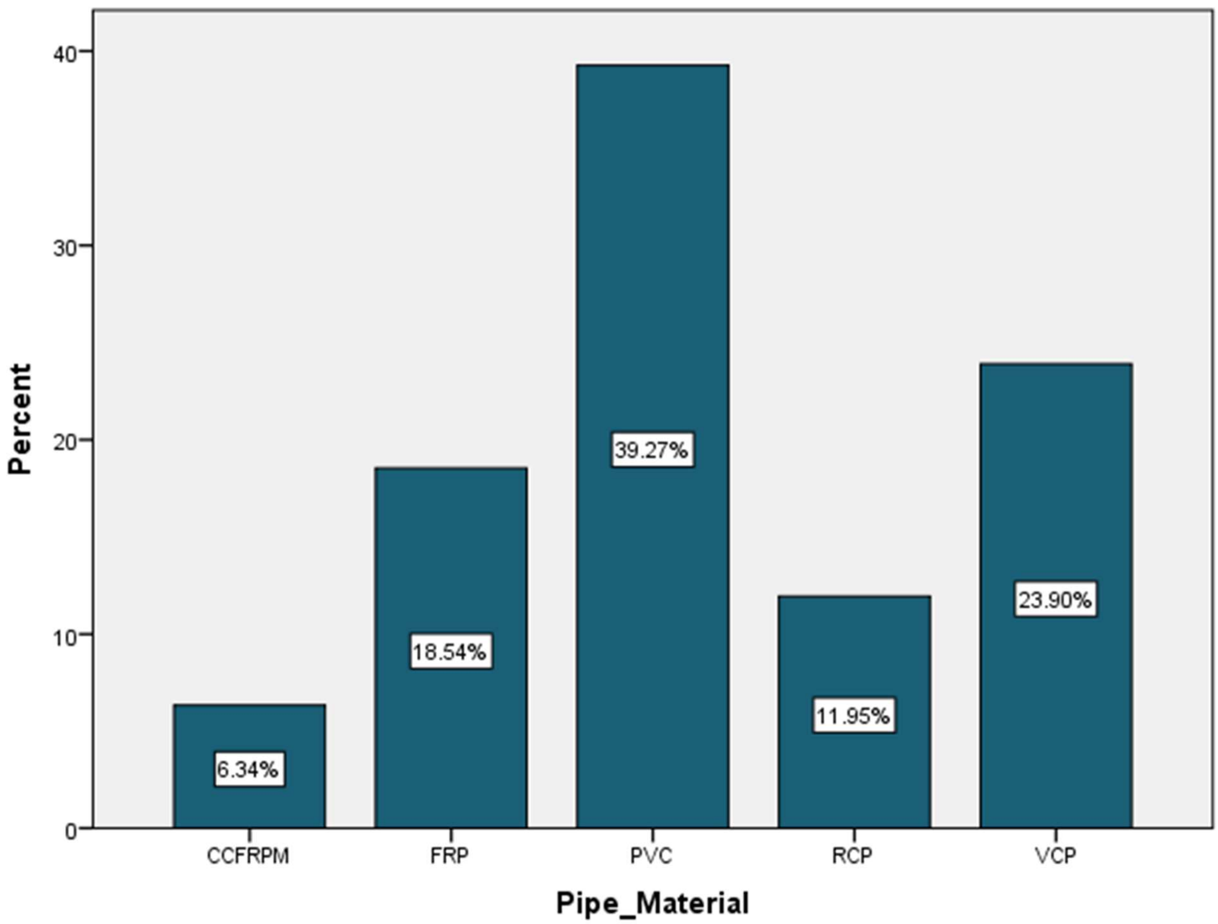

Figure 4.19: Distribution associated with various pipe materials

4.13 Binary operational condition grading of sewer pipes

Binary operational (O&M) condition gradings of sewer pipes are obtained based on the previously stated operational condition grading of pipes. This section contains the distribution associated with O&M binary condition gradings. Based on the following figure, it is observed that 92.93% of the sewer pipes are in condition grading 0 and only 7.07% of pipes are in condition grading 1. In this dissertation, it is assumed that operational condition gradings 1, 2, and 3 denote that pipes do not show severe problems, and therefore, it is assumed that they belong to binary condition grading 0; in other words, for these pipes, it is assumed that the failure criterion is not met and they have not failed. Moreover, sewer pipes which are in operational condition gradings 4 and 5, are assumed to have failed. It should be noted that these assumptions are solely for the purpose of illustration and can therefore be subjected to changes based on the decision makers priorities and criteria for defining failure of sewer pipes.

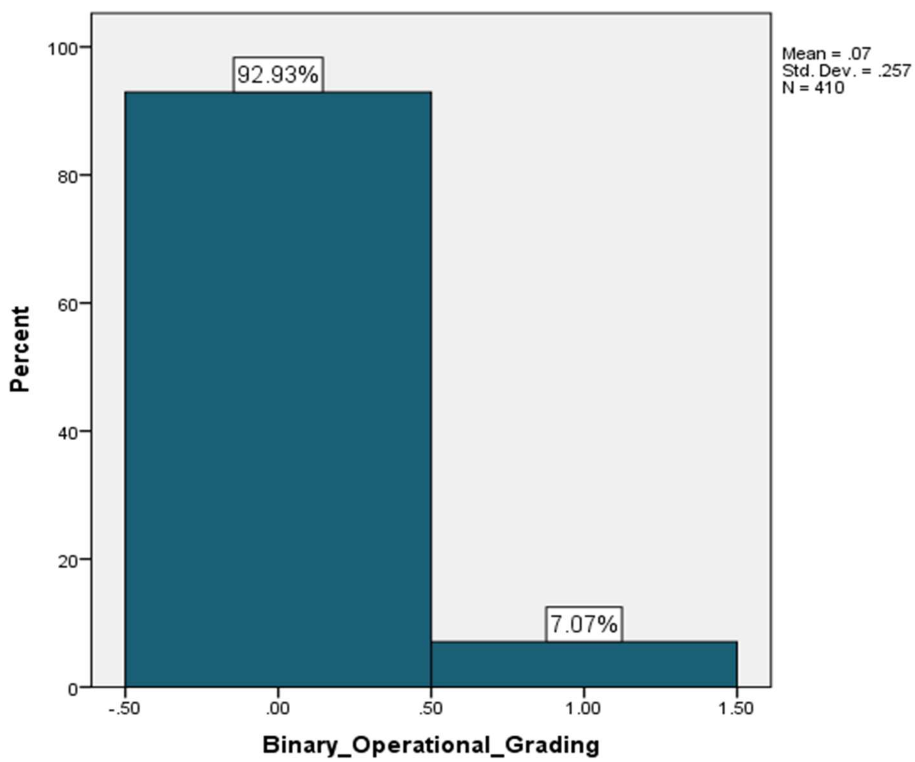

Figure 4.20: Distribution associated with binary operational condition grading

Furthermore, the following figure shows the details of distribution of binary operational gradings of sewer pipes with respect to various pipe materials. Based on this figure, reinforced concrete pipes have the highest percentage of pipes in binary operational condition grading 1 with 28.57% of RCP pipes in this condition state; moreover, vitrified clay pipes have the second highest percentage of binary operational condition 1 with 13.27%. Similar to previous observations, all pipes in CCFRPM category are in binary operational condition state 0.

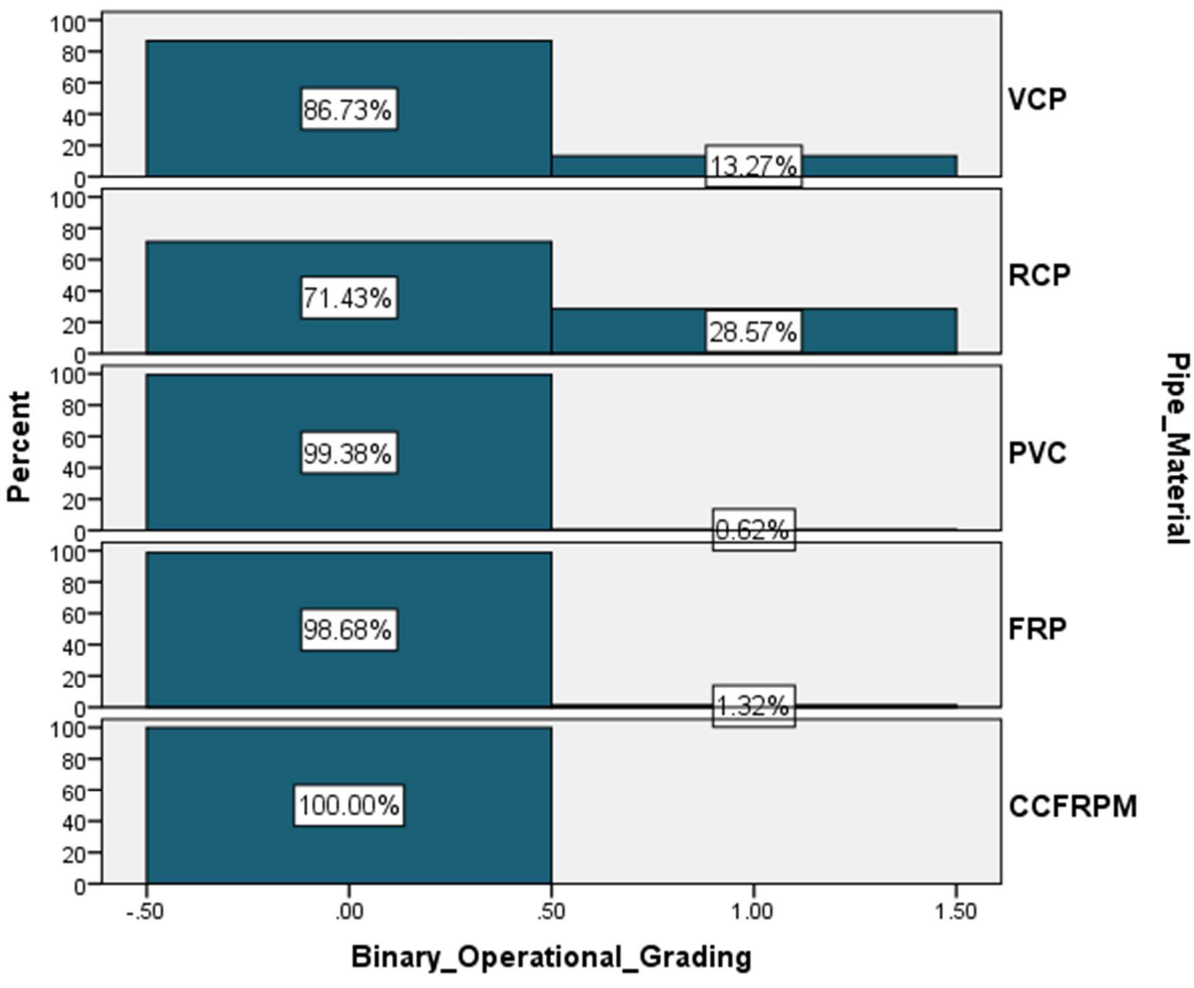

Figure 4.21: Distribution of binary operational condition grading for different pipe materials

4.14 Binary structural condition grading of sewer pipes

Similar to the binary operational condition gradings of sewer pipes discussed earlier, binary structural condition gradings of pipes are also obtained using structural condition grading of sewer pipes. The following figure illustrates the distribution associated with binary structural condition grading of pipes. The same assumption utilized for categorizing pipes in binary operational condition grading of pipes is applied here as well. In other words, pipes in structural condition states 4 and 5 are categorized as binary structural grading 1 and if the structural condition states of pipes are less than 4, they are categorized in binary structural condition 0. As stated earlier, these assumptions are utilized for illustration purposes and can be altered based on decision maker's priorities and defining various failure criteria. The following figure shows that the majority of the sewer pipes, i.e.90.49% of assets, are in binary structural condition grading 0.

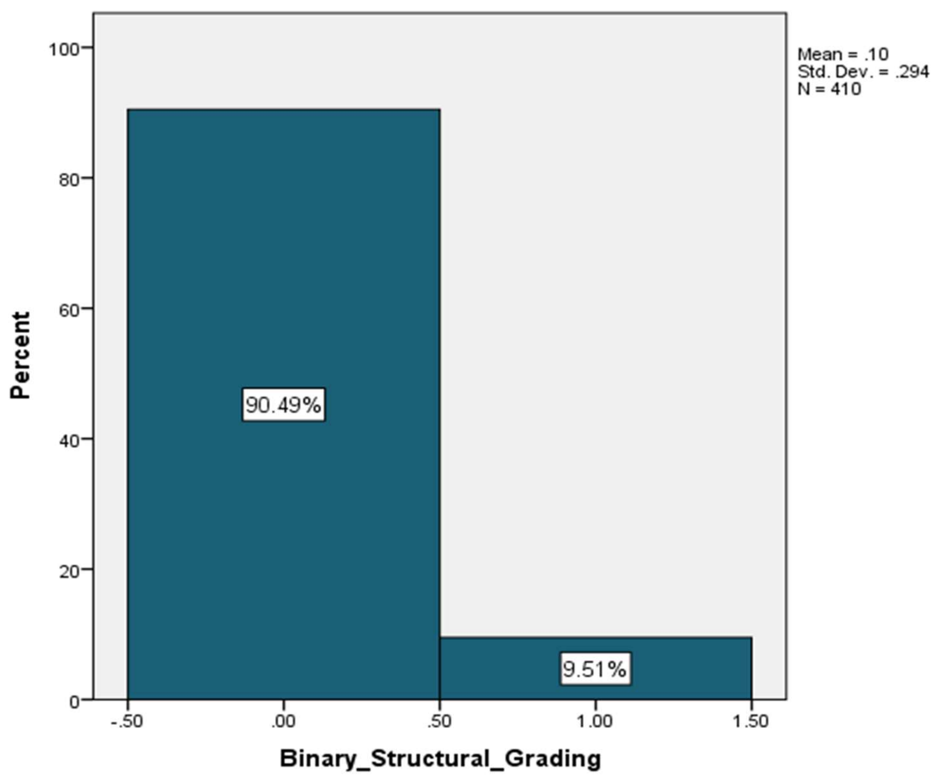

Figure 4.22: Distribution associated with binary structural condition grading

The following figure shows the percentage of sewer pipes in binary structural condition gradings 0 and 1 for various pipe materials. Similar to binary operational condition gradings of pipes, vitrified clay pipes and reinforced concrete pipes have the highest percentage of pipes in binary structural grading 1 and all CCFRPM pipes are in binary structural state 0; however, unlike binary operational grading, vitrified clay pipes have the highest percentage of pipes in binary structural condition state 1, with 24.49% of pipes in this condition state. The second highest frequency of pipes in binary structural grading 1 belongs to reinforced concrete pipes with 10.20% of sewer pipes in this condition grading.

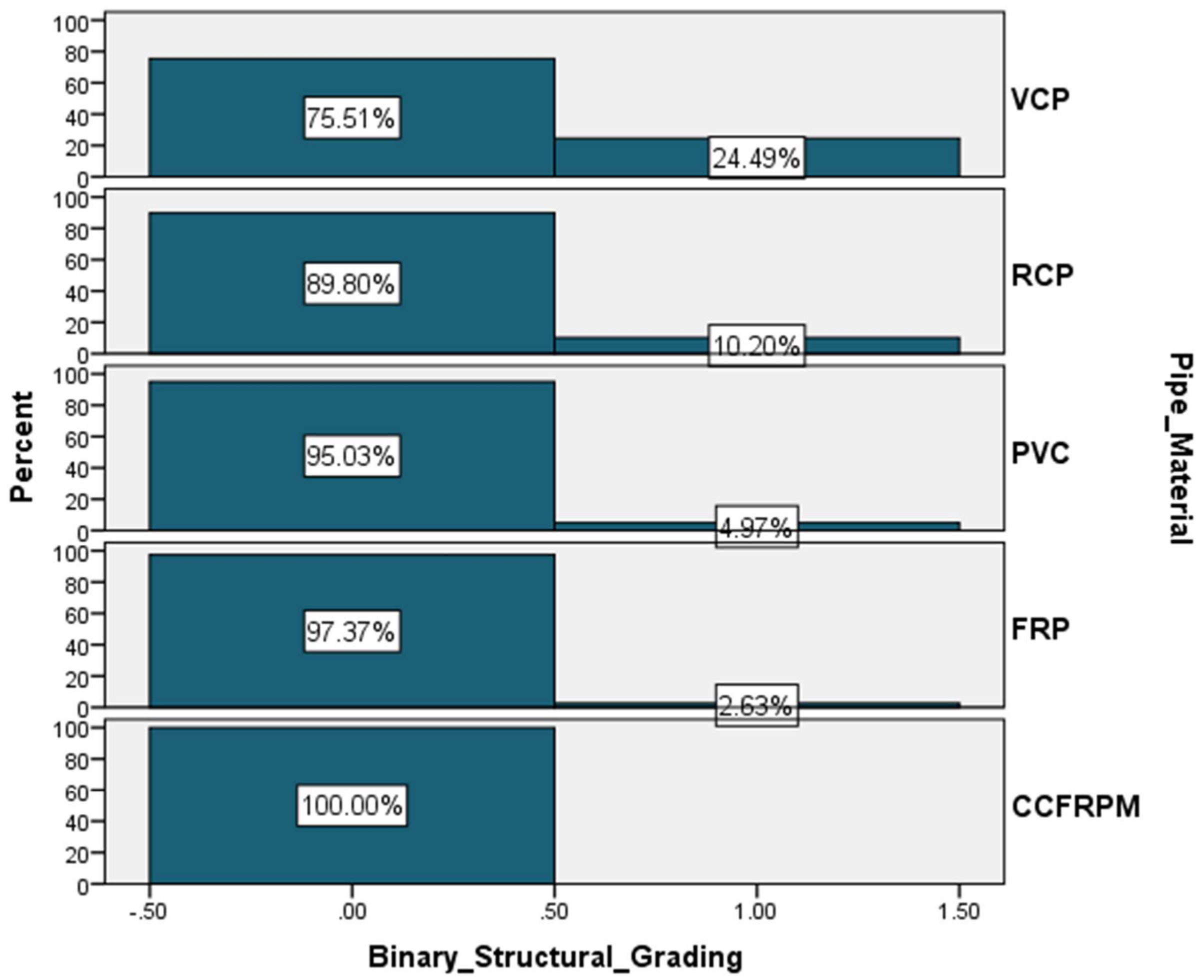

Figure 4.23: Distribution of binary structural condition grading for different pipe materials

4.15 Binary overall condition grading of sewer pipes

Binary overall condition gradings of pipes are achieved based on overall condition gradings of sewer pipes. The designation of condition states to binary condition gradings are similar to binary operational and structural condition grading of pipes. If overall condition gradings are less than 4, the sewer pipes are assumed to be in binary condition state 0; otherwise, they are assumed to be in binary overall condition state 1. As stated earlier, these designations may be subjected to changes based on different failure criteria. Based on the below figure, it is observed that the majority of sewer pipes are in binary overall condition state 0 with 84.15% of pipes in this condition state.

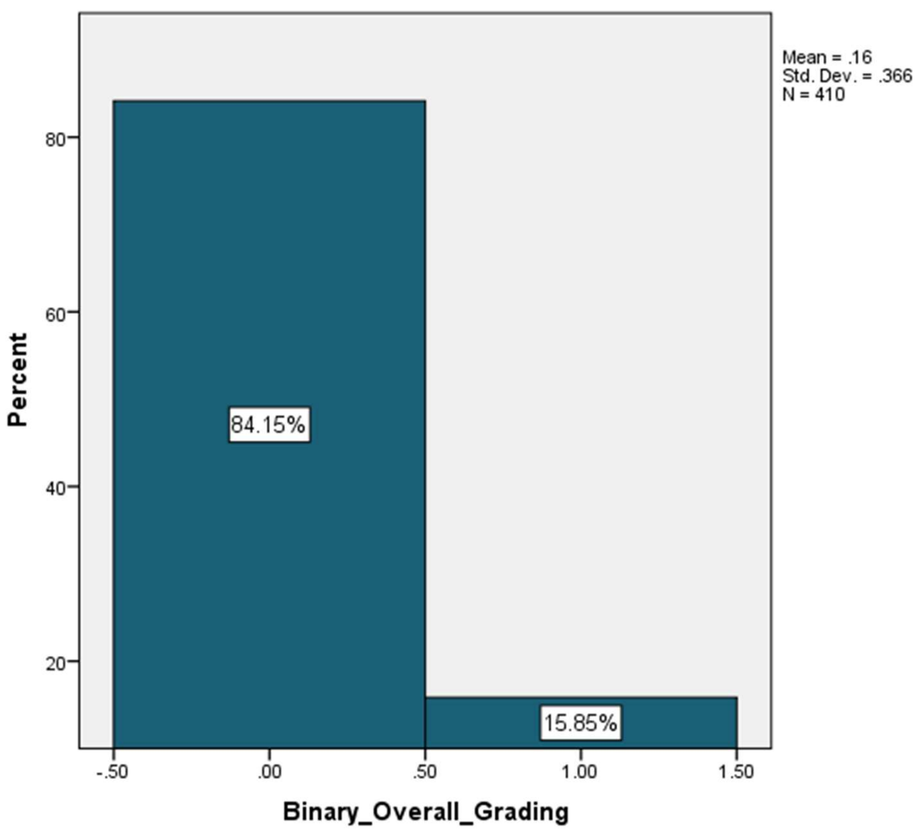

Figure 4.24: Distribution associated with binary overall condition grading

The following figure contains the shows the percentages of sewer pipes in binary overall gradings 0 and 1 with respect to various materials used in pipes. Similar to results from structural and operational binary gradings, observations made based on the following figure demonstrate that reinforced concrete pipes and vitrified clay pipes have the highest and the second highest percentages of sewer pipes in binary overall condition grading 1 with 36.73% and 35.71% of their associated sewer pipes in this condition state, respectively. Additionally, it is observed that none of the CCFRPM pipes are in binary overall condition state 1.

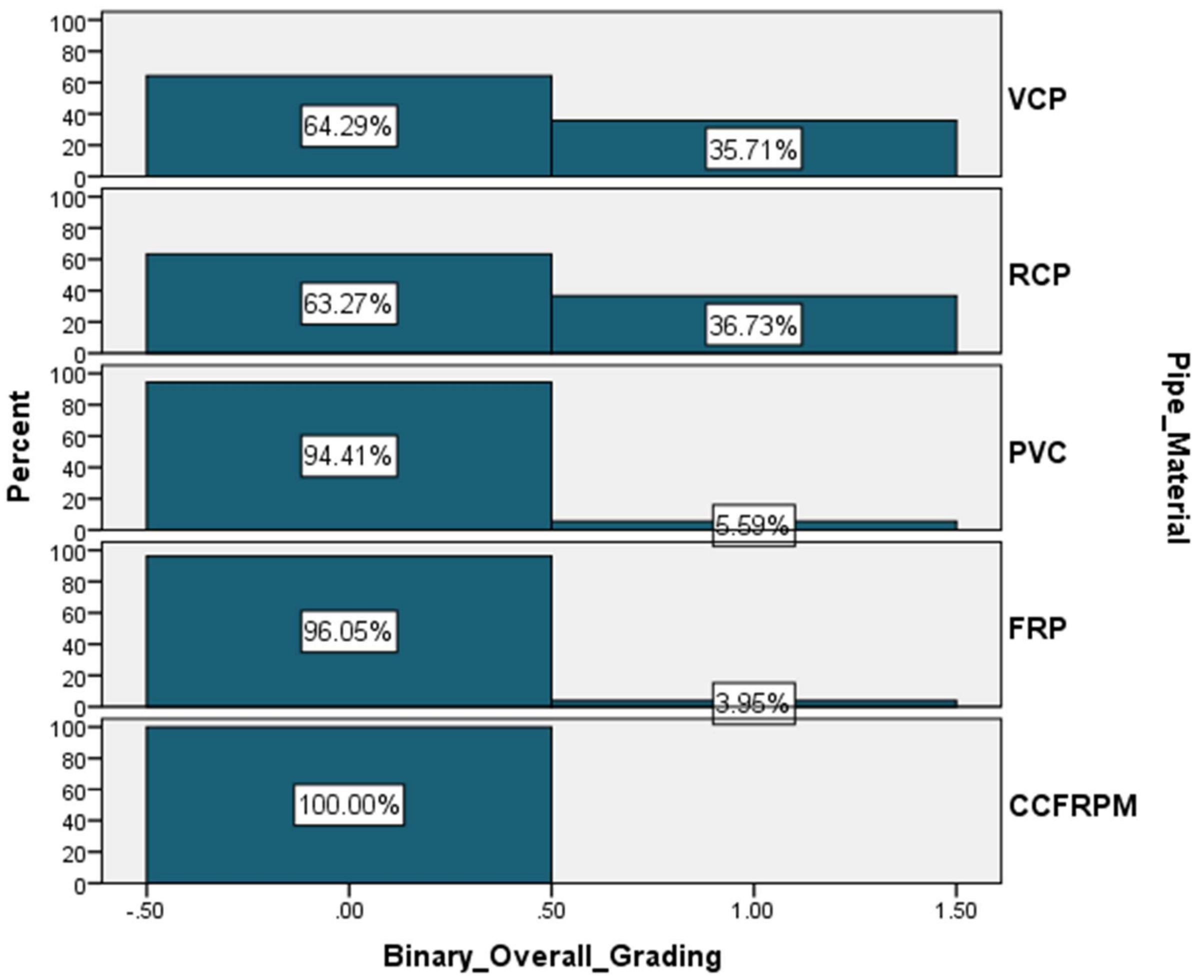

Figure 4.25: Distribution of binary overall condition grading for different pipe materials

4.16 Various condition gradings of sewer pipes with respect to age of assets

In this section, operational (O&M), structural, and overall condition gradings of sewer pipes with respect to age of pipes are examined. Based on the following figures, it is observed that for all condition grading categories, as the age of assets increases, the percentage of pipes in categories 4 and 5 emerge. Furthermore, with regards to the structural condition grading of sewer pipes, it is observed that for assets with younger ages, the proportion of pipes in condition gradings 1 and 2 are greater compared to sewer pipes with higher ages. This trend is observed for condition state 2 in operational and overall categories as well. These graphs are illustrative of the effect of age of sewer pipes in their corresponding condition gradings. Based on these graphs, it is realized that when compared to younger pipes, older pipes have transitioned to higher condition gradings. However, in various cases throughout these graphs sudden drops and increases are observed with regards to condition states of pipes. These observations are due to the fact that the pipes presented in these graphs have different properties, and therefore, the effect of all independent variables need to be incorporated in order to investigate the effect of age in condition grading of sewer pipes. This task is performed through developing various modeling approaches in this dissertation.

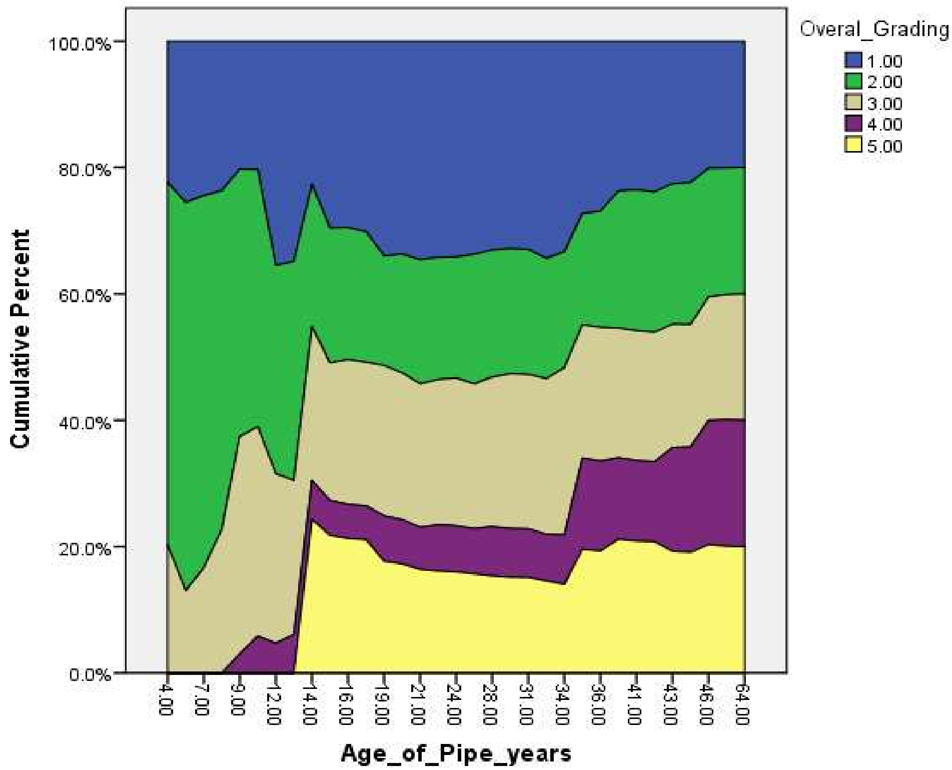

Figure 4.26: Overall condition grading of assets with respect to age of assets

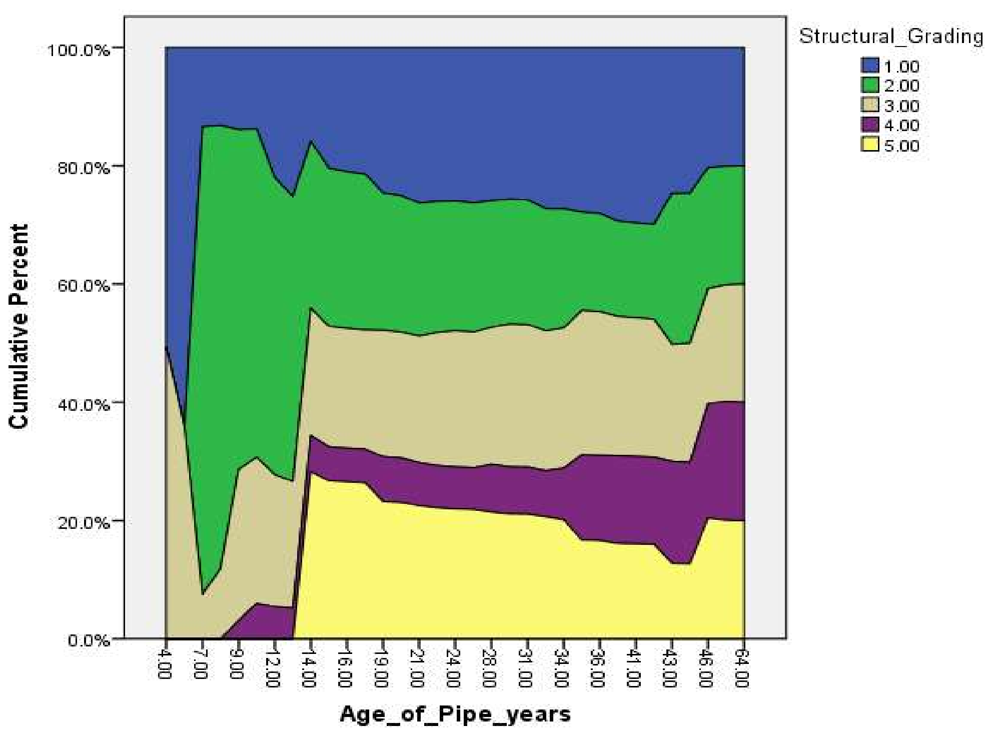

Figure 4.27: Structural condition grading of assets with respect to age of assets

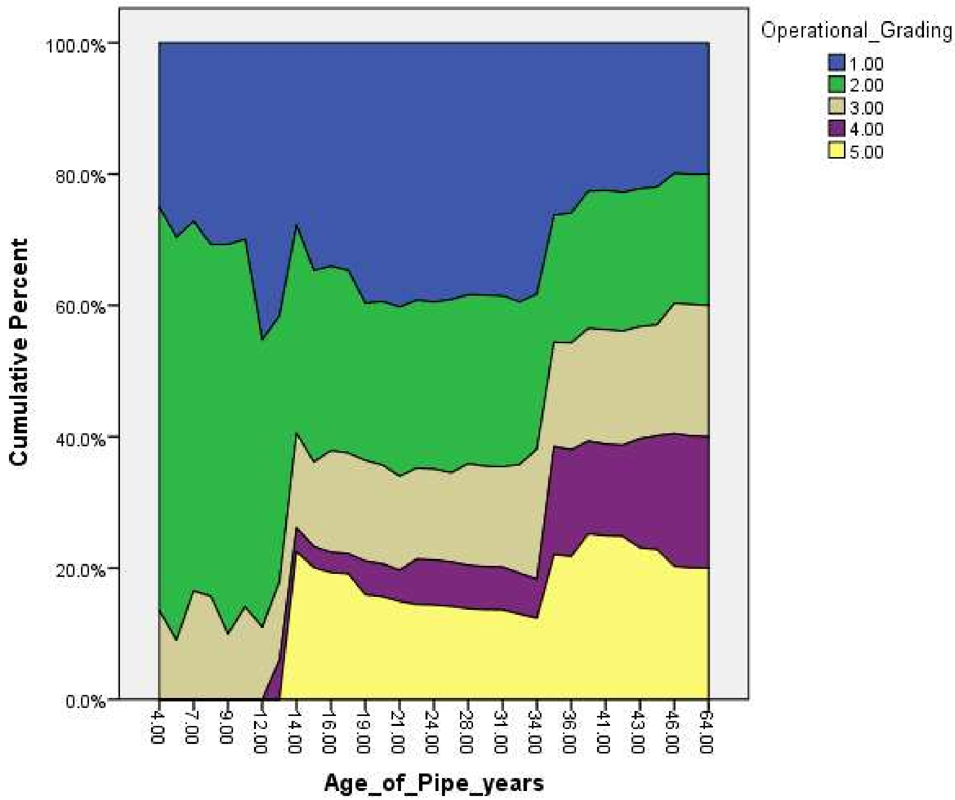

Figure 4.28: Operational (O&M) condition grading of assets with respect to age of assets

4.17 Binary condition gradings of sewer pipes with respect to age of assets

Binary condition gradings of sewer pipes in operational (O&M), structural, and overall categories and with respect to age of assets are presented in the following figures. Based on these figures, it is observed that for all condition grading categories, when the age of sewer pipes increase, the condition gradings transition from binary condition state 0 to condition state 1; therefore, this illustrates the deterioration of sewer pipes due to aging of pipes. Additionally, in these graphs, it is realized that the percentage of sewer pipes in binary condition grading 1 associated with different grading categories are increasing as the age of assets increase, and on the other hand, the percentage of pipes designated with condition state 0 in various binary condition gradings are declining as well. It is further observed that based on these graphs, when considering binary structural condition states, sewer pipes transition to binary condition state 1 in earlier ages compared to binary operational grading. Therefore, it is realized that compared to sever operational (O&M) defects, severe structural defects occur earlier in these sewer pipes. It should be noted that the assumption utilized in designation of binary condition states are the same stated earlier, and can be changed based on decision maker's priorities and failure criteria.

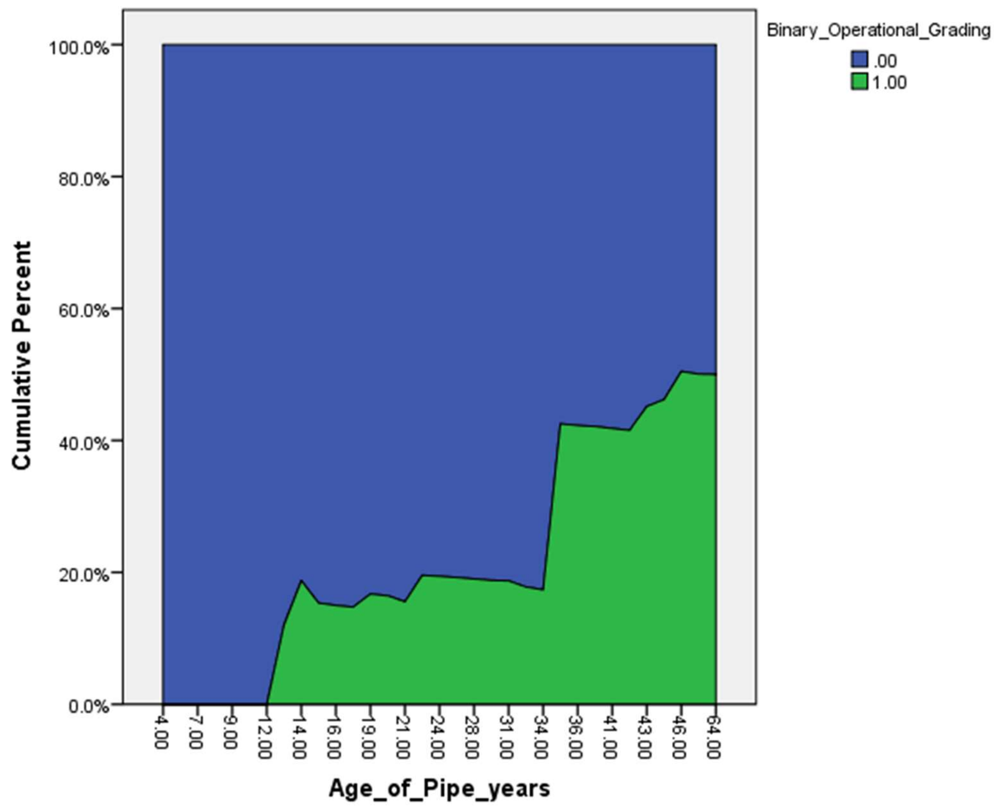

Figure 4.29: Binary operational condition grading of assets with respect to age of assets

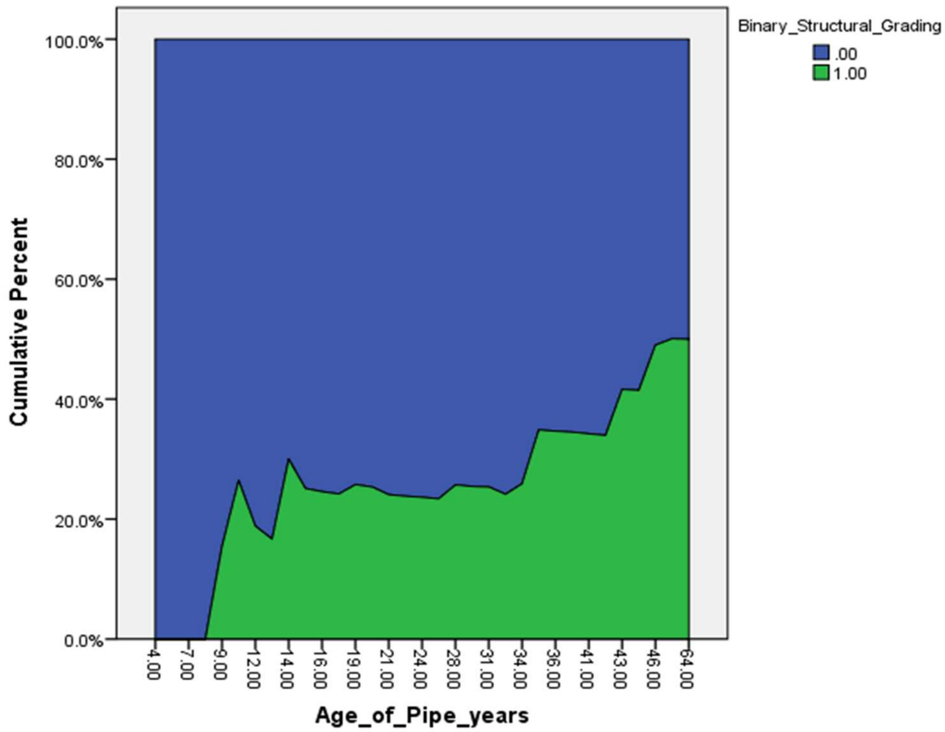

Figure 4.30: Binary structural condition grading of assets with respect to age of assets

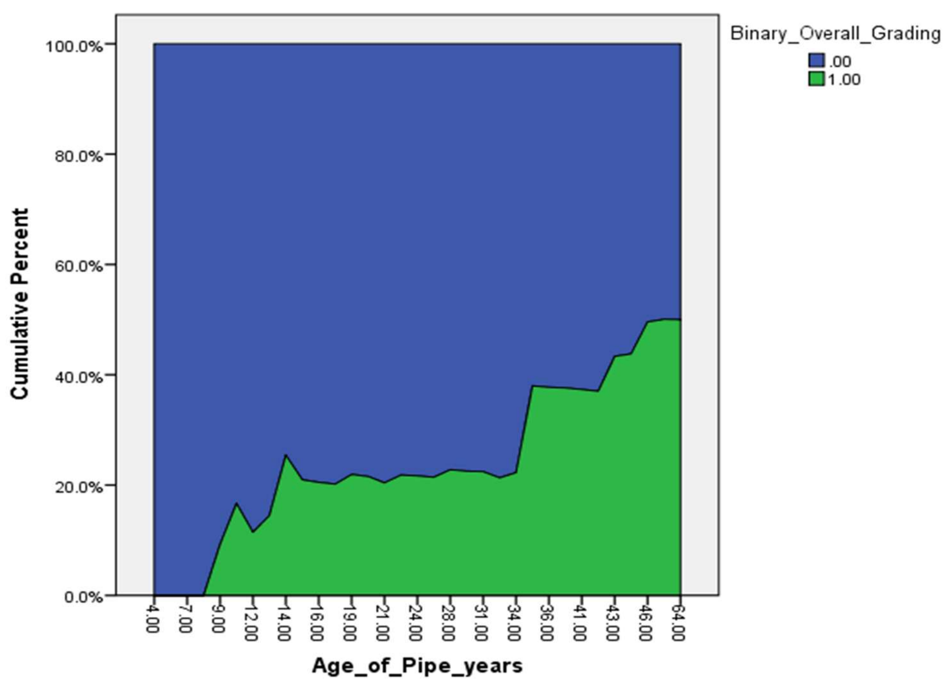

Figure 4.31: Binary overall condition grading of assets with respect to age of assets

4.18 Spearman rank correlation between predictors

In order to identify the changes in a predictor due to changes transpired in another predictor, the correlation analysis is conducted. When the correlation value between two predictors is observed to be high, it thus suggests that with high accuracy, the value of the predictor can be estimated using the other predictor. When independent variables with high correlation values are utilized in a regression model, even though the multiple determination coefficient, R^2, will be at a high level, multicollinearity can occur and therefore, significance test of the model variables can be impacted. On the other hand, in case it is desired to obtain the maximum value of R^2, multicollinearity might not result in an issue [Meyers et al. 2006, Salman 2010].

The typical correlation analysis is conducted through Pearson correlation; however, Pearson correlation can be utilized for linear relationship between variables; furthermore, if extreme data instances are present, Pearson correlation may be impacted by these data. Additionally, in Pearson correlation, it is assumed that the two variables of interest are also bivariately normal distributed. However, herein, based on the distributions obtained for the independent variables, it is observed that it is more suitable to utilize Spearman rank correlation. Therefore, based on the available data set, and for correlation analysis of the predictors used in deterioration modeling, it is decided to use Spearman rank correlation, which is a non-parametric correlation coefficient. If outliers are present in the data set, Spearman rank correlation will not be impacted and furthermore, this non-parametric correlation coefficient can also be utilized for nonlinear relationship between predictors as well. Spearman correlation coefficient can be obtained based on the following equation [Gravetter and Walnau 2004, Salman 2010]:

$$r_S = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\left((\sum x^2 - \frac{(\sum x)^2}{n})(\sum y^2 - \frac{(\sum y)^2}{n})\right)^{0.5}} \qquad (4.1)$$

The following table demonstrates Spearman rank correlation between various independent variables. For instance, based on these correlation values, it is observed that pipe diameter and the average flow depth in pipe are significantly correlated at the 0.01 level; furthermore, their associated correlation coefficient is +0.732, which indicated a high correlation between diameter of the pipe and average flow depth in pipe. Additionally, the positive sign of the correlation coefficient indicates that the larger the diameter of the sewer pipe is, the greater average flow depth is observed in the sewer pipe. Moreover, considering slope of the pipe and average flow velocity in the pipe, it is realized that these predictors are also significantly correlated at the 0.01 level; the correlation coefficient obtained for these independent variable is +0.283 which once again illustrates that for sewer pipes with higher values of pipe slope, the average velocity of the flow was also a greater value.

However, when considering pipe slope and average flow depth in pipe, it is observed that these predictors are significantly correlated at the 0.01 level as well and the correlation coefficient pertaining to these predictors is -0.463. The negative sign of the correlation coefficient indicates that for sewer pipes where pipe slope was greater, the average flow depth in pipe was lower.

Table 4.1: Spearman rank correlation between predictors of the model

Correlations								
		Pipe Diameter in	Age of Pipe years	Pipe Slope	Length of Pipe ft	Average Flow in Pipe percent full	Average Velocity ft/s	Average Flow Depth
Pipe Diameter in	Correlation Coefficient	1	-.120*	-.289**	-0.033	.278**	.321**	.732**
	Sig. (2-tailed)	.	0.015	0	0.503	0	0	0
	N	410	410	410	410	410	410	410
Age of Pipe years	Correlation Coefficient	-.120*	1	-.274**	.326**	.197**	-0.069	0.031
	Sig. (2-tailed)	0.015	.	0	0	0	0.161	0.525
	N	410	410	410	410	410	410	410
Pipe Slope	Correlation Coefficient	-.289**	-.274**	1	-.146**	-.424**	.283**	-.463**
	Sig. (2-tailed)	0	0	.	0.003	0	0	0
	N	410	410	410	410	410	410	410
Length of Pipe ft	Correlation Coefficient	-0.033	.326**	-.146**	1	0.066	0.021	0.044
	Sig. (2-tailed)	0.503	0	0.003	.	0.18	0.678	0.378
	N	410	410	410	410	410	410	410
Average Flow in Pipe percent full	Correlation Coefficient	.278**	.197**	-.424**	0.066	1	-.203**	.815**
	Sig. (2-tailed)	0	0	0	0.18	.	0	0
	N	410	410	410	410	410	410	410
Average Velocity ft/s	Correlation Coefficient	.321**	-0.069	.283**	0.021	-.203**	1	0.036
	Sig. (2-tailed)	0	0.161	0	0.678	0	.	0.466
	N	410	410	410	410	410	410	410
Average Flow Depth	Correlation Coefficient	.732**	0.031	-.463**	0.044	.815**	0.036	1
	Sig. (2-tailed)	0	0.525	0	0.378	0	0.466	.
	N	410	410	410	410	410	410	410
*. Correlation is significant at the 0.05 level (2-tailed).								
**. Correlation is significant at the 0.01 level (2-tailed).								

4.19 Significance of categorical variables

In order to determine the significance of categorical variables, cross table analysis was utilized. The categorical variable considered herein is the sewer pipe material; moreover, the cross tables were constructed considering operational (O&M) condition grading, structural condition grading and overall condition grading. In addition to these condition gradings, cross tables utilizing the associated binary condition gradings for each of the operational, structural and overall states were also taken into account. In order to determine significance of categorical variables, the expected and the observed frequencies for each individual cell are compared to one another. The expected frequency for various cells can be obtained based on the following equation [Cramer 1994, Salman 2010]:

$$Expected\ Frequency = \frac{Row\ Total\ \times Column\ Total}{Grand\ Total} \tag{4.2}$$

Furthermore, considering a cross table contains n cells, the chi-square statistic can be computer based on the following equation [Cramer 1994, Salman 2010]:

$$X^2 = \sum_{i=1}^{n} \frac{(Observed\ Frequency - Expected\ Frequency)^2}{Expected\ Frequency} \tag{4.3}$$

Additionally, assuming r and c denoted the number of rows and columns, respectively, therefore, the degrees of freedom associated with the critical chi-square is equal to (c-1)×(r-1); and this value will be compared to the chi-square computed based on the cross table analysis.

Based on the following tables, it is observed that considering cross table analyses based upon operational, structural, overall, and their associated binary condition gradings, sewer pipe material is found to be significant at 0.05 level in all cases.

Table 4.2: Cross table of pipe material and operational (O&M) grading

		Operational Grading					Total
		1	2	3	4	5	
CCFRPM	Count	26	0	0	0	0	26
	% within Pipe Material	100.00%	0.00%	0.00%	0.00%	0.00%	100.00%
	% within Operational Grading	10.70%	0.00%	0.00%	0.00%	0.00%	6.30%
	% of Total	6.30%	0.00%	0.00%	0.00%	0.00%	6.30%
FRP	Count	69	4	2	1	0	76
	% within Pipe Material	90.80%	5.30%	2.60%	1.30%	0.00%	100.00%
	% within Operational Grading	28.40%	4.50%	4.00%	4.00%	0.00%	18.50%
	% of Total	16.80%	1.00%	0.50%	0.20%	0.00%	18.50%
PVC	Count	95	46	19	0	1	161
	% within Pipe Material	59.00%	28.60%	11.80%	0.00%	0.60%	100.00%
	% within Operational Grading	39.10%	52.30%	38.00%	0.00%	25.00%	39.30%
	% of Total	23.20%	11.20%	4.60%	0.00%	0.20%	39.30%
RCP	Count	12	14	9	12	2	49
	% within Pipe Material	24.50%	28.60%	18.40%	24.50%	4.10%	100.00%
	% within Operational Grading	4.90%	15.90%	18.00%	48.00%	50.00%	12.00%
	% of Total	2.90%	3.40%	2.20%	2.90%	0.50%	12.00%
VCP	Count	41	24	20	12	1	98
	% within Pipe Material	41.80%	24.50%	20.40%	12.20%	1.00%	100.00%
	% within Operational Grading	16.90%	27.30%	40.00%	48.00%	25.00%	23.90%
	% of Total	10.00%	5.90%	4.90%	2.90%	0.20%	23.90%
Total	Count	243	88	50	25	4	410
	% within Pipe Material	59.30%	21.50%	12.20%	6.10%	1.00%	100.00%
	% within Operational Grading	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	% of Total	59.30%	21.50%	12.20%	6.10%	1.00%	100.00%

Table 4.3: Chi-squared tests for operational grading and pipe material

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	124.682[a]	16	.000
Likelihood Ratio	137.386	16	.000
N of Valid Cases	410		
a. 9 cells (36.0%) have expected count less than 5. The minimum expected count is .25.			

Table 4.4: Cross table of pipe material and structural grading

		Structural Grading					Total
		1	2	3	4	5	
CCFRPM	Count	26	0	0	0	0	26
	% within Pipe_Material	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%
	% within Structural_Grading	8.5%	0.0%	0.0%	0.0%	0.0%	6.3%
	% of Total	6.3%	0.0%	0.0%	0.0%	0.0%	6.3%
FRP	Count	69	0	5	2	0	76
	% within Pipe_Material	90.8%	0.0%	6.6%	2.6%	0.0%	100.0%
	% within Structural_Grading	22.5%	0.0%	7.9%	5.4%	0.0%	18.5%
	% of Total	16.8%	0.0%	1.2%	0.5%	0.0%	18.5%
PVC	Count	134	1	18	8	0	161
	% within Pipe_Material	83.2%	0.6%	11.2%	5.0%	0.0%	100.0%
	% within Structural_Grading	43.8%	50.0%	28.6%	21.6%	0.0%	39.3%
	% of Total	32.7%	0.2%	4.4%	2.0%	0.0%	39.3%
RCP	Count	33	0	11	5	0	49
	% within Pipe_Material	67.3%	0.0%	22.4%	10.2%	0.0%	100.0%
	% within Structural_Grading	10.8%	0.0%	17.5%	13.5%	0.0%	12.0%
	% of Total	8.0%	0.0%	2.7%	1.2%	0.0%	12.0%
VCP	Count	44	1	29	22	2	98
	% within Pipe_Material	44.9%	1.0%	29.6%	22.4%	2.0%	100.0%
	% within Structural_Grading	14.4%	50.0%	46.0%	59.5%	100.0%	23.9%
	% of Total	10.7%	0.2%	7.1%	5.4%	0.5%	23.9%
Total	Count	306	2	63	37	2	410
	% within Pipe_Material	74.6%	0.5%	15.4%	9.0%	0.5%	100.0%
	% within Structural_Grading	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total	74.6%	0.5%	15.4%	9.0%	0.5%	100.0%

Table 4.5: Chi-squared tests for structural grading and pipe material

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	78.710[a]	16	.000
Likelihood Ratio	80.918	16	.000
N of Valid Cases	410		
a. 13 cells (52.0%) have expected count less than 5. The minimum expected count is .13.			

Table 4.6: Cross table of pipe material and overall grading

		Overall_Grading					Total
		1	2	3	4	5	
CCFRPM	Count	26	0	0	0	0	26
	% within Pipe_Material	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%
	% within Overall_Grading	12.3%	0.0%	0.0%	0.0%	0.0%	6.3%
	% of Total	6.3%	0.0%	0.0%	0.0%	0.0%	6.3%
FRP	Count	63	3	7	3	0	76
	% within Pipe_Material	82.9%	3.9%	9.2%	3.9%	0.0%	100.0%
	% within Overall_Grading	29.7%	5.5%	9.0%	5.1%	0.0%	18.5%
	% of Total	15.4%	0.7%	1.7%	0.7%	0.0%	18.5%
PVC	Count	87	32	33	8	1	161
	% within Pipe_Material	54.0%	19.9%	20.5%	5.0%	0.6%	100.0%
	% within Overall_Grading	41.0%	58.2%	42.3%	13.6%	16.7%	39.3%
	% of Total	21.2%	7.8%	8.0%	2.0%	0.2%	39.3%
RCP	Count	10	10	11	16	2	49
	% within Pipe_Material	20.4%	20.4%	22.4%	32.7%	4.1%	100.0%
	% within Overall_Grading	4.7%	18.2%	14.1%	27.1%	33.3%	12.0%
	% of Total	2.4%	2.4%	2.7%	3.9%	0.5%	12.0%
VCP	Count	26	10	27	32	3	98
	% within Pipe_Material	26.5%	10.2%	27.6%	32.7%	3.1%	100.0%
	% within Overall_Grading	12.3%	18.2%	34.6%	54.2%	50.0%	23.9%
	% of Total	6.3%	2.4%	6.6%	7.8%	0.7%	23.9%
Total	Count	212	55	78	59	6	410
	% within Pipe_Material	51.7%	13.4%	19.0%	14.4%	1.5%	100.0%
	% within Overall_Grading	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total	51.7%	13.4%	19.0%	14.4%	1.5%	100.0%

Table 4.7: Chi-squared tests for overall grading and pipe material

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	136.410[a]	16	.000
Likelihood Ratio	147.993	16	.000
N of Valid Cases	410		
a. 8 cells (32.0%) have expected count less than 5. The minimum expected count is .38.			

Table 4.8: Cross table of pipe material and binary operational grading

			Binary_Operational_Grading		Total
			0	1	
Pipe_Material	CCFRPM	Count	26	0	26
		% within Pipe_Material	100.00%	0.00%	100.00%
		% within Binary_Operational_Grading	6.80%	0.00%	6.30%
		% of Total	6.30%	0.00%	6.30%
	FRP	Count	75	1	76
		% within Pipe_Material	98.70%	1.30%	100.00%
		% within Binary_Operational_Grading	19.70%	3.40%	18.50%
		% of Total	18.30%	0.20%	18.50%
	PVC	Count	160	1	161
		% within Pipe_Material	99.40%	0.60%	100.00%
		% within Binary_Operational_Grading	42.00%	3.40%	39.30%
		% of Total	39.00%	0.20%	39.30%
	RCP	Count	35	14	49
		% within Pipe_Material	71.40%	28.60%	100.00%
		% within Binary_Operational_Grading	9.20%	48.30%	12.00%
		% of Total	8.50%	3.40%	12.00%
	VCP	Count	85	13	98
		% within Pipe_Material	86.70%	13.30%	100.00%
		% within Binary_Operational_Grading	22.30%	44.80%	23.90%
		% of Total	20.70%	3.20%	23.90%
Total		Count	381	29	410
		% within Pipe_Material	92.90%	7.10%	100.00%
		% within Binary_Operational_Grading	100.00%	100.00%	100.00%
		% of Total	92.90%	7.10%	100.00%

Table 4.9: Chi-squared tests for binary operational grading and pipe material

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	56.180[a]	4	.000
Likelihood Ratio	51.383	4	.000
N of Valid Cases	410		
a. 2 cells (20.0%) have expected count less than 5. The minimum expected count is 1.84.			

Table 4.10: Cross table of pipe material and binary structural grading

			Binary_Structural_Grading		Total
			0	1	
Pipe_Material	CCFRPM	Count	26	0	26
		% within Pipe_Material	100.00%	0.00%	100.00%
		% within Binary_Structural_Grading	7.00%	0.00%	6.30%
		% of Total	6.30%	0.00%	6.30%
	FRP	Count	74	2	76
		% within Pipe_Material	97.40%	2.60%	100.00%
		% within Binary_Structural_Grading	19.90%	5.10%	18.50%
		% of Total	18.00%	0.50%	18.50%
	PVC	Count	153	8	161
		% within Pipe_Material	95.00%	5.00%	100.00%
		% within Binary_Structural_Grading	41.20%	20.50%	39.30%
		% of Total	37.30%	2.00%	39.30%
	RCP	Count	44	5	49
		% within Pipe_Material	89.80%	10.20%	100.00%
		% within Binary_Structural_Grading	11.90%	12.80%	12.00%
		% of Total	10.70%	1.20%	12.00%
	VCP	Count	74	24	98
		% within Pipe_Material	75.50%	24.50%	100.00%
		% within Binary_Structural_Grading	19.90%	61.50%	23.90%
		% of Total	18.00%	5.90%	23.90%
Total		Count	371	39	410
		% within Pipe_Material	90.50%	9.50%	100.00%
		% within Binary_Structural_Grading	100.00%	100.00%	100.00%
		% of Total	90.50%	9.50%	100.00%

Table 4.11: Chi-squared tests for binary structural grading and pipe material

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	36.343[a]	4	.000
Likelihood Ratio	34.144	4	.000
N of Valid Cases	410		
a. 2 cells (20.0%) have expected count less than 5. The minimum expected count is 2.47.			

Table 4.12: Cross table of pipe material and binary overall grading

			Binary_Overall_Grading		Total
			0	1	
Pipe_Material	CCFRPM	Count	26	0	26
		% within Pipe_Material	100.00%	0.00%	100.00%
		% within Binary_Overall_Grading	7.50%	0.00%	6.30%
		% of Total	6.30%	0.00%	6.30%
	FRP	Count	73	3	76
		% within Pipe_Material	96.10%	3.90%	100.00%
		% within Binary_Overall_Grading	21.20%	4.60%	18.50%
		% of Total	17.80%	0.70%	18.50%
	PVC	Count	152	9	161
		% within Pipe_Material	94.40%	5.60%	100.00%
		% within Binary_Overall_Grading	44.10%	13.80%	39.30%
		% of Total	37.10%	2.20%	39.30%
	RCP	Count	31	18	49
		% within Pipe_Material	63.30%	36.70%	100.00%
		% within Binary_Overall_Grading	9.00%	27.70%	12.00%
		% of Total	7.60%	4.40%	12.00%
	VCP	Count	63	35	98
		% within Pipe_Material	64.30%	35.70%	100.00%
		% within Binary_Overall_Grading	18.30%	53.80%	23.90%
		% of Total	15.40%	8.50%	23.90%
Total		Count	345	65	410
		% within Pipe_Material	84.10%	15.90%	100.00%
		% within Binary_Overall_Grading	100.00%	100.00%	100.00%
		% of Total	84.10%	15.90%	100.00%

Table 4.13: Chi-squared tests for binary overall grading and pipe material

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	70.680[a]	4	.000
Likelihood Ratio	71.676	4	.000
N of Valid Cases	410		
a. 1 cells (10.0%) have expected count less than 5. The minimum expected count is 4.12.			

Chapter 5 : Results Associated with Developed Models

5.1 Results Based on Statistical Modeling

In this section of the dissertation, the results pertaining to statistical deterioration models are presented. For all deterioration models, the development of the model is based upon the (binary) overall condition grading of the sewer pipe, depending on the approach utilized in modeling the deterioration of assets.

5.1.1 Binomial logistic regression

The results obtained based on binomial logistic regression model are presented below. Based on these results, it is observed that Omnibus tests of model coefficients yield desirable result; furthermore, based on Hosmer and Lemeshow test, it is observed that binomial logistic regression is a suitable approach for modeling the deterioration of the sewer pipes considered in this study. Based on the classification table obtained for binomial logistic regression, it is realized that the accuracy associated with binary overall grading 0 is 95.7%, whereas the accuracy associated binary overall grading 1 is 16.9%. Moreover, the overall accuracy of the model is 83.2%.

Additionally, it is observed that the age of the pipe is a significant independent variable and its coefficient is 0.060; therefore, this indicates that when the age of sewer pipe increases by one year, the odds ratio will be increased by 6.18%.

Table 5.1: Table containing categorical variables codings (binomial logistic regression)

		Frequency	Parameter coding			
			(1)	(2)	(3)	(4)
Pipe_Material	CCFRPM	26	1	0	0	0
	FRP	76	0	1	0	0
	PVC	161	0	0	1	0
	RCP	49	0	0	0	1
	VCP	98	0	0	0	0

Table 5.2: Table illustrating Omnibus tests of model coefficients (binomial logistic regression)

		Chi-square	df	Sig.
Step 1	Step	100.546	11	.000
	Block	100.546	11	.000
	Model	100.546	11	.000

Table 5.3: Nagelkerke and Cox & Snell R-Square values (binomial logistic regression)

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	257.987[a]	0.217	0.373
a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.			

Table 5.4: Hosmer and Lemeshow test (binomial logistic regression)

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	7.541	8	0.480

Table 5.5: Classification table (binomial logistic regression)

Classification Table[a]					
	Observed		Predicted		
			Binary_Overall_Grading		Percentage Correct
			0	1	
Step 1	Binary_Overall_Grading	0	330	15	95.7
		1	54	11	16.9
	Overall Percentage				83.2

a. The cut value is .500

Table 5.6: Variables in the equation (binomial logistic regression)

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Pipe_Material			14.553	4	0.006			
	Pipe_Material(1)	-19.565	7768.602	.000	1	0.998	0	0	.
	Pipe_Material(2)	-2.944	1.054	7.798	1	0.005	0.053	0.007	0.416
	Pipe_Material(3)	-1.500	0.517	8.417	1	0.004	0.223	0.081	0.615
	Pipe_Material(4)	-0.741	0.612	1.465	1	0.226	0.477	0.144	1.582
	Pipe_Diameter_in	0.250	0.099	6.329	1	0.012	1.284	1.057	1.559
	Age_of_Pipe_years	0.060	0.017	12.4	1	.000	1.062	1.027	1.098
	Pipe_Slope	0.196	0.124	2.471	1	0.116	1.216	0.953	1.552
	Length_of_Pipe_ft	.000	0.001	0.509	1	0.475	1	0.999	1.002
	Average_Velocity_ft_per_s	0.062	0.211	0.087	1	0.768	1.064	0.703	1.611
	Average_Flow_Depth	-0.701	0.31	5.119	1	0.024	0.496	0.27	0.911
	Average_Flow_in_Pipe_percent_full	0.194	0.082	5.549	1	0.018	1.214	1.033	1.428
	Constant	-10.146	2.831	12.849	1	0	0		

a. Variable(s) entered on step 1: Pipe_Material, Pipe_Diameter_in, Age_of_Pipe_years, Pipe_Slope, Length_of_Pipe_ft, Average_Velocity_ft_per_s, Average_Flow_Depth, Average_Flow_in_Pipe_percent_full.

5.1.2 Multinomial logistic regression

The following tables illustrate the results based upon multinomial logistic regression. The reference category used in developing the deterioration model is overall condition grading 1. Based on the classification table obtained through multinomial logistic regression, it is observed that the overall percentage of accuracy pertaining to this model is 59.3%; furthermore, it is observed that the percentage accuracy of the model associated with overall condition gradings 1, 2, 3, 4, and 5 are 88.7%, 1.8%, 11.5%, 72.9%, and 33.3%, respectively. Therefore, based on multinomial logistic regression, the highest and the lowest percentages of accuracy are associated with overall condition gradings 1 and 2, respectively. Additionally, based on the results pertaining to Goodness-of-Fit test and the likelihood ratio test available within model fitting information table, multinomial logistic regression is found to be a suitable approach for developing deterioration model of the sewer pipes considered in this study.

Table 5.7: Table illustrating model fitting information (multinomial logistic regression)

Model Fitting Information				
Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.953			
Final	828.911	210.042	44	.000

Table 5.8: Results of Goodness-of-Fit tests (multinomial logistic regression)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	1681.7	1592	0.058
Deviance	828.911	1592	1.000

Table 5.9: Table containing Pseudo R-Square values (multinomial logistic regression)

Pseudo R-Square	
Cox and Snell	0.401
Nagelkerke	0.435
McFadden	0.202

Table 5.10: Table illustrating likelihood ratio tests (multinomial logistic regression)

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	828.911[a]	.000	0	.
Pipe_Diameter_in	840.254	11.344	4	0.023
Age_of_Pipe_years	845.503	16.592	4	0.002
Pipe_Slope	833.619	4.709	4	0.319
Length_of_Pipe_ft	831.142	2.232	4	0.693
Average_Flow_in_Pipe_percent_full	846.989	18.078	4	0.001
Average_Velocity_ft_per_s	829.806	0.896	4	0.925
Average_Flow_Depth	844.915	16.004	4	0.003
Pipe_Material	892.261	63.350	16	.000
The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.				
a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.				

Table 5.11: Parameter estimates for overall gradings 2 and 3 (multinomial logistic regression)

Parameter Estimates									
Overall_Grading[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
2	Intercept	0.046	3.019	0	1	0.988			
	Pipe_Diameter_in	-0.027	0.107	0.065	1	0.799	0.973	0.789	1.2
	Age_of_Pipe_years	0.012	0.014	0.65	1	0.42	1.012	0.984	1.04
	Pipe_Slope	-0.184	0.213	0.745	1	0.388	0.832	0.548	1.264
	Length_of_Pipe_ft	0	0.001	0.002	1	0.962	1	0.999	1.001
	Average_Flow_in_Pipe_percent_full	-0.13	0.094	1.914	1	0.167	0.878	0.731	1.056
	Average_Velocity_ft_per_s	-0.138	0.213	0.416	1	0.519	0.871	0.574	1.324
	Average_Flow_Depth	0.429	0.326	1.732	1	0.188	1.535	0.811	2.907
	[Pipe_Material=CCFRPM]	-20.901	0	.	1	.	8.37E-10	8.37E-10	8.37E-10
	[Pipe_Material=FRP]	-3.55	1.134	9.8	1	0.002	0.029	0.003	0.265
	[Pipe_Material=PVC]	0.231	0.562	0.169	1	0.681	1.26	0.419	3.791
	[Pipe_Material=RCP]	0.211	0.755	0.078	1	0.779	1.236	0.281	5.428
	[Pipe_Material=VCP]	0[b]	.	.	0	.	.	.	.
3	Intercept	0.36	2.641	0.019	1	0.891			
	Pipe_Diameter_in	-0.059	0.096	0.37	1	0.543	0.943	0.781	1.139
	Age_of_Pipe_years	0.006	0.013	0.189	1	0.664	1.006	0.981	1.031
	Pipe_Slope	-0.208	0.212	0.962	1	0.327	0.812	0.536	1.231
	Length_of_Pipe_ft	0	0.001	0.529	1	0.467	1	0.999	1.002
	Average_Flow_in_Pipe_percent_full	-0.087	0.076	1.296	1	0.255	0.917	0.789	1.065
	Average_Velocity_ft_per_s	-0.003	0.205	0	1	0.986	0.997	0.667	1.489
	Average_Flow_Depth	0.425	0.282	2.28	1	0.131	1.53	0.881	2.657
	[Pipe_Material=CCFRPM]	-21.176	9838.481	0	1	0.998	6.36E-10	0	.[c]
	[Pipe_Material=FRP]	-2.94	0.913	10.364	1	0.001	0.053	0.009	0.317
	[Pipe_Material=PVC]	-0.414	0.456	0.824	1	0.364	0.661	0.271	1.615
	[Pipe_Material=RCP]	-0.221	0.666	0.11	1	0.74	0.802	0.217	2.959
	[Pipe_Material=VCP]	0[b]	.	.	0	.	.	.	.

a. The reference category is: 1.00.

b. This parameter is set to zero because it is redundant.

c. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.

Table 5.12: Parameter estimates for overall gradings 4 and 5 (multinomial logistic regression)

Parameter Estimates									
Overall_Grading[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
4	Intercept	-8.673	3.243	7.152	1	0.007			
	Pipe_Diameter_in	0.192	0.114	2.821	1	0.093	1.212	0.968	1.516
	Age_of_Pipe_years	0.069	0.019	12.778	1	.000	1.071	1.031	1.112
	Pipe_Slope	0.154	0.131	1.39	1	0.238	1.167	0.903	1.509
	Length_of_Pipe_ft	0	0.001	0.567	1	0.452	1	0.999	1.002
	Average_Flow_in_Pipe_percent_full	0.108	0.096	1.258	1	0.262	1.114	0.922	1.345
	Average_Velocity_ft_per_s	0.068	0.238	0.081	1	0.776	1.07	0.671	1.706
	Average_Flow_Depth	-0.319	0.36	0.782	1	0.377	0.727	0.359	1.473
	[Pipe_Material=CCFRPM]	-20.712	0	.	1	.	1.01E-09	1.01E-09	1.01E-09
	[Pipe_Material=FRP]	-4.075	1.201	11.51	1	0.001	0.017	0.002	0.179
	[Pipe_Material=PVC]	-1.543	0.587	6.91	1	0.009	0.214	0.068	0.675
	[Pipe_Material=RCP]	-0.735	0.72	1.043	1	0.307	0.479	0.117	1.966
	[Pipe_Material=VCP]	0[b]	.	.	0	.	.	.	.
5	Intercept	-16.967	7.342	5.341	1	0.021			
	Pipe_Diameter_in	0.557	0.264	4.453	1	0.035	1.746	1.041	2.93
	Age_of_Pipe_years	0.017	0.049	0.121	1	0.728	1.017	0.924	1.12
	Pipe_Slope	-0.904	2.818	0.103	1	0.748	0.405	0.002	101.376
	Length_of_Pipe_ft	0.002	0.002	1.605	1	0.205	1.002	0.999	1.005
	Average_Flow_in_Pipe_percent_full	0.429	0.259	2.742	1	0.098	1.536	0.924	2.552
	Average_Velocity_ft_per_s	-0.312	0.607	0.265	1	0.607	0.732	0.223	2.404
	Average_Flow_Depth	-1.775	1.003	3.131	1	0.077	0.169	0.024	1.211
	[Pipe_Material=CCFRPM]	-19.075	9667.955	0	1	0.998	5.20E-09	0	.[c]
	[Pipe_Material=FRP]	-19.836	2574.812	0	1	0.994	2.43E-09	0	.[c]
	[Pipe_Material=PVC]	-1.552	1.559	0.992	1	0.319	0.212	0.01	4.494
	[Pipe_Material=RCP]	-0.898	1.673	0.288	1	0.591	0.407	0.015	10.815
	[Pipe_Material=VCP]	0[b]	.	.	0	.	.	.	.
a. The reference category is: 1.00.									
b. This parameter is set to zero because it is redundant.									
c. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.									

Table 5.13: Classification table obtained for multinomial logistic regression

Classification						
Observed	Predicted					
	1	2	3	4	5	Percent Correct
1	188	0	7	17	0	88.70%
2	38	1	4	12	0	1.80%
3	45	2	9	22	0	11.50%
4	14	0	2	43	0	72.90%
5	1	0	1	2	2	33.30%
Overall Percentage	69.80%	0.70%	5.60%	23.40%	0.50%	59.30%

5.1.3 Ordinal regression with Logit link function

Utilizing ordinal regression with Logit link function with all independent variables, the following tables are obtained. Based on these results, it is observed that Goodness-of-Fit test (Pearson as well as Deviance) and likelihood ratio test available within model fitting information table are both satisfied. However, with regards to test of parallel lines which is conducted for verifying the proportional odds assumption associated with ordinal regression, it is realized that this test did not yield desirable results. Therefore, the full model is not suitable to be used for modeling deterioration of sewer pipes. However, it was realized that by removing the average flow in pipe (%full) from the deterioration model, the test of parallel lines did yield satisfactory results; and therefore, the proportional odds assumption was satisfied. It should be noted that Goodness-of-Fit test (both Pearson and Deviance) as well as likelihood ratio test available within model fitting information table were still satisfied in the new model. The results pertaining to both modeling approaches are presented in the following tables.

Model including all independent variables:

The full model did not pass the test of parallel lines; therefore, the proportional odds assumption is not satisfied.

Table 5.14: Table containing model fitting information (ordinal regression using Logit link function: full model)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.95			
Final	885.222	153.73	11	.000
Link function: Logit.				

Table 5.15: Results of Goodness-of-Fit tests (ordinal regression using Logit link function: full model)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	1419.47	1625	1.000
Deviance	885.222	1625	1.000
Link function: Logit.			

Table 5.16: Table containing Pseduo R-Square values (ordinal regression using Logit link function: full model)

Pseudo R-Square	
Cox and Snell	0.313
Nagelkerke	0.340
McFadden	0.148
Link function: Logit.	

Table 5.17: Parameter estimates for ordinal regression using Logit link function: full model

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[Overall_Grading = 1.00]	3.936	1.365	8.316	1	0.004	1.261	6.611
	[Overall_Grading = 2.00]	4.671	1.369	11.641	1	0.001	1.988	7.355
	[Overall_Grading = 3.00]	5.988	1.382	18.764	1	.000	3.279	8.698
	[Overall_Grading = 4.00]	8.799	1.485	35.102	1	.000	5.888	11.71
Location	Pipe_Diameter_in	0.143	0.049	8.326	1	0.004	0.046	0.239
	Age_of_Pipe_years	0.025	0.009	7.635	1	0.006	0.007	0.043
	Pipe_Slope	-0.01	0.094	0.012	1	0.912	-0.194	0.173
	Length_of_Pipe_ft	.000	0	0.409	1	0.522	-0.001	0.001
	Average_Flow_in_Pipe_percent_full	0.088	0.037	5.583	1	0.018	0.015	0.161
	Average_Velocity_ft_per_s	-0.02	0.134	0.021	1	0.884	-0.282	0.243
	Average_Flow_Depth	-0.281	0.141	3.947	1	0.047	-0.558	-0.004
	[Pipe_Material=CCFRPM]	-22.853	0	.	1	.	-22.853	-22.853
	[Pipe_Material=FRP]	-3.605	0.609	35.026	1	.000	-4.799	-2.411
	[Pipe_Material=PVC]	-0.877	0.313	7.858	1	0.005	-1.49	-0.264
	[Pipe_Material=RCP]	-0.705	0.432	2.657	1	0.103	-1.552	0.143
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Logit.								
a. This parameter is set to zero because it is redundant.								

Table 5.18: Results of test of parallel lines for ordinal regression using Logit link function: full model

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	885.222			
General	834.475[b]	50.748[c]	33	0.025
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Logit.				
b. The log-likelihood value cannot be further increased after maximum number of step-halving.				
c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.				

Model without average flow in pipe (%full):

In the following, the average flow in pipe (%full) is removed from the model and thus the test of parallel lines is satisfied; furthermore, it is observed that the age of sewer pipe is found to be a significant independent variable as well.

Table 5.19: Table containing model fitting information (ordinal regression using Logit link function)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.95			
Final	889.666	149.286	10	.000
Link function: Logit.				

Table 5.20: Results obtained based on Goodness-of-Fit tests (ordinal regression using Logit link function)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	1439.06	1626	1.000
Deviance	889.666	1626	1.000
Link function: Logit.			

Table 5.21: Table containing Pseudo R-Square values for ordinal regression using Logit link function

Pseudo R-Square	
Cox and Snell	0.305
Nagelkerke	0.331
McFadden	0.144
Link function: Logit.	

Table 5.22: Parameter estimates obtained based on ordinal regression using Logit link function

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	1.163	0.7	2.758	1	0.097	-0.21	2.535
	[Overall_Grading = 2.00]	1.898	0.703	7.295	1	0.007	0.521	3.275
	[Overall_Grading = 3.00]	3.206	0.715	20.081	1	.000	1.804	4.608
	[Overall_Grading = 4.00]	5.946	0.834	50.782	1	.000	4.311	7.581
Location	Pipe_Diameter_in	0.043	0.025	2.918	1	0.088	-0.006	0.092
	Age_of_Pipe_years	0.022	0.009	6.247	1	0.012	0.005	0.04
	Average_Flow_Depth	0.033	0.04	0.679	1	0.41	-0.045	0.111
	Pipe_Slope	-0.019	0.094	0.041	1	0.839	-0.203	0.165
	Length_of_Pipe_ft	.000	0	0.36	1	0.548	-0.001	0.001
	Average_Velocity_ft_per_s	0.05	0.13	0.144	1	0.704	-0.206	0.305
	[Pipe_Material=CCFRPM]	-22.867	0	.	1	.	-22.867	-22.867
	[Pipe_Material=FRP]	-3.401	0.601	31.993	1	.000	-4.579	-2.222
	[Pipe_Material=PVC]	-1.054	0.304	12.002	1	0.001	-1.651	-0.458
	[Pipe_Material=RCP]	-0.578	0.428	1.819	1	0.177	-1.417	0.262
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Logit.								
a. This parameter is set to zero because it is redundant.								

Table 5.23: Results of test of parallel lines obtained based on ordinal regression using Logit link function

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	889.666			
General	849.629[b]	40.037[c]	30	0.104
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Logit.				
b. The log-likelihood value cannot be further increased after maximum number of step-halving.				
c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.				

5.1.4 Ordinal regression with Probit link function

The following tables demonstrate the results obtained by using ordinal regression with Probit link function. Based on these results, it is realized that both Goodness-of-Fit test (Pearson and Deviance) as well as likelihood ratio test available within model fitting information table are satisfied. Furthermore, the proportional odds assumption is also validated through test of parallel lines. Based on the following results, it is observed that similar to previous modeling techniques, the age of the sewer pipes is a significant independent variable.

Table 5.24: Table containing model fitting information (ordinal regression using Probit link function)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.95			
Final	.000	1038.953	11	.000
Link function: Probit.				

Table 5.25: Results of Goodness-of-Fit tests (ordinal regression using Probit link function)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	1588.129	1625	0.739
Deviance	883.156	1625	1.000
Link function: Probit.			

Table 5.26: Table containing Pseudo R-Square values for ordinal regression using Probit link function

Pseudo R-Square	
Cox and Snell	0.921
Nagelkerke	1.000
McFadden	1.000
Link function: Probit.	

Table 5.27: Parameter estimates obtained based on ordinal regression using Probit link function

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	2.32	0.788	8.663	1	0.003	0.775	3.865
	[Overall_Grading = 2.00]	2.761	0.79	12.21	1	.000	1.212	4.309
	[Overall_Grading = 3.00]	3.538	0.795	19.791	1	.000	1.979	5.097
	[Overall_Grading = 4.00]	4.99	0.831	36.068	1	.000	3.362	6.619
Location	Pipe_Diameter_in	0.082	0.029	8.183	1	0.004	0.026	0.138
	Age_of_Pipe_years	0.015	0.005	8.212	1	0.004	0.005	0.026
	Pipe_Slope	0.001	0.055	.000	1	0.985	-0.107	0.109
	Length_of_Pipe_ft	.000	.000	0.518	1	0.472	0	0.001
	Average_Flow_in_Pipe_percent_full	0.051	0.021	5.772	1	0.016	0.009	0.093
	Average_Velocity_ft_per_s	-0.005	0.078	0.004	1	0.949	-0.159	0.149
	Average_Flow_Depth	-0.163	0.081	4.009	1	0.045	-0.323	-0.003
	[Pipe_Material=CCFRPM]	-6.131	327.222	.000	1	0.985	-647.474	635.212
	[Pipe_Material=FRP]	-2.042	0.34	36.123	1	.000	-2.708	-1.376
	[Pipe_Material=PVC]	-0.523	0.185	7.966	1	0.005	-0.886	-0.16
	[Pipe_Material=RCP]	-0.387	0.254	2.316	1	0.128	-0.885	0.111
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Probit.								
a. This parameter is set to zero because it is redundant.								

Table 5.28: Results of test of parallel lines obtained based on ordinal regression using Probit link function

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	.000			
General	.000[b]	.000	33	1.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Probit.				
b. The log-likelihood value is practically zero. There may be a complete separation in the data. The maximum likelihood estimates do not exist.				

5.1.5 Ordinal regression with Negative Log-Log link function

Utilizing ordinal regression with Negative Log-Log link function for service life estimation of sewer pipes, it was observed that although the developed model satisfied Goodness-of-Fit test (Pearson and Deviance) as well as likelihood ratio test available within model fitting information table, however, the results from test of parallel lines illustrated that the proportional odds assumption was not satisfied for the sewer pipes considered in this study. It was further observed that upon removal of the average flow in pipe (%full) as well as the average velocity of flow in pipe from the model, it was observed that in addition to satisfying the Goodness-of-Fit test and the likelihood ratio test available within model fitting information table, based upon the results obtained from test of parallel lines, the proportional odds assumption was also satisfied. It is further observed that the age of sewer pipes is a significant variable of this model. The results pertaining to the model with all independent variables and the model without average flow in pipe (%full) and average velocity of flow are presented below.

Model developed by all independent variable:

Proportional odds assumption was not validated for this model as the full model did not pass the test of parallel lines. The following tables include the results from this model:

Table 5.29: Table containing model fitting information (ordinal regression using Negative Log-Log link function: full model)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.953			
Final	894.185	144.768	11	.000
Link function: Negative Log-log.				

Table 5.30: Results of Goodness-of-Fit tests (ordinal regression using Negative Log-Log link function: full model)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	1368.292	1625	1.000
Deviance	894.185	1625	1.000
Link function: Negative Log-log.			

Table 5.31: Table containing Pseudo R-Square values for ordinal regression using Negative Log-Log link function: full model)

Pseudo R-Square	
Cox and Snell	0.297
Nagelkerke	0.323
McFadden	0.139
Link function: Negative Log-log.	

Table 5.32: Parameter estimates obtained based on ordinal regression using Negative Log-Log link function: full model)

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	2.537	0.896	8.027	1	0.005	0.782	4.293
	[Overall_Grading = 2.00]	3.053	0.9	11.513	1	0.001	1.29	4.817
	[Overall_Grading = 3.00]	4.067	0.912	19.899	1	0	2.28	5.853
	[Overall_Grading = 4.00]	6.612	1	43.677	1	0	4.651	8.572
Location	Pipe_Diameter_in	0.081	0.032	6.394	1	0.011	0.018	0.144
	Age_of_Pipe_years	0.02	0.007	9.259	1	0.002	0.007	0.033
	Pipe_Slope	-0.004	0.071	0.004	1	0.953	-0.143	0.134
	Length_of_Pipe_ft	7.29E-05	0	0.075	1	0.785	0	0.001
	Average_Flow_in_Pipe_percent_full	0.053	0.024	5.084	1	0.024	0.007	0.1
	Average_Velocity_ft_per_s	-0.109	0.094	1.35	1	0.245	-0.293	0.075
	Average_Flow_Depth	-0.166	0.091	3.314	1	0.069	-0.345	0.013
	[Pipe_Material=CCFRPM]	-21.851	0	.	1	.	-21.851	-21.851
	[Pipe_Material=FRP]	-2.29	0.441	26.946	1	0	-3.154	-1.425
	[Pipe_Material=PVC]	-0.505	0.215	5.545	1	0.019	-0.926	-0.085
	[Pipe_Material=RCP]	-0.234	0.288	0.659	1	0.417	-0.799	0.331
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Negative Log-log.								
a. This parameter is set to zero because it is redundant.								

Table 5.33: Results of test of parallel lines obtained based on ordinal regression using Negative Log-Log link function: full model)

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	894.185			
General	785.466[b]	108.719[c]	33	.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Negative Log-log.				
b. The log-likelihood value cannot be further increased after maximum number of step-halving.				
c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.				

Model developed without average flow in pipe (%full) and average velocity of flow:

By removing both the average flow in pipe (%full) and the average velocity of flow in pipe from the model, the test of parallel lines is satisfied; therefore, proportional odds assumption is valid for this model. As observed in the following results, the age of pipe is determined as a significant independent variable.

Table 5.34: Table containing model fitting information (ordinal regression using Negative Log-Log link function)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.953			
Final	897.571	141.382	9	.000
Link function: Negative Log-log.				

Table 5.35: Results of Goodness-of-Fit tests (ordinal regression using Negative Log-Log link function)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	1396.114	1627	1.000
Deviance	897.571	1627	1.000
Link function: Negative Log-log.			

Table 5.36: Table containing Pseudo R-Square values for ordinal regression using Negative Log-Log link function

Pseudo R-Square	
Cox and Snell	0.292
Nagelkerke	0.317
McFadden	0.136
Link function: Negative Log-log.	

Table 5.37: Parameter estimates obtained based on ordinal regression using Negative Log-Log link function

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	0.951	0.465	4.187	1	0.041	0.04	1.861
	[Overall_Grading = 2.00]	1.465	0.469	9.755	1	0.002	0.546	2.384
	[Overall_Grading = 3.00]	2.474	0.482	26.359	1	.000	1.53	3.419
	[Overall_Grading = 4.00]	5.003	0.622	64.664	1	.000	3.784	6.223
Location	Pipe_Diameter_in	0.017	0.016	1.054	1	0.305	-0.015	0.049
	Age_of_Pipe_years	0.017	0.006	7.291	1	0.007	0.005	0.03
	Average_Flow_Depth	0.034	0.025	1.897	1	0.168	-0.014	0.083
	Pipe_Slope	-0.023	0.071	0.104	1	0.747	-0.162	0.116
	Length_of_Pipe_ft	6.22E-05	.000	0.055	1	0.815	.000	0.001
	[Pipe_Material=CCFRPM]	-21.913	.000	.	1	.	-21.913	-21.913
	[Pipe_Material=FRP]	-2.197	0.431	25.937	1	.000	-3.042	-1.351
	[Pipe_Material=PVC]	-0.602	0.207	8.466	1	0.004	-1.007	-0.196
	[Pipe_Material=RCP]	-0.182	0.279	0.426	1	0.514	-0.73	0.365
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Negative Log-log.								
a. This parameter is set to zero because it is redundant.								

Table 5.38: Results of test of parallel lines obtained based on ordinal regression using Negative Log-Log link function

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	897.571			
General	875.982[b]	21.589[c]	27	0.758
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Negative Log-log.				
b. The log-likelihood value cannot be further increased after maximum number of step-halving.				
c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.				

5.1.6 Ordinal regression with Complementary Log-Log link function

When ordinal regression with Complementary Log-Log link function is utilized for obtaining the deterioration model in sewer pipes, it is realized that only Deviance Goodness-of-Fit test yielded desirable result; it was further observed that both likelihood ratio test available within model fitting information table as well as test of parallel lines, illustrated satisfactory results. Therefore, the proportional odds assumption using Complementary Log-Log link function is validated. The following tables contain results from this deterioration model:

Table 5.39: Table containing model fitting information (ordinal regression using Complementary Log-Log link function)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.95			
Final	.000	1038.95	11	.000
Link function: Complementary Log-log.				

Table 5.40: Results of Goodness-of-Fit tests (ordinal regression using Complementary Log-Log link function)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	3650.789	1625	.000
Deviance	887.914	1625	1.000
Link function: Complementary Log-log.			

Table 5.41: Table containing Pseudo R-Square values for ordinal regression using Complementary Log-Log link function

Pseudo R-Square	
Cox and Snell	0.921
Nagelkerke	1.000
McFadden	1.000
Link function: Complementary Log-log.	

Table 5.42: Parameter estimates obtained based on ordinal regression using Complementary Log-Log link function

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[Overall_Grading = 1.00]	1.787	0.795	5.058	1	0.025	0.23	3.344
	[Overall_Grading = 2.00]	2.255	0.796	8.032	1	0.005	0.695	3.814
	[Overall_Grading = 3.00]	3.006	0.8	14.126	1	.000	1.438	4.574
	[Overall_Grading = 4.00]	4.152	0.821	25.578	1	.000	2.543	5.761
Location	Pipe_Diameter_in	0.074	0.029	6.472	1	0.011	0.017	0.13
	Age_of_Pipe_years	0.014	0.005	7.532	1	0.006	0.004	0.025
	Pipe_Slope	0.005	0.052	0.009	1	0.923	-0.097	0.107
	Length_of_Pipe_ft	.000	.000	1.075	1	0.3	.000	0.001
	Average_Flow_in_Pipe_percent_full	0.045	0.022	4.081	1	0.043	0.001	0.088
	Average_Velocity_ft_per_s	0.056	0.074	0.583	1	0.445	-0.088	0.201
	Average_Flow_Depth	-0.143	0.083	2.974	1	0.085	-0.306	0.02
	[Pipe_Material=CCFRPM]	-2.698	0.627	18.491	1	.000	-3.927	-1.468
	[Pipe_Material=FRP]	-1.951	0.316	38.135	1	.000	-2.57	-1.332
	[Pipe_Material=PVC]	-0.587	0.187	9.894	1	0.002	-0.952	-0.221
	[Pipe_Material=RCP]	-0.479	0.25	3.656	1	0.056	-0.97	0.012
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.

Link function: Complementary Log-log.

a. This parameter is set to zero because it is redundant.

Table 5.43: Results of test of parallel lines obtained based on ordinal regression using Complementary Log-Log link function

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	.000			
General	.000[b]	.000	33	1.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Complementary Log-log.				
b. The log-likelihood value is practically zero. There may be a complete separation in the data. The maximum likelihood estimates do not exist.				

5.1.7 Ordinal regression with Cauchit link function

Tables below illustrate the results obtained from ordinal regression utilizing Cauchit link function. Based on these results, it is observed that even though the Goodness-of-Fit test (Deviance and Pearson) and the likelihood ratio test available within model fitting information table both were satisfied, however, the test of parallel lines did not yield desirable results and therefore the proportional odds assumption cannot be validated. Additionally, removal of independent variables (as observed for Logit and Negative Log-Log link functions) did not yield satisfactory results either. Furthermore, the results associated with Cauchit link functions are solely presented for illustration and comparison purposes.

Table 5.44: Table containing model fitting information (ordinal regression using Cauchit link function)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1038.953			
Final	909.646	129.307	11	.000
Link function: Cauchit.				

Table 5.45: Results of Goodness-of-Fit tests (ordinal regression using Cauchit link function)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	1449.17	1625	0.999
Deviance	909.646	1625	1.000
Link function: Cauchit.			

Table 5.46: Table containing Pseudo R-Square values for ordinal regression using Cauchit link function

Pseudo R-Square	
Cox and Snell	0.270
Nagelkerke	0.294
McFadden	0.124
Link function: Cauchit.	

Table 5.47: Parameter estimates obtained based on ordinal regression using Cauchit link function

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	0.201	1.204	0.028	1	0.867	-2.158	2.56
	[Overall_Grading = 2.00]	0.885	1.203	0.541	1	0.462	-1.473	3.243
	[Overall_Grading = 3.00]	2.19	1.224	3.202	1	0.074	-0.209	4.588
	[Overall_Grading = 4.00]	14.933	4.968	9.034	1	0.003	5.196	24.671
Location	Pipe_Diameter_in	0.024	0.044	0.307	1	0.58	-0.061	0.109
	Age_of_Pipe_years	0.015	0.008	3.783	1	0.052	0	0.031
	Pipe_Slope	-0.125	0.103	1.493	1	0.222	-0.326	0.076
	Length_of_Pipe_ft	.000	.000	0.518	1	0.472	0	0.001
	Average_Flow_in_Pipe_percent_full	-0.029	0.032	0.853	1	0.356	-0.092	0.033
	Average_Velocity_ft_per_s	0.057	0.12	0.229	1	0.632	-0.177	0.292
	Average_Flow_Depth	0.116	0.123	0.893	1	0.345	-0.125	0.357
	[Pipe_Material=CCFRPM]	-1220.282	.000	.	1	.	-1220.282	-1220.282
	[Pipe_Material=FRP]	-3.549	0.707	25.16	1	.000	-4.935	-2.162
	[Pipe_Material=PVC]	-0.987	0.297	11.029	1	0.001	-1.57	-0.405
	[Pipe_Material=RCP]	-0.792	0.404	3.846	1	0.05	-1.583	0
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Cauchit.								
a. This parameter is set to zero because it is redundant.								

Table 5.48: Results of test of parallel lines obtained based on ordinal regression using Cauchit link function

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	909.646			
General	.000[b]	909.646	33	.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Cauchit.				
b. The log-likelihood value is practically zero. There may be a complete separation in the data. The maximum likelihood estimates do not exist.				

5.1.8 Binomial logistic regression considering initial condition gradings

In this section, the deterioration model for sewer pipes is developed based on binomial logistic regression and by considering the initial condition grading of sewer pipes. It is assumed that sewer pipes were initially at perfect condition (binary overall condition grading 0) and no defects due to inherent defects within sewer pipes or due to installation of sewer pipes, etc. were introduced to the pipes at the start of their service life. Based on the results presented in the following tables, it is realized that Ombinus tests of model coefficients and Hosmer and Lemeshow test both yield satisfactory results, and therefore, this approach seems to be suitable for developing deterioration model in sewer pipes. The classification table obtained herein shows that the percentage of accuracy associated with binary overall condition grading 0 is 97.6%; however, the percentage of accuracy for condition grading 1 is 16.9%.

The overall accuracy of the model is found to be 91.2%. It is also observed that the overall accuracy of the model (i.e. 91.2%) is greater than the model developed based on binomial logistic regression but without considering the initial condition gradings of sewer pipes (i.e. 83.2%). Based on this model, the age of sewer pipe is found to be a significant independent variable. Furthermore, in this model, the coefficient of the age of pipe is 0.100, whereas based on binomial logistic regression and when the initial condition gradings of sewer pipes were not accounted for, this value was 0.060. This indicates that by considering the initial condition grading of sewer pipes, the deterioration rate is increasing. In this model, increasing the age of sewer pipe by one year results in 10.52% increase in the odds ratio, whereas without considering the initial condition grading of pipes, this value corresponds to 6.18%.

Table 5.49: Table containing categorical variables codings (binomial logistic regression and considering initial condition gradings)

Categorical Variables Codings						
		Frequency	Parameter coding			
			(1)	(2)	(3)	(4)
Pipe_Material	CCFRPM	52	1	0	0	0
	FRP	152	0	1	0	0
	PVC	322	0	0	1	0
	RCP	98	0	0	0	1
	VCP	196	0	0	0	0

Table 5.50: Table illustrating Omnibus tests of model coefficients (binomial logistic regression and considering initial condition gradings)

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	181.542	11	.000
	Block	181.542	11	.000
	Model	181.542	11	.000

Table 5.51: Nagelkerke and Cox & Snell R-Square values (binomial logistic regression and considering initial condition gradings)

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	272.703[a]	0.199	0.467
a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.			

Table 5.52: Hosmer and Lemeshow test (binomial logistic regression and considering initial condition gradings)

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	3.135	8	0.926

Table 5.53: Classification table (binomial logistic regression and considering initial condition gradings)

Classification Table[a]					
	Observed		Predicted		
			Binary_Overall_Grading		Percentage Correct
			0	1	
Step 1	Binary_Overall_Grading	0	737	18	97.6
		1	54	11	16.9
	Overall Percentage				91.2
a. The cut value is .500					

Table 5.54: Variables in the equation (binomial logistic regression and considering initial condition gradings)

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Age_of_Pipe_years	0.1	0.014	50.778	1	0	1.105	1.075	1.136
	Pipe_Material			7.997	4	0.092			
	Pipe_Material(1)	-17.694	5441.162	0	1	0.997	0	0	.
	Pipe_Material(2)	-1.706	0.976	3.056	1	0.08	0.182	0.027	1.23
	Pipe_Material(3)	-1.111	0.497	5.001	1	0.025	0.329	0.124	0.872
	Pipe_Material(4)	-0.274	0.593	0.213	1	0.645	0.761	0.238	2.432
	Pipe_Slope	0.218	0.12	3.321	1	0.068	1.244	0.984	1.572
	Average_Flow_in_Pipe_percent_full	0.127	0.041	9.863	1	0.002	1.136	1.049	1.23
	Length_of_Pipe_ft	0	0.001	0.326	1	0.568	1	0.999	1.001
	Average_Velocity_ft_per_s	0.056	0.206	0.073	1	0.787	1.057	0.706	1.583
	Pipe_Diameter_in	0.168	0.06	7.855	1	0.005	1.182	1.052	1.329
	Average_Flow_Depth	-0.437	0.166	6.93	1	0.008	0.646	0.467	0.894
	Constant	-9.709	1.75	30.764	1	0	0		

a. Variable(s) entered on step 1: Age_of_Pipe_years, Pipe_Material, Pipe_Slope, Average_Flow_in_Pipe_percent_full, Length_of_Pipe_ft, Average_Velocity_ft_per_s, Pipe_Diameter_in, Average_Flow_Depth.

5.1.9 Multinomial logistic regression considering initial condition gradings

The following table contain the results obtained based on multinomial logistic regression and by taking the initial condition grading of sewer pipes into account. As stated earlier, it is assumed that the initial overall condition grading of sewer pipes is 1 (similar to binary overall condition grading 0) and they were initially in perfect condition. The results illustrate that both Goodness-of-Fit test and the likelihood ratio test available within model fitting information table yield satisfactory results; therefore, multinomial logistic regression considering initial condition gradings of sewer pipes seems to be a suitable technique to be used as a deterioration model. Moreover, considering the classification table, the results show that the percentage accuracy of this model for overall condition gradings 1, 2, 3, 4, and 5 are 96.8%, 9.1%, 3.8%, 72.9%, and 16.7%, respectively. Compared to multinomial logistic regression without considering the initial condition gradings of sewer pipes, wherein the overall accuracy of the model was 59.3%, herein the overall accuracy of the model is computed to be 79.8% which shows an increase by 20.5% in the overall accuracy.

Table 5.55: Table containing model fitting information (multinomial logistic regression and considering initial condition gradings)

Model Fitting Information				
Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1377.578			
Final	953.664	423.914	44	.000

Table 5.56: Results of Goodness-of-Fit tests (multinomial logistic regression and considering initial condition gradings)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	2621.255	3232	1.000
Deviance	953.664	3232	1.000

Table 5.57: Table containing Pseudo R-Square values (multinomial logistic regression and considering initial condition gradings function)

Pseudo R-Square	
Cox and Snell	0.404
Nagelkerke	0.496
McFadden	0.308

Table 5.58: Table illustrating likelihood ratio tests (multinomial logistic regression and considering initial condition gradings)

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	953.664[a]	0	0	.
Pipe_Diameter_in	962.157	8.494	4	0.075
Age_of_Pipe_years	1227.68	274.013	4	0
Pipe_Slope	957.43	3.766	4	0.439
Length_of_Pipe_ft	957.647	3.983	4	0.408
Average_Velocity_ft_per_s	955.798	2.134	4	0.711
Average_Flow_Depth	965.537	11.874	4	0.018
Average_Flow_in_Pipe_percent_full	968.892	15.228	4	0.004
Pipe_Material	989.329	35.665	16	0.003
The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.				
a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.				

Table 5.59: Parameter estimates for overall gradings 2 and 3 (multinomial logistic regression and considering initial condition gradings)

Parameter Estimates									
Overall_Grading[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
2	Intercept	-3.076	2.626	1.372	1	0.241			
	Pipe_Diameter_in	-0.001	0.092	0	1	0.991	0.999	0.834	1.197
	Age_of_Pipe_years	0.091	0.011	69.592	1	.000	1.095	1.072	1.118
	Pipe_Slope	-0.103	0.181	0.323	1	0.57	0.902	0.633	1.286
	Length_of_Pipe_ft	-0.001	0.001	2.62	1	0.105	0.999	0.998	1
	Average_Velocity_ft_per_s	-0.24	0.197	1.481	1	0.224	0.787	0.534	1.158
	Average_Flow_Depth	0.268	0.273	0.969	1	0.325	1.308	0.766	2.232
	Average_Flow_in_Pipe_percent_full	-0.084	0.081	1.081	1	0.298	0.92	0.785	1.077
	[Pipe_Material=CCFRPM]	-18.884	9601.375	0	1	0.998	6.29E-09	0	.[b]
	[Pipe_Material=FRP]	-1.968	1.031	3.642	1	0.056	0.14	0.019	1.055
	[Pipe_Material=PVC]	0.944	0.519	3.305	1	0.069	2.571	0.929	7.117
	[Pipe_Material=RCP]	0.271	0.71	0.146	1	0.703	1.311	0.326	5.276
	[Pipe_Material=VCP]	0[c]	.	.	0	.	.	.	.
3	Intercept	-3.131	2.05	2.332	1	0.127			
	Pipe_Diameter_in	-0.015	0.076	0.039	1	0.843	0.985	0.849	1.143
	Age_of_Pipe_years	0.088	0.009	88.131	1	.000	1.092	1.072	1.112
	Pipe_Slope	-0.113	0.172	0.429	1	0.512	0.893	0.638	1.251
	Length_of_Pipe_ft	0	0.001	0.764	1	0.382	1	0.999	1.001
	Average_Velocity_ft_per_s	-0.171	0.182	0.88	1	0.348	0.843	0.59	1.205
	Average_Flow_Depth	0.23	0.209	1.209	1	0.272	1.259	0.835	1.897
	Average_Flow_in_Pipe_percent_full	-0.034	0.057	0.361	1	0.548	0.967	0.865	1.08
	[Pipe_Material=CCFRPM]	-19.108	8138.529	0	1	0.998	5.03E-09	0	.[b]
	[Pipe_Material=FRP]	-1.439	0.799	3.244	1	0.072	0.237	0.05	1.135
	[Pipe_Material=PVC]	0.331	0.397	0.697	1	0.404	1.393	0.64	3.031
	[Pipe_Material=RCP]	-0.252	0.608	0.172	1	0.678	0.777	0.236	2.557
	[Pipe_Material=VCP]	0[c]	.	.	0	.	.	.	.
a. The reference category is: 1.00.									
b. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.									
c. This parameter is set to zero because it is redundant.									

Table 5.60: Parameter estimates for overall gradings 4 and 5 (multinomial logistic regression and considering initial condition gradings)

Parameter Estimates									
Overall_Grading[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
4	Intercept	-9.239	2.076	19.8	1	.000			
	Pipe_Diameter_in	0.137	0.072	3.636	1	0.057	1.147	0.996	1.321
	Age_of_Pipe_years	0.136	0.016	69.589	1	.000	1.146	1.11	1.183
	Pipe_Slope	0.19	0.128	2.211	1	0.137	1.209	0.941	1.553
	Length_of_Pipe_ft	0	0.001	0.067	1	0.796	1	0.999	1.001
	Average_Velocity_ft_per_s	-0.037	0.229	0.026	1	0.872	0.964	0.616	1.509
	Average_Flow_Depth	-0.212	0.201	1.116	1	0.291	0.809	0.545	1.199
	Average_Flow_in_Pipe_percent_full	0.08	0.051	2.454	1	0.117	1.083	0.98	1.197
	[Pipe_Material=CCFRPM]	-18.587	8817.956	0	1	0.998	8.47E-09	0	.[b]
	[Pipe_Material=FRP]	-2.23	1.072	4.325	1	0.038	0.108	0.013	0.879
	[Pipe_Material=PVC]	-1.017	0.558	3.324	1	0.068	0.362	0.121	1.079
	[Pipe_Material=RCP]	-0.437	0.673	0.421	1	0.516	0.646	0.173	2.417
	[Pipe_Material=VCP]	0[c]	.	.	0	.	.	.	.
5	Intercept	-14.925	4.13	13.062	1	.000			
	Pipe_Diameter_in	0.369	0.17	4.746	1	0.029	1.447	1.038	2.017
	Age_of_Pipe_years	0.111	0.04	7.539	1	0.006	1.117	1.032	1.21
	Pipe_Slope	-0.444	2.233	0.039	1	0.842	0.642	0.008	51.077
	Length_of_Pipe_ft	0.001	0.001	0.592	1	0.442	1.001	0.998	1.004
	Average_Velocity_ft_per_s	-0.306	0.577	0.281	1	0.596	0.736	0.238	2.283
	Average_Flow_Depth	-1.091	0.545	4.004	1	0.045	0.336	0.115	0.978
	Average_Flow_in_Pipe_percent_full	0.236	0.081	8.477	1	0.004	1.266	1.08	1.484
	[Pipe_Material=CCFRPM]	-17.177	0	.	1	.	3.47E-08	3.47E-08	3.47E-08
	[Pipe_Material=FRP]	-17.145	2451.012	0	1	0.994	3.58E-08	0	.[b]
	[Pipe_Material=PVC]	-0.868	1.511	0.33	1	0.566	0.42	0.022	8.12
	[Pipe_Material=RCP]	-0.544	1.666	0.106	1	0.744	0.581	0.022	15.213
	[Pipe_Material=VCP]	0[c]	.	.	0	.	.	.	.
a. The reference category is: 1.00.									
b. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.									
c. This parameter is set to zero because it is redundant.									

Table 5.61: Classification table obtained for multinomial logistic regression (considering initial condition gradings)

Classification						
Observed	Predicted					
	1	2	3	4	5	Percent Correct
1	602	1	1	17	1	96.8%
2	38	5	1	11	0	9.1%
3	55	3	3	17	0	3.8%
4	14	0	2	43	0	72.9%
5	2	0	0	3	1	16.7%
Overall Percentage	86.7%	1.1%	0.9%	11.1%	0.2%	79.8%

5.1.10 Ordinal regression using Logit link function and considering initial condition gradings

Results associated with ordinal regression utilizing Logit link function and considering sewer pipes are initially in perfect condition gradings, i.e. overall condition grading 1, are presented in the following tables. Based on these results, it is observed that the Goodness-of-Fit test (Pearson and Deviance) and likelihood ratio test available within model fitting information table both yield satisfactory results. Furthermore, based on test of parallel lines, it is realized that the proportional odds assumption is also validated. The age of sewer pipe has been found as a significant predictor in this deterioration model as well.

Table 5.62: Table containing model fitting information (ordinal regression using Logit link function and considering initial condition gradings)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1377.578			
Final	1011.01	366.567	11	.000
Link function: Logit.				

Table 5.63: Results of Goodness-of-Fit tests (ordinal regression using Logit link function and considering initial condition gradings)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	2197.07	3265	1.000
Deviance	1011.01	3265	1.000
Link function: Logit.			

Table 5.64: Parameter estimates obtained based on ordinal regression using Logit link function and considering initial condition gradings

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	5.923	1.077	30.256	1	.000	3.813	8.034
	[Overall_Grading = 2.00]	6.583	1.084	36.855	1	.000	4.457	8.708
	[Overall_Grading = 3.00]	7.876	1.1	51.25	1	.000	5.719	10.032
	[Overall_Grading = 4.00]	10.656	1.176	82.063	1	.000	8.351	12.962
Location	Age_of_Pipe_years	0.093	0.007	196.858	1	.000	0.08	0.106
	Average_Flow_Depth	-0.217	0.109	3.944	1	0.047	-0.431	-0.003
	Length_of_Pipe_ft	0	0	0.553	1	0.457	-0.001	0
	Pipe_Slope	0.028	0.086	0.105	1	0.746	-0.14	0.196
	Average_Flow_in_Pipe_percent_full	0.076	0.028	7.502	1	0.006	0.022	0.131
	Pipe_Diameter_in	0.121	0.04	9.269	1	0.002	0.043	0.199
	Average_Velocity_ft_per_s	-0.111	0.124	0.802	1	0.371	-0.353	0.131
	[Pipe_Material=CCFRPM]	-20.826	0	.	1	.	-20.826	-20.826
	[Pipe_Material=FRP]	-2.091	0.547	14.617	1	.000	-3.164	-1.019
	[Pipe_Material=PVC]	-0.02	0.271	0.005	1	0.942	-0.55	0.511
	[Pipe_Material=RCP]	-0.394	0.398	0.98	1	0.322	-1.175	0.386
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Logit.								
a. This parameter is set to zero because it is redundant.								

Table 5.65: Results of test of parallel lines obtained based on ordinal regression using Logit link function and considering initial condition gradings

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	1011.011			
General	973.157[b]	37.854[c]	33	0.257
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Logit.				
b. The log-likelihood value cannot be further increased after maximum number of step-halving.				
c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.				

5.1.11 Ordinal regression using Probit link function and considering initial condition gradings

Considering initial condition gradings of sewer pipes (overall condition grading of 1 for all assets), and by using Probit link function, the following results are obtained. The Goodness-of-Fit test (Pearson and Deviance) as well as the likelihood ratio test available within model fitting information table both presented desirable results. However, when validating proportional odds assumption through test of parallel lines, satisfactory results were not obtained and therefore this assumption cannot be validated. By eliminating the average flow in pipe (%full) from the deterioration model, in addition to yielding satisfactory results associated with the Goodness-of-Fit test (Pearson and Deviance) and the likelihood ratio test available within model fitting information table, the result obtained from test of parallel lines also indicated that the proportional odds assumption was validated too. The tables below, demonstrate the outcome of both models with and without average flow in pipe (%full). Moreover, it is realized that the age of sewer pipe is a significant predictor in these models.

Model with all independent variables:

The following tables are associated with the ordinal regression model with Probit function developed utilizing all independent variables including average flow in pipe (%full). Through test of parallel lines, the proportional odds assumption could not be validated.

Table 5.66: Table containing model fitting information (ordinal regression using Probit link function and considering initial condition gradings: full model)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1377.578			
Final	.000	1377.578	11	.000
Link function: Probit.				

Table 5.67: Results of Goodness-of-Fit tests (ordinal regression using Probit link function and considering initial condition gradings: full model)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	2600.21	3265	1.000
Deviance	1003.72	3265	1.000
Link function: Probit.			

Table 5.68: Table containing Pseudo R-Square values (ordinal regression using Probit link function and considering initial condition gradings: full model)

Pseudo R-Square	
Cox and Snell	0.814
Nagelkerke	1.000
McFadden	1.000
Link function: Probit.	

Table 5.69: Parameter estimates obtained based on ordinal regression using Probit link function and considering initial condition gradings: full model

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	3.631	0.606	35.907	1	.000	2.444	4.819
	[Overall_Grading = 2.00]	4.006	0.609	43.269	1	.000	2.812	5.2
	[Overall_Grading = 3.00]	4.74	0.617	59.078	1	.000	3.531	5.948
	[Overall_Grading = 4.00]	6.16	0.646	90.782	1	.000	4.892	7.427
Location	Age_of_Pipe_years	0.055	0.004	238.966	1	.000	0.048	0.061
	Average_Flow_Depth	-0.142	0.063	5.123	1	0.024	-0.264	-0.019
	Length_of_Pipe_ft	0	0	0.77	1	0.38	-0.001	0
	Pipe_Slope	0.012	0.049	0.066	1	0.797	-0.083	0.108
	Average_Flow_in_Pipe_percent_full	0.049	0.016	9.373	1	0.002	0.018	0.08
	Pipe_Diameter_in	0.074	0.023	10.784	1	0.001	0.03	0.118
	Average_Velocity_ft_per_s	-0.056	0.069	0.671	1	0.413	-0.192	0.079
	[Pipe_Material=CCFRPM]	-4.482	110.391	0.002	1	0.968	-220.845	211.88
	[Pipe_Material=FRP]	-1.151	0.294	15.351	1	.000	-1.726	-0.575
	[Pipe_Material=PVC]	0.066	0.152	0.185	1	0.667	-0.233	0.364
	[Pipe_Material=RCP]	-0.202	0.225	0.81	1	0.368	-0.643	0.238
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Probit.								
a. This parameter is set to zero because it is redundant.								

Table 5.70: Results of test of parallel lines obtained based on ordinal regression using Probit link function and considering initial condition gradings: full model

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	.000			
General	33.729[b]	.[c]	33	.
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Probit.				
b. The log-likelihood value cannot be further increased after maximum number of step-halving.				
c. The log-likelihood value of the general model is smaller than that of the null model. This is because convergence cannot be attained or ascertained in estimating the general model. Therefore, the test of parallel lines cannot be performed.				

Model without average flow in pipe (%full).:

The results associated with the ordinal regression model with Probit function developed without average flow in pipe (%full) are presented in the following tables. The proportional odds assumption is validated in this model, and the age of sewer pipe is observed to be a significant independent variable.

Table 5.71: Table containing model fitting information (ordinal regression using Probit link function and considering initial condition gradings)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1377.58			
Final	.000	1377.58	10	.000
Link function: Probit.				

Table 5.72: Results of Goodness-of-Fit tests (ordinal regression using Probit link function and considering initial condition gradings)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	2784.73	3266	1.000
Deviance	1012.33	3266	1.000
Link function: Probit.			

Table 5.73: Table containing Pseudo R-Square values for ordinal regression using Probit link function and considering initial condition gradings

Pseudo R-Square	
Cox and Snell	0.814
Nagelkerke	1.000
McFadden	1.000
Link function: Probit.	

Table 5.74: Parameter estimates obtained based on ordinal regression using Probit link function and considering initial condition gradings

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	2.078	0.327	40.369	1	.000	1.437	2.719
	[Overall_Grading = 2.00]	2.452	0.33	55.097	1	.000	1.804	3.099
	[Overall_Grading = 3.00]	3.177	0.339	87.739	1	.000	2.512	3.842
	[Overall_Grading = 4.00]	4.538	0.38	142.951	1	.000	3.794	5.282
Location	Age_of_Pipe_years	0.053	0.003	235.375	1	.000	0.046	0.06
	Average_Flow_Depth	0.039	0.02	3.771	1	0.052	0	0.077
	Pipe_Diameter_in	0.016	0.012	1.746	1	0.186	-0.008	0.041
	Pipe_Slope	0.01	0.049	0.039	1	0.843	-0.086	0.105
	Length_of_Pipe_ft	0	0	1.126	1	0.289	-0.001	0
	Average_Velocity_ft_per_s	-0.013	0.067	0.038	1	0.846	-0.145	0.119
	[Pipe_Material=CCFRPM]	-4.455	111.554	0.002	1	0.968	-223.098	214.188
	[Pipe_Material=FRP]	-1.03	0.29	12.602	1	.000	-1.599	-0.461
	[Pipe_Material=PVC]	-0.029	0.148	0.039	1	0.842	-0.319	0.261
	[Pipe_Material=RCP]	-0.125	0.221	0.32	1	0.572	-0.559	0.308
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Probit.								
a. This parameter is set to zero because it is redundant.								

Table 5.75: Results of test of parallel lines obtained based on ordinal regression using Probit link function and considering initial condition gradings

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	.000			
General	.000[b]	.000	30	1.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Probit.				
b. The log-likelihood value is practically zero. There may be a complete separation in the data. The maximum likelihood estimates do not exist.				

5.1.12 Ordinal regression using Complementary Log-Log link function and considering initial condition gradings

The following tables are obtained by considering initial condition gradings of assets (i.e. overall condition grading of 1 is assumed for all sewer pipes at the beginning of their service lives). Based on the following results, the Deviance Goodness-of-Fit test yielded satisfactory result; additionally, both likelihood ratio test available within model fitting information table as well as test of parallel lines were satisfactory too. Therefore, proportional odds assumption is validated for this model. These results also illustrate that the age of sewer pipe is a significant predictor of the model as well.

Table 5.76: Table containing model fitting information (ordinal regression using Complementary Log-Log link function and considering initial condition gradings)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1377.578			
Final	.000	1377.578	11	.000
Link function: Complementary Log-log.				

Table 5.77: Results of Goodness-of-Fit tests (ordinal regression using Complementary Log-Log link function and considering initial condition gradings)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	11474.448	3265	.000
Deviance	1010.252	3265	1.000
Link function: Complementary Log-log.			

Table 5.78: Table containing Pseudo R-Square values for ordinal regression using Complementary Log-Log link function and considering initial condition gradings

Pseudo R-Square	
Cox and Snell	0.814
Nagelkerke	1.000
McFadden	1.000
Link function: Complementary Log-log.	

Table 5.79: Parameter estimates obtained based on ordinal regression using Complementary Log-Log link function and considering initial condition gradings

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	3.159	0.556	32.236	1	.000	2.069	4.25
	[Overall_Grading = 2.00]	3.511	0.559	39.395	1	.000	2.415	4.608
	[Overall_Grading = 3.00]	4.18	0.569	53.905	1	.000	3.064	5.296
	[Overall_Grading = 4.00]	5.301	0.592	80.143	1	.000	4.141	6.462
Location	Age_of_Pipe_years	0.051	0.003	228.645	1	.000	0.044	0.058
	Average_Flow_Depth	-0.14	0.059	5.609	1	0.018	-0.255	-0.024
	Pipe_Diameter_in	0.069	0.021	10.799	1	0.001	0.028	0.11
	Pipe_Slope	-0.006	0.042	0.023	1	0.879	-0.089	0.076
	Length_of_Pipe_ft	0	0	1.642	1	0.2	-0.001	0
	Average_Velocity_ft_per_s	-0.006	0.059	0.011	1	0.916	-0.122	0.109
	Average_Flow_in_Pipe_percent_full	0.05	0.015	10.585	1	0.001	0.02	0.08
	[Pipe_Material=CCFRPM]	-1.166	0.4	8.488	1	0.004	-1.95	-0.382
	[Pipe_Material=FRP]	-1.002	0.242	17.102	1	.000	-1.477	-0.527
	[Pipe_Material=PVC]	0.139	0.132	1.107	1	0.293	-0.12	0.398
	[Pipe_Material=RCP]	-0.258	0.197	1.708	1	0.191	-0.644	0.129
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Complementary Log-log.								
a. This parameter is set to zero because it is redundant.								

Table 5.80: Results of test of parallel lines obtained based on ordinal regression using Complementary Log-Log link function and considering initial condition gradings

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	.000			
General	.000[b]	.000	33	1.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Complementary Log-log.				
b. The log-likelihood value is practically zero. There may be a complete separation in the data. The maximum likelihood estimates do not exist.				

5.1.13 Ordinal regression using Negative Log-Log link function and considering initial condition gradings

The results pertaining to ordinal regression utilizing Negative Log-Log link function and by considering perfect condition for the sewer pipes at the beginning of their service lives, are presented in this section. Based upon the outcome of this model, it is found that the Goodness-of-Fit test (Deviance and Pearson), likelihood ratio test available within model fitting information table and the test of parallel lines all yield satisfactory results. Hence, considering proportional odds assumption is also validated, this deterioration model seems to be suitable. Moreover, the independent variable denoting the age of assets is observed to be a significant variable of the model. The following tables show the results achieved based on this model.

Table 5.81: Table containing model fitting information (ordinal regression using Negative Log-Log link function and considering initial condition gradings)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1377.578			
Final	1020.842	356.736	11	.000
Link function: Negative Log-log.				

Table 5.82: Results of Goodness-of-Fit tests (ordinal regression using Negative Log-Log link function and considering initial condition gradings)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	2199.367	3265	1.000
Deviance	1020.842	3265	1.000
Link function: Negative Log-log.			

Table 5.83: Table containing Pseudo R-Square values for ordinal regression using Negative Log-Log link function and considering initial condition gradings

Pseudo R-Square	
Cox and Snell	0.353
Nagelkerke	0.434
McFadden	0.259
Link function: Negative Log-log.	

Table 5.84: Parameter estimates obtained based on ordinal regression using Negative Log-Log link function and considering initial condition gradings

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	4.378	0.779	31.608	1	.000	2.852	5.905
	[Overall_Grading = 2.00]	4.877	0.783	38.782	1	.000	3.342	6.412
	[Overall_Grading = 3.00]	5.897	0.792	55.468	1	.000	4.345	7.449
	[Overall_Grading = 4.00]	8.438	0.884	91.099	1	.000	6.705	10.17
Location	Age_of_Pipe_years	0.072	0.005	194.162	1	.000	0.062	0.082
	Average_Flow_Depth	-0.141	0.078	3.264	1	0.071	-0.293	0.012
	Pipe_Diameter_in	0.077	0.028	7.363	1	0.007	0.021	0.132
	Pipe_Slope	0.036	0.067	0.287	1	0.592	-0.095	0.166
	Length_of_Pipe_ft	0	0	0.385	1	0.535	-0.001	0
	Average_Velocity_ft_per_s	-0.202	0.091	4.893	1	0.027	-0.381	-0.023
	Average_Flow_in_Pipe_percent_full	0.049	0.02	6.19	1	0.013	0.01	0.087
	[Pipe_Material=CCFRPM]	-20.38	0	.	1	.	-20.38	-20.38
	[Pipe_Material=FRP]	-1.267	0.435	8.492	1	0.004	-2.12	-0.415
	[Pipe_Material=PVC]	0.066	0.204	0.106	1	0.745	-0.334	0.467
	[Pipe_Material=RCP]	0.064	0.287	0.05	1	0.823	-0.499	0.627
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Negative Log-log.								
a. This parameter is set to zero because it is redundant.								

Table 5.85: Results of test of parallel lines obtained based on ordinal regression using Negative Log-Log link function and considering initial condition gradings

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	1020.842			
General	994.760[b]	26.082[c]	33	0.798
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Negative Log-log.				
b. The log-likelihood value cannot be further increased after maximum number of step-halving.				
c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.				

5.1.14 Ordinal regression using Cauchit link function and considering initial condition gradings

In this section, the deterioration model is constructed through Cauchit link function and assuming sewer pipes are initially in perfect condition (i.e. overall condition grading 1). Based on the results presented in the following tables, it is observed that both the Goodness-of-Fit test (Deviance and Pearson) and the likelihood ratio test available within model fitting information table show satisfactory outcomes for this model. On the other hand, test of parallel lines indicate that the proportional odds assumption is not validated for this deterioration model. Moreover, it is observed that the age of sewer pipe is a significant predictor of this deterioration model. Although the proportional odds assumption was not satisfied in this model, however, for illustration and comparison purposes this model is presented herein.

Table 5.86: Table containing model fitting information (ordinal regression using Cauchit link function and considering initial condition gradings)

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1377.578			
Final	1079.734	297.844	11	.000
Link function: Cauchit.				

Table 5.87: Results of Goodness-of-Fit tests (ordinal regression using Cauchit link function: full model and considering initial condition gradings)

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	2473.949	3265	1.000
Deviance	1079.734	3265	1.000
Link function: Cauchit.			

Table 5.88: Table containing Pseudo R-Square values for ordinal regression using Cauchit link function and considering initial condition gradings

Pseudo R-Square	
Cox and Snell	0.305
Nagelkerke	0.374
McFadden	0.216
Link function: Cauchit.	

Table 5.89: Parameter estimates obtained based on ordinal regression using Cauchit link function and considering initial condition gradings

Parameter Estimates								
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Overall_Grading = 1.00]	2.832	1.183	5.732	1	0.017	0.514	5.15
	[Overall_Grading = 2.00]	3.612	1.195	9.133	1	0.003	1.27	5.955
	[Overall_Grading = 3.00]	5.346	1.238	18.643	1	.000	2.919	7.773
	[Overall_Grading = 4.00]	21.96	5.734	14.665	1	.000	10.721	33.199
Location	Age_of_Pipe_years	0.095	0.01	90.212	1	.000	0.075	0.115
	Average_Flow_Depth	0.033	0.114	0.083	1	0.773	-0.191	0.257
	Pipe_Diameter_in	0.043	0.042	1.008	1	0.315	-0.041	0.126
	Pipe_Slope	-0.111	0.135	0.669	1	0.413	-0.376	0.155
	Length_of_Pipe_ft	0	0	0.084	1	0.772	-0.001	0.001
	Average_Velocity_ft_per_s	-0.238	0.133	3.183	1	0.074	-0.499	0.023
	Average_Flow_in_Pipe_percent_full	0.001	0.03	0	1	0.985	-0.058	0.059
	[Pipe_Material=CCFRPM]	-2206.815	0	.	1	.	-2206.815	-2206.815
	[Pipe_Material=FRP]	-2.316	0.692	11.187	1	0.001	-3.673	-0.959
	[Pipe_Material=PVC]	-0.824	0.298	7.657	1	0.006	-1.408	-0.24
	[Pipe_Material=RCP]	-0.8	0.426	3.53	1	0.060	-1.635	0.035
	[Pipe_Material=VCP]	0[a]	.	.	0	.	.	.
Link function: Cauchit.								
a. This parameter is set to zero because it is redundant.								

Table 5.90: Results of test of parallel lines obtained based on ordinal regression using Cauchit link function and considering initial condition gradings

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	1079.734			
General	.000[b]	1079.734	33	.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.				
a. Link function: Cauchit.				
b. The log-likelihood value is practically zero. There may be a complete separation in the data. The maximum likelihood estimates do not exist.				

5.2 Results Based on LightGBM Method

In this section, the deterioration model developed based on LightGBM is presented. By using the test size of 20% in LightGBM model, the number of estimators equal to 1000, and the number of leaves equal to 31, and using the maximum depth of 5, the following tables demonstrate the confusion matrices for both training and testing sets of data. Furthermore, it should be noted that in developing LightGBM model, it is assumed that at the beginning of their service lives, sewer pipes are in perfect condition gradings.

Table 5.91: Confusion matrix for LightGBM (testing set)

	Predictions	
Observations	0	1
0	*153*	*3*
1	*7*	*1*

Table 5.92: Confusion matrix for LightGBM (training set)

	Predictions	
Observations	0	1
0	*599*	*0*
1	*0*	*57*

Once the confusion matrices of training and testing sets are obtained, the next step is to find the feature importance pertaining to the independent variables of the model. The graph demonstrating the feature importance associated with different features is presented below:

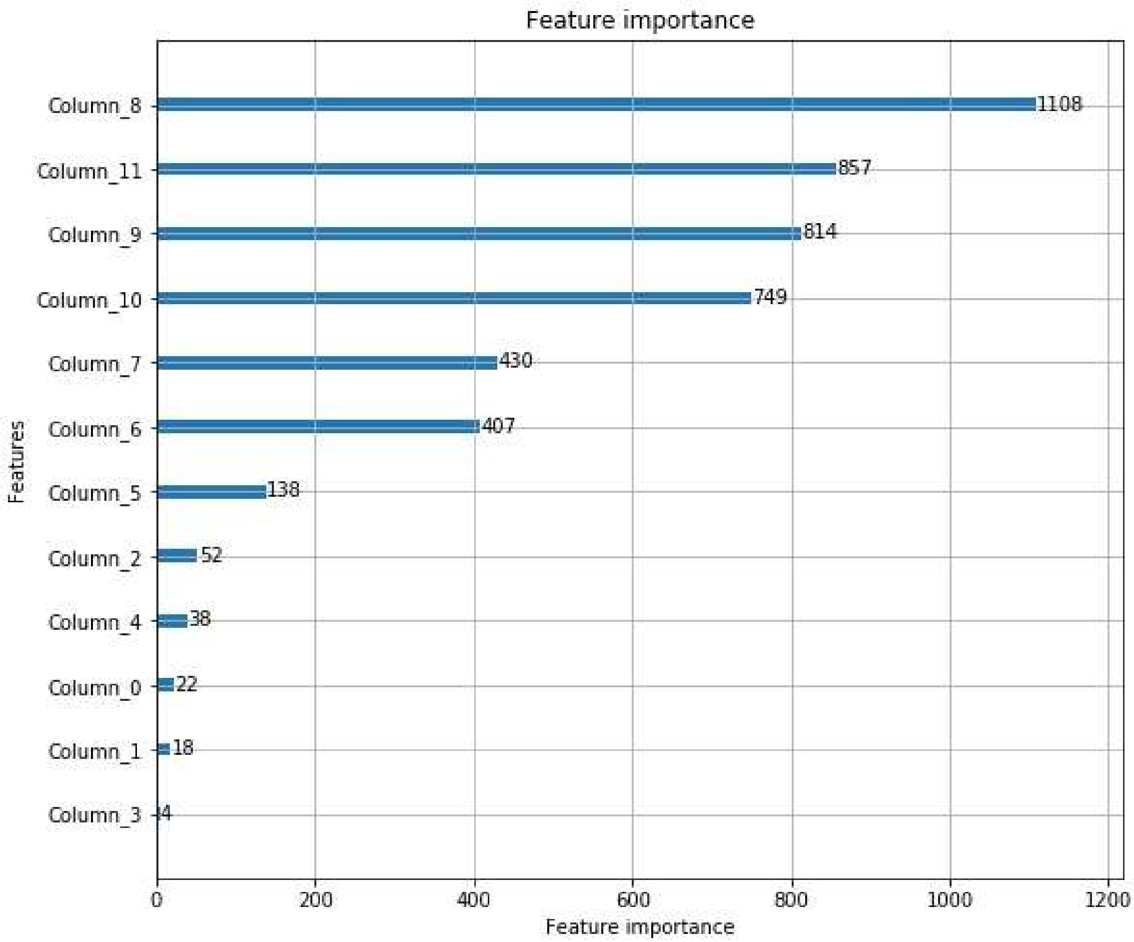

Figure 5.1: Feature importance associated with each predictor obtained utilizing LightGBM model

In the above figure, various independent variables of the developed model corresponding to each of the feature columns are as follows:

Table 5.93: Feature descriptions associated with different feature labels (LightGBM)

Feature Label	Feature Description
Column_0	CCFRPM
Column_1	FRP
Column_2	PVC
Column_3	RCP
Column_4	VCP
Column_5	Pipe Diameter (in)
Column_6	Pipe Age (years)
Column_7	Pipe Slope
Column_8	Pipe Length (ft)
Column_9	Average Flow Pipe % Full
Column_10	Average Velocity (ft/s)
Column_11	Average Flow Pipe Flow Depth

Based on the results obtained from feature importance, it is thus concluded that based on the model developed utilizing LightGBM, the most important independent variables of the models are in the following order:

Table 5.94: Table illustrating the order of feature descriptions from highest to lowest based on LightGBM

Feature Label	Feature Description
Column_8	Pipe Length (ft)
Column_11	Average Flow Pipe Flow Depth
Column_9	Average Flow Pipe % Full
Column_10	Average Velocity (ft/s)
Column_7	Pipe Slope
Column_6	Pipe Age (years)
Column_5	Pipe Diameter (in)
Column_2	PVC
Column_4	VCP
Column_0	CCFRPM
Column_1	FRP
Column_3	RCP

Once the significance of each individual independent variables of the model are obtained, next the receiver operating characteristic (ROC) curve of the model is presented.

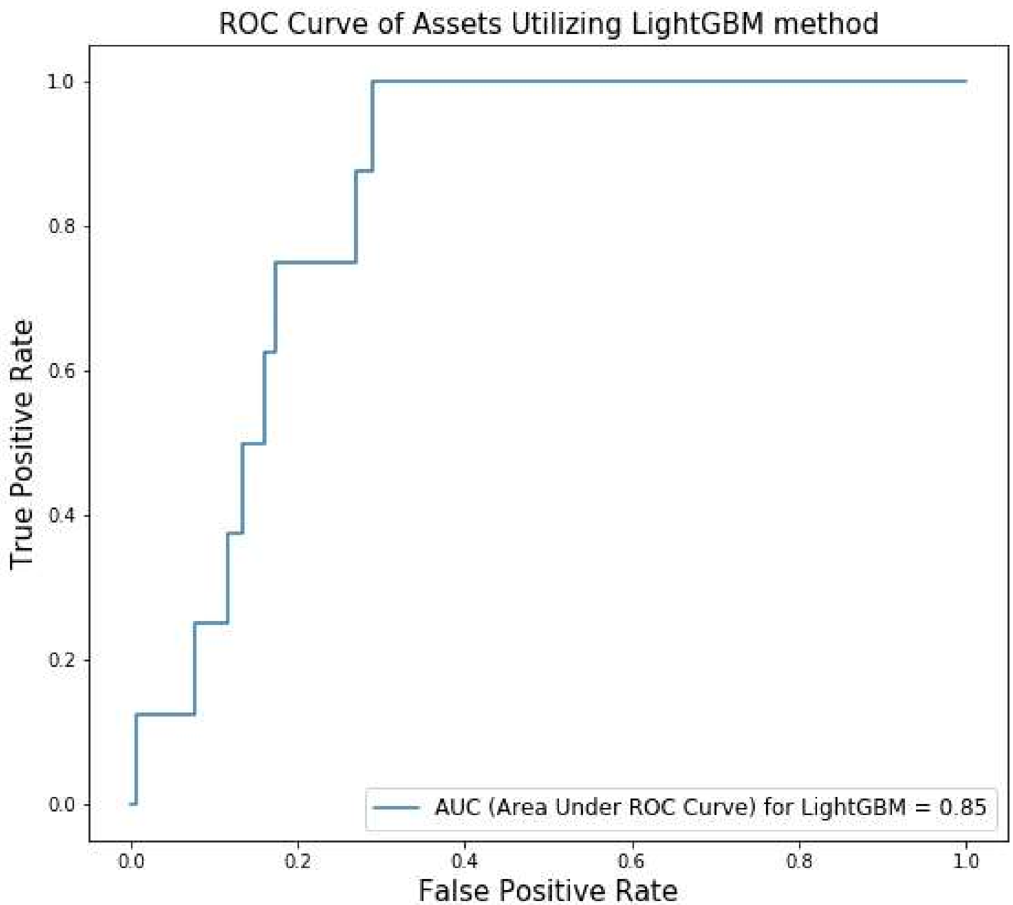

Figure 5.2: ROC curve obtained based on LightGBM model

The corresponding values of false positive rates and true positive rates for LightGBM method are presented in the table below.

Table 5.95: True positive and false positive rates obtained based on LightGBM

LightGBM Model	
False Positive Rate	True Positive Rate
0	0
0.00641026	0
0.00641026	0.125
0.0769231	0.125
0.0769231	0.25
0.115385	0.25
0.115385	0.375
0.134615	0.375
0.134615	0.5
0.160256	0.5
0.160256	0.625
0.173077	0.625
0.173077	0.75
0.269231	0.75
0.269231	0.875
0.288462	0.875
0.288462	1
0.794872	1
0.807692	1
1	1

It is observed that as illustrated in the receiver operating characteristic curve obtained based on LightGBM model, the area under the ROC curve is equal to 0.85.

5.3 Results Based on CatBoost Method

This section demonstrates the model developed based on CatBoost approach. Similar to LightGBM, the test size utilized in developing CatBoost model is also 20%. When using

CatBoost, the number of iterations is equal to 500, and the values corresponding to depth and the learning rate of the model are set to 5 and 0.2, respectively. The following tables illustrate the confusion matrices corresponding to training set as well as testing set. Similar to LightGBM model, when developing CatBoost model, it is assumed that the sewer pipes initially (i.e. at the beginning of their service lives) have perfect condition gradings.

Table 5.96: Confusion matrix for CatBoost (testing set)

	Predictions	
Observations	0	1
0	*151*	*5*
1	*7*	*1*

Table 5.97: Confusion matrix for CatBoost (training set)

	Predictions	
Observations	0	1
0	*599*	*0*
1	*0*	*57*

In CatBoost approach, similar to LightGBM model, after obtaining the confusion matrices pertaining to training and testing sets, the feature importance associated with each of the predictors of the model is presented in this section. The graph illustrated below contains the feature importance for various features utilized in CatBoost model.

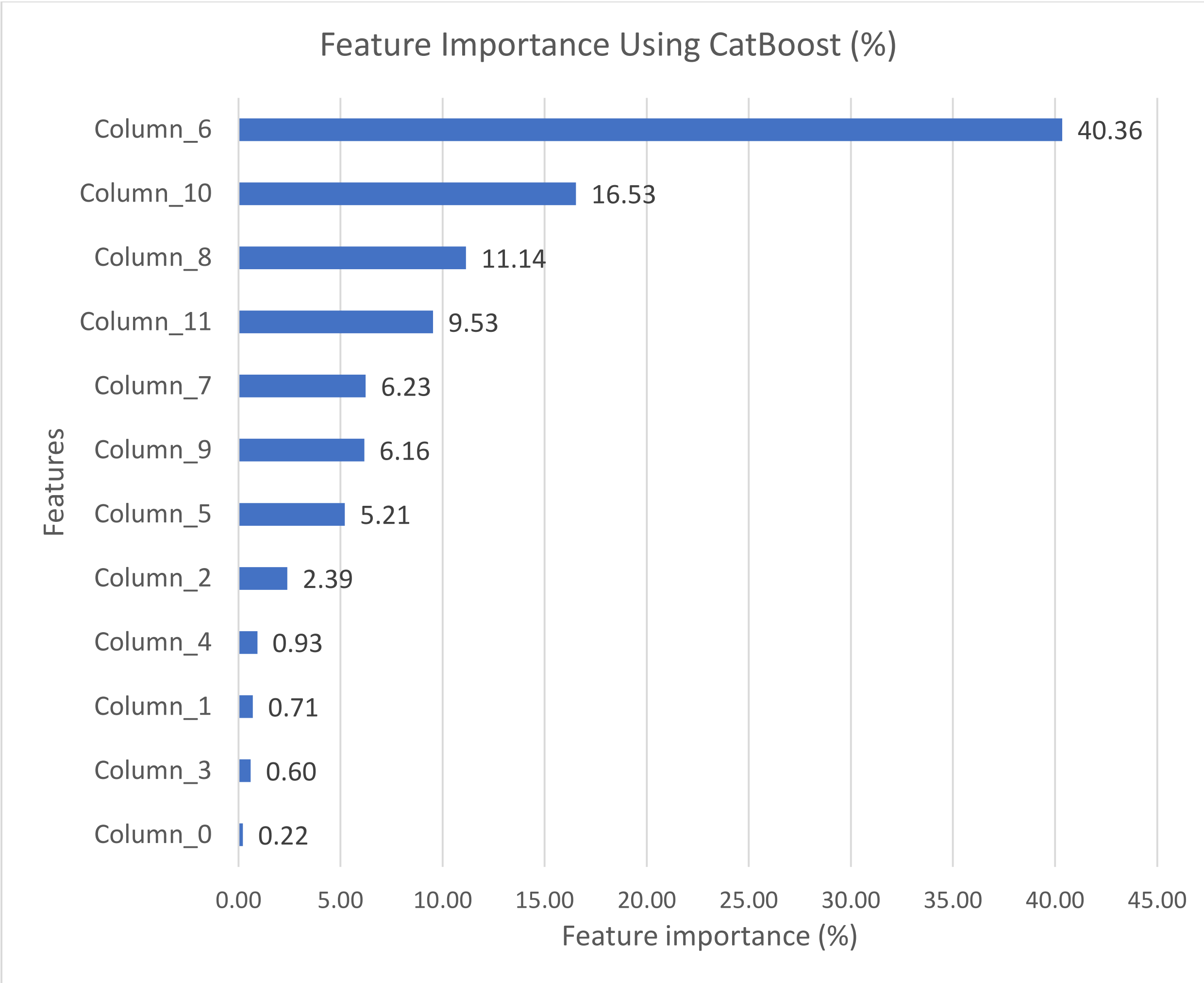

Figure 5.3: Feature importance associated with each predictor obtained based on CatBoost model

In this graph, each of the columns correspond to a different independent variable of the CatBoost model. The predictors and their associated columns are the same as presented earlier for LightGBM model. It is observed that based on the feature importance of each of the independent variables used in the CatBoost model, the predictors with highest to lowest values of feature importance are as presented in the following table:

Table 5.98: Table illustrating the order of feature descriptions from highest to lowest based on CatBoost model

Feature Label	Feature Description	Feature Importance (%)
Column_6	Pipe Age (years)	40.3572
Column_10	Average Velocity (ft/s)	16.5263
Column_8	Pipe Length (ft)	11.1402
Column_11	Average Flow Pipe Flow Depth	9.53081
Column_7	Pipe Slope	6.22713
Column_9	Average Flow Pipe % Full	6.16123
Column_5	Pipe Diameter (in)	5.20763
Column_2	PVC	2.39256
Column_4	VCP	0.929102
Column_1	FRP	0.705604
Column_3	RCP	0.603758
Column_0	CCFRPM	0.218479

The receiver operating characteristic (ROC) curve obtained based on CatBoost model is presented in the following figure.

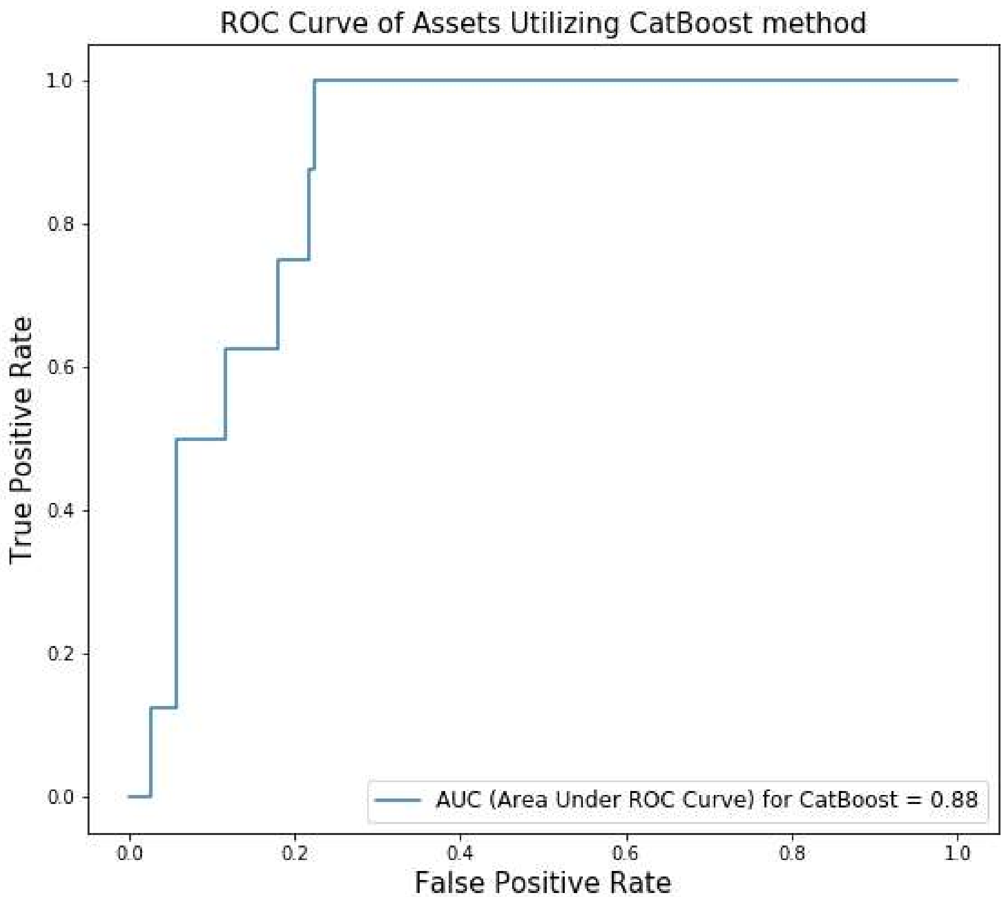

Figure 5.4: ROC curve associated with CatBoost model

The following table illustrates the points (false positive rates and true positive rates) corresponding to the receiver operating characteristic curve obtained based on CatBoost model:

Table 5.99: True positive and false positive rates obtained based on CatBoost model

CatBoost Model	
False Positive Rate	True Positive Rate
0	0
0.00641026	0
0.025641	0
0.025641	0.125
0.0576923	0.125
0.0576923	0.5
0.115385	0.5
0.115385	0.625
0.179487	0.625
0.179487	0.75
0.217949	0.75
0.217949	0.875
0.224359	0.875
0.224359	1
1	1

Based on the observations made in the receiver operating characteristic curve obtained by utilizing CatBoost approach, the area under this curve has a value of 0.88.

When comparing CatBoost and LightGBM, it is observed that the area under the ROC curve for LightGBM is equal to 0.85, whereas this area for CatBoost is equal to 0.88. Therefore, this demonstrates that when comparing these models based on their corresponding area under the ROC curve, CatBoost shows a better performance in modeling the set of data.

5.4 Comparing LightGBM and CatBoost models

In this section, considering the observations made based on the results pertaining to feature importance and receiver operating characteristic curves, LightGBM and CatBoost are compared.

5.4.1 Feature importance

Based on the models constructed using LightGBM and CatBoost, the significant features pertaining to each model are compared in the following table. In this table, it is observed that in CatBoost model, the feature which has the highest feature importance is the age of sewer pipe, whereas based on LightGBM approach, pipe length has the highest feature importance. However, based on both methods features such as pipe slope, pipe dimeter, PVC, and VCP are ranked similar among all features.

Table 5.100: Table comparing the order of feature descriptions (highest to lowest) for LightGBM and CatBoost models

LightGBM Model		CatBoost Model	
Feature Label	Feature Description	Feature Label	Feature Description
Column_8	Pipe Length (ft)	Column_6	Pipe Age (years)
Column_11	Average Flow Pipe Flow Depth	Column_10	Average Velocity (ft/s)
Column_9	Average Flow Pipe % Full	Column_8	Pipe Length (ft)
Column_10	Average Velocity (ft/s)	Column_11	Average Flow Pipe Flow Depth
Column_7	Pipe Slope	Column_7	Pipe Slope
Column_6	Pipe Age (years)	Column_9	Average Flow Pipe % Full
Column_5	Pipe Diameter (in)	Column_5	Pipe Diameter (in)
Column_2	PVC	Column_2	PVC
Column_4	VCP	Column_4	VCP
Column_0	CCFRPM	Column_1	FRP
Column_1	FRP	Column_3	RCP
Column_3	RCP	Column_0	CCFRPM

5.4.2 ROC comparison

Based on the results associated with receiver operating characteristic curves for both LightGBM and CatBoost models, and as shown in the following figure , it is observed that except for the beginning part of the graphs and a small portion after that, CatBoost seems to be a more suitable model for the sewer pipe data set. In other words, when the false positive rate is less than 0.0256 and when the false positive rate is between 0.1731 and 0.1795, LightGBM seems to be a more suitable model comparted to CatBoost; however, in all other cases, the model developed through CatBoost will be a more suitable model for sewer pipe data set.

Furthermore, the area under the receiver operating characteristic curves for LightGBM and CatBoost are 0.85 and 0.88, respectively. The greater area under the receiver operating characteristic curve in CatBoost model demonstrates a more suitable model for sewer pipe data set.

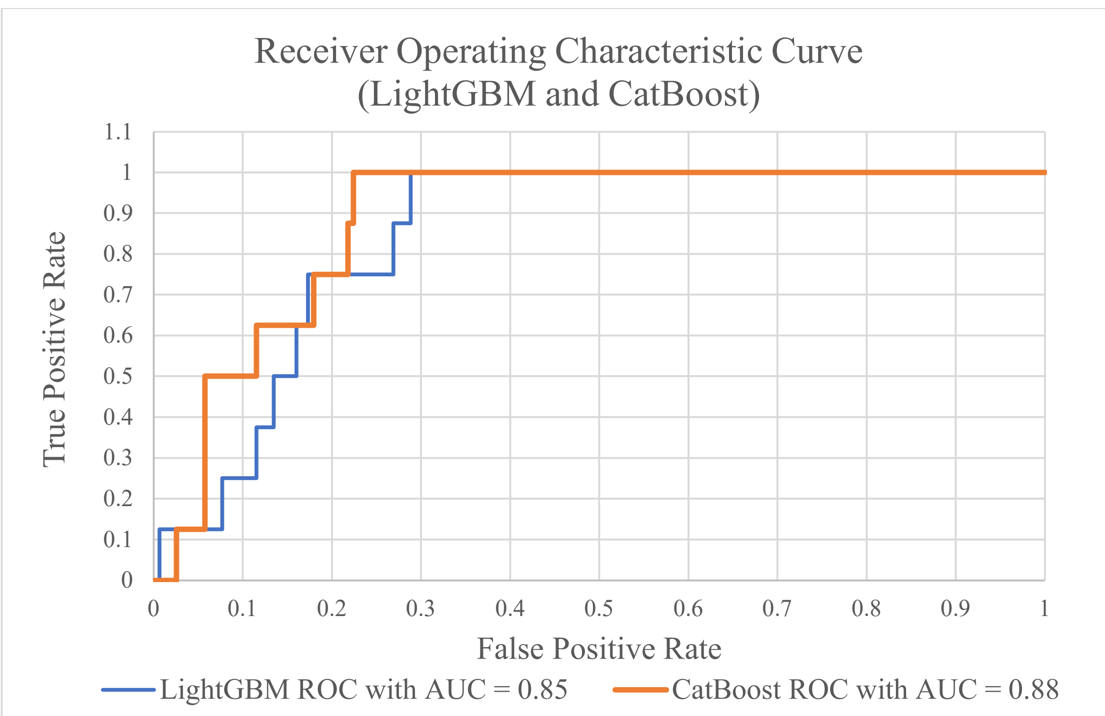

Figure 5.5: Receiver operating characteristic curves for LightGBM and CatBoost models

5.5 Results of Service Life of Pipe Using LightGBM and CatBoost

The probability of failure with respect to age of a vitrified clay sewer pipe obtained based on both LightGBM model and CatBoost model is presented in the following figure.

Figure 5.6: Probability of failure with respect to age of sewer pipe obtained based on LightGBM and CatBoost models

As observed in the above figure, the service life obtained for this vitrified clay pipe is equal to 44 years utilizing LightGBM as well as CatBoost model. However, in parts of this figure, the probability of failure with respect to age obtained based on LightGBM model is observed to be greater compared to the probability of failure calculated based on CatBoost approach.

In case there are more available data representing the condition of assets for various ages, the abrupt transitions in probability values which are observed in the above figure illustrating the probability of failure with respect to age of asset, and obtained through both

LightGBM as well as Catboost approaches, can become smoother transitions. However, herein, since the conditions of assets were available only for one age, therefore, in the graphs demonstrating the probability of failure with respect to age of asset, and obtained based on LightGBM and CatBoost, jumps in the probability values are observed.

In the following section, the results associated with statistical models are included as well.

5.6 Results of Service Life of Pipe Using Artificial Intelligence Based and Statistical models

In the following figure, the results obtained based on binomial and multinomial logistic regressions are included as well. It is observed that the service lives obtained through binomial logistic regression and multinomial logistic regression are 48 and 47 years, respectively.

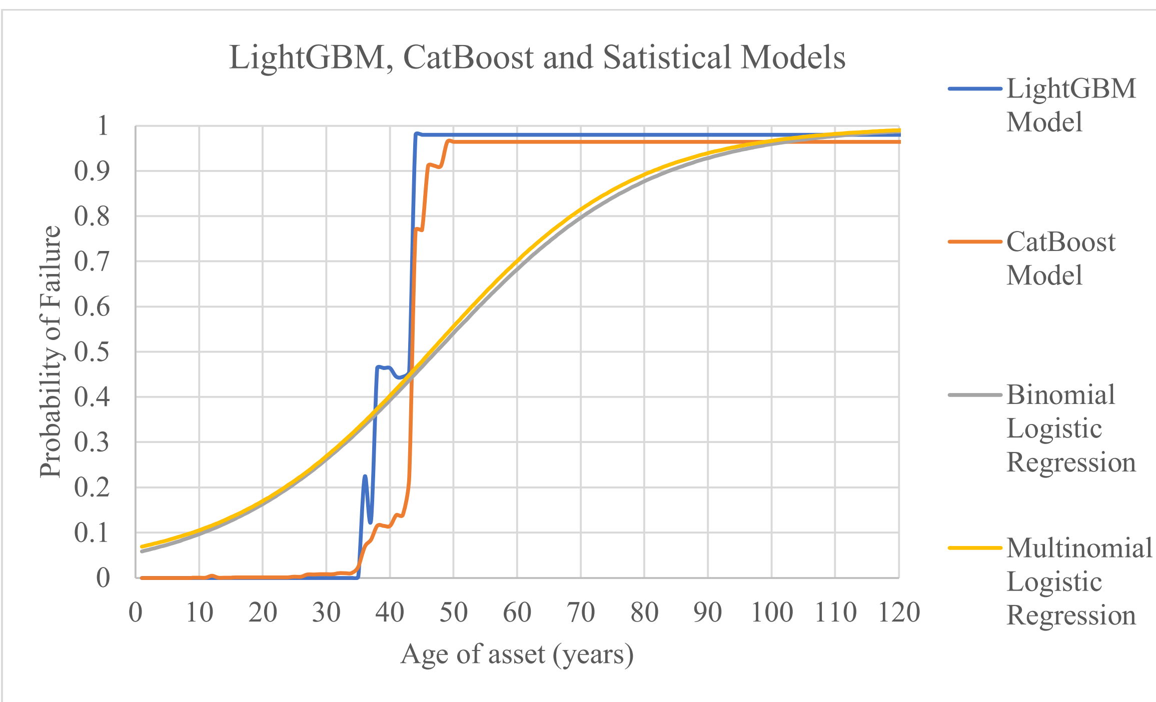

Figure 5.7: Probability of failure with respect to age of sewer pipe (including binomial and multinomial logistic regressions)

Furthermore, results pertaining to binomial logistic regression and multinomial logistic regression considering initial condition ratings of the sewer pipe are illustrated in the following figure as well. In these two additional models, it is assumed that at the beginning of their service lives, the sewer pipes are at perfect conditions.

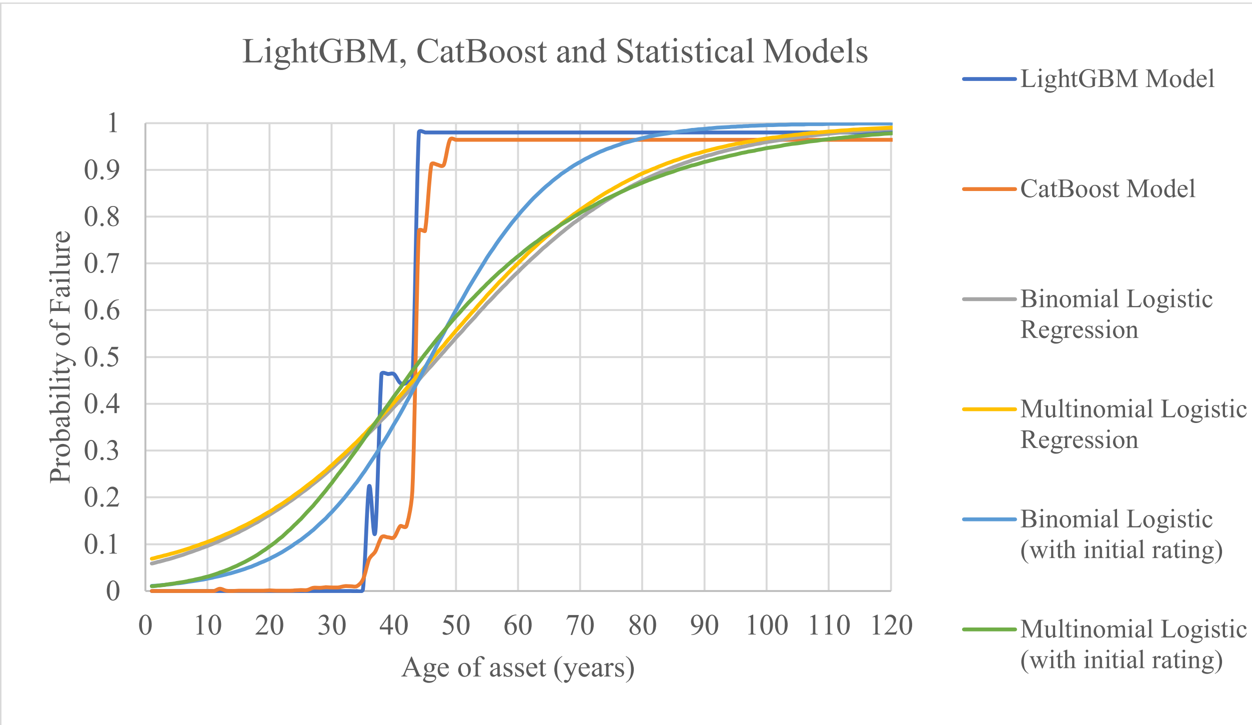

Figure 5.8: Probability of failure with respect to age of sewer pipe (including binomial and multinomial logistic regressions with initial condition grading)

Based on the models developed through binomial logistic regression and multinomial logistic regression, and assuming that the sewer pipes were initially in perfect condition ratings, the service life of the asset is observed to be 46 and 45 years, respectively.

Ordinal regression by using the logit function as the associated link function and by considering the initial rating of sewer pipes is presented below as well. Based on this figure, it is observed that the service life obtained based on ordinal regression with logit link function is equal to 46 years.

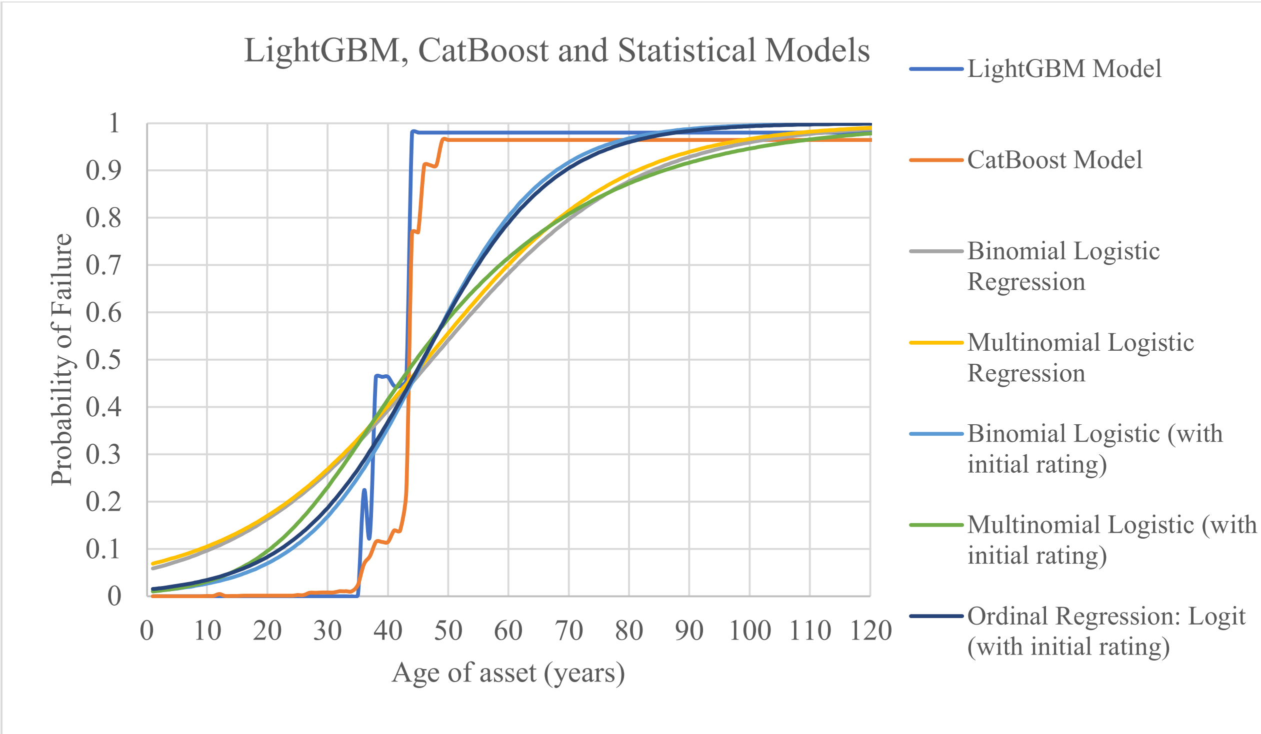

Figure 5.9: Probability of failure with respect to age of sewer pipe (including ordinal regression)

5.7 Effect of population on service life of gravity sewer pipes

In order to consider the effect of population growth, the annual rate of population growth is assumed to be 0.01. This assumption is solely for the purpose of illustrating the effect of population growth on the probability of failure of sewer pipes throughout their age as well as its impact on the service life of sewer pipes. The effect of increase in population is expected to be observed through the volume of the sewer flow which enters the sewer pipe. Therefore, considering this occurrence in sewer pipes the impact of increase in population could be implemented on the probability of failure of sewer pipes with respect to their age as well as their associated service lives.

5.7.1 Effect of population on service life of gravity sewer pipes utilizing artificial intelligence-based models

The following figure demonstrates the effect of population growth on the probability of failure of a sewer pipe based on CatBoost modeling. In this figure, the asset is made up of vitrified clay. Based on this figure it is observed that the probability of failure of asset with respect to age of sewer pipe is subjected to increase. Furthermore, it is observed that the service life of sewer pipe without consideration of population growth is 44 years, whereas by considering the population growth the service life of this asset is decreased to 42 years.

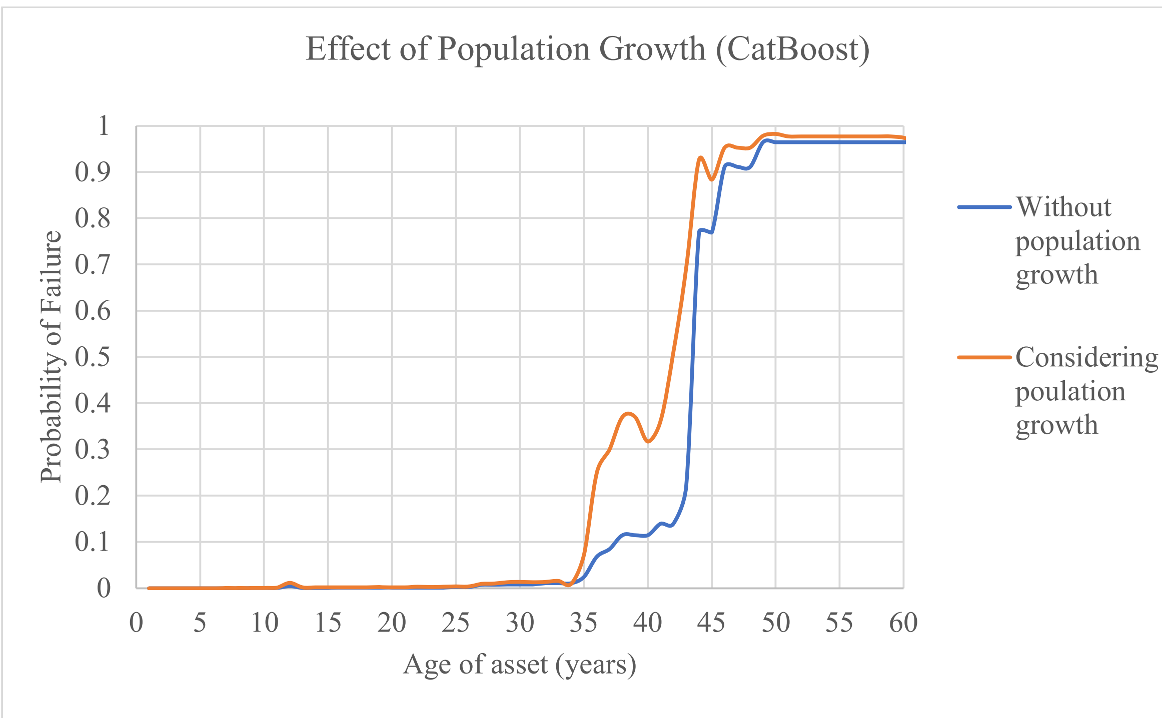

Figure 5.10: Effect of population growth on probability of failure (utilizing CatBoost model)

Additionally, the figure below, demonstrates the impact of population growth for the vitrified clay pipe when LightGBM model is used instead.

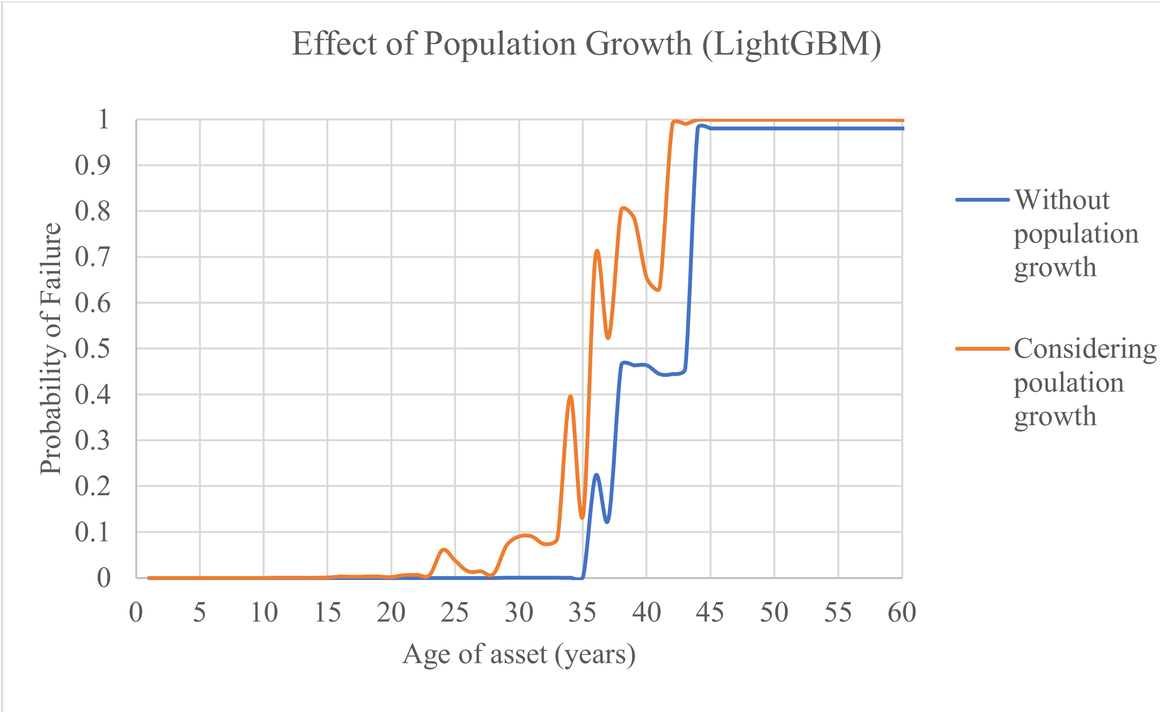

Figure 5.11: Influence of population growth on probability of failure (based on LightGBM model)

Based on the observation made from the LightGBM model, similar to CatBoost approach, in comparison to the LightGBM model without consideration of population growth, when the population growth is taken into account, the probability of failure of sewer pipe is thus increased and therefore, the service life of the sewer pipe is subjected to reduction. Herein when the population growth is not considered, based on LightGBM model, the service life of the sewer pipe is 44 years; however, when the population growth is considered for the sewer pipe, the service life obtained based on LightGBM model is decreased to 36 years.

5.7.2 Effect of population on service life of gravity sewer pipes utilizing statistical models

The following figure illustrates binomial logistic regression model for a vitrified clay pipe, with and without consideration of population growth (considering annual rate of growth is equal to 0.01). Based on this figure, it is observed that the probability of failure with respect to age of sewer pipe is increased when population growth is taken into consideration. Moreover, it is observed that due to effect of population growth, the service life of this sewer pipe is decreased from 52 years to 51 years.

Figure 5.12: Effect of population growth on probability of failure with respect to age of asset (Binomial Logistic Regression)

Similarly, the following figure shows multinomial logistic regression model of the pipe. As it is observed in this figure, similar to binomial logistic regression model, the probability of failure of the VCP sewer pipe is increased when population growth is taken into consideration. Additionally, by considering the population growth, the service life of this sewer pipe is subjected to reduction from 51 years to 40 years.

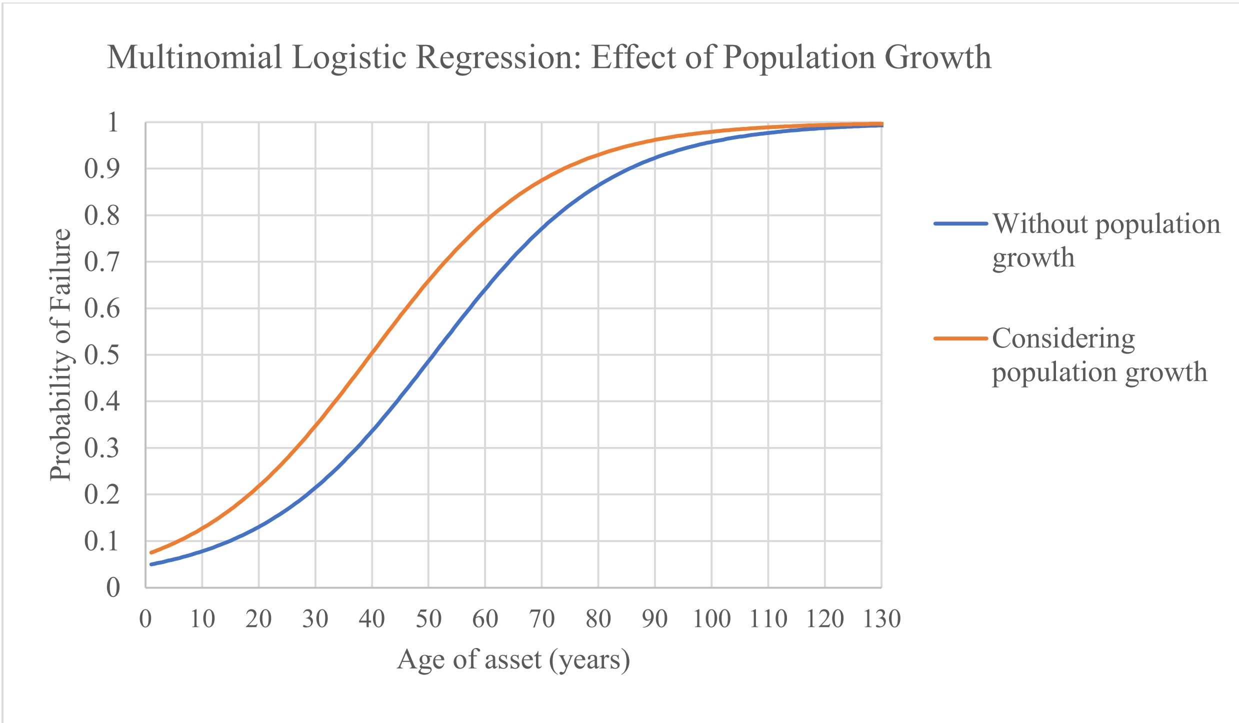

Figure 5.13: Influence of population growth on probability of failure (Multinomial Logistic Regression)

Next, by assuming that the pipe is in perfect condition at the beginning of its service life, and by using binomial logistic regression the following figure presents the probability of failure of the pipe with respect to age of asset. This figure shows that, similar to previous observations, when the increase in population is considered, the probability of failure and the service life of the VCP sewer pipe are subjected to increase and reduction, respectively; It is further realized that the service life of this asset with and without the effect of population growth has values of 48 and 49 years, respectively.

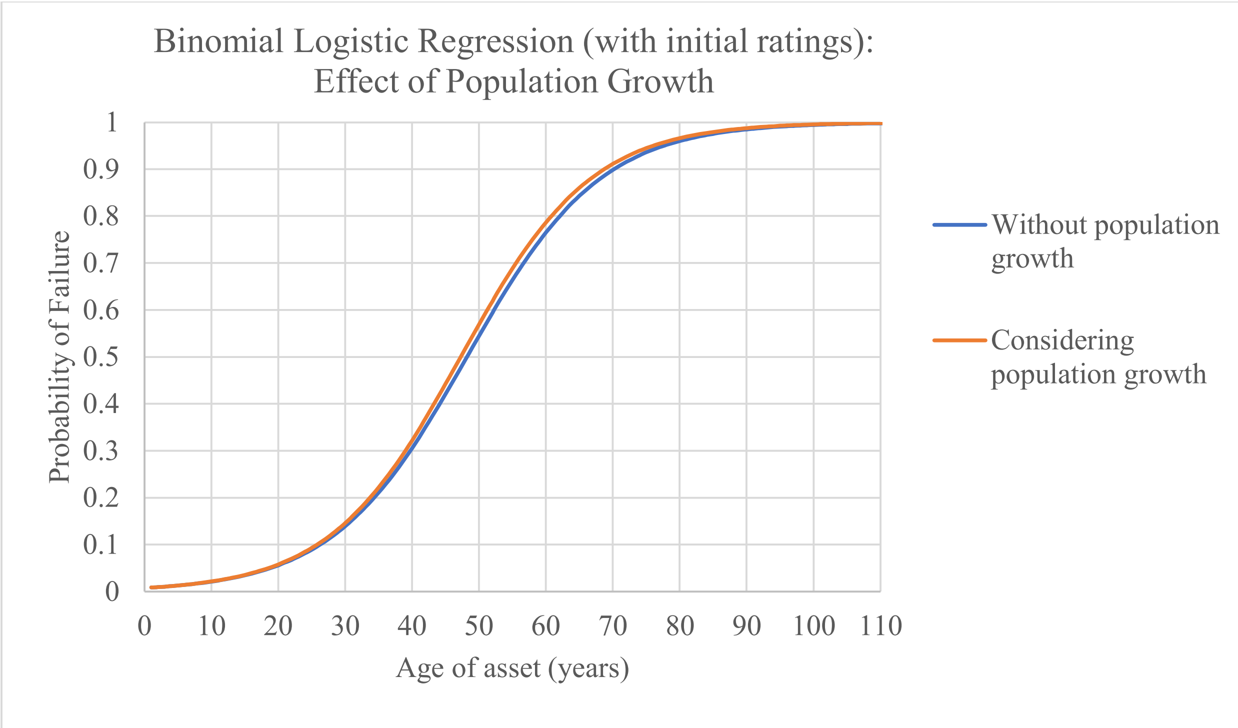

Figure 5.14: Effect of population growth on probability of failure with respect to age (Binomial Logistic Regression considering initial ratings)

Utilizing multinomial logistic regression and assuming that the asset has perfect condition at the beginning of its service life, the probability of failure of the sewer pipe with respect to its age is shown in the following figure. As seen in this figure, effect of population growth includes increase in probability of failure of the VCP pipe as well as decrease in service life. The values obtained for service life of this asset with and without the effect of population growth are equal to 40 and 48 years, respectively.

Figure 5.15: Influence of population growth on probability of failure (Multinomial Logistic Regression considering initial ratings)

Results pertaining to modeling the asset via ordinal regressions with logit, complementary log-log, and probit functions as the corresponding link functions are presented in the following figures. In these models, it is assumed that the condition of the pipe at the beginning of its service has been perfect. When taking the population growth into consideration, the probability of failure of the sewer pipe with respect to age of the pipe is increased in the ordinal regressions as well. When the logit function is used as the link function, the service life of the asset without taking into account the population growth has a value of 49 years. However, when the population growth is considered, the service life of the pipe is decreased to 46 years. When the link function of the ordinal regression is complementary log-log function, the service life with and without considering the population growth are 39 and 48 years, respectively. Using Probit function as the link function of the regression and when the population growth is accounted for, the service life is decreased from 49 year to 42 years.

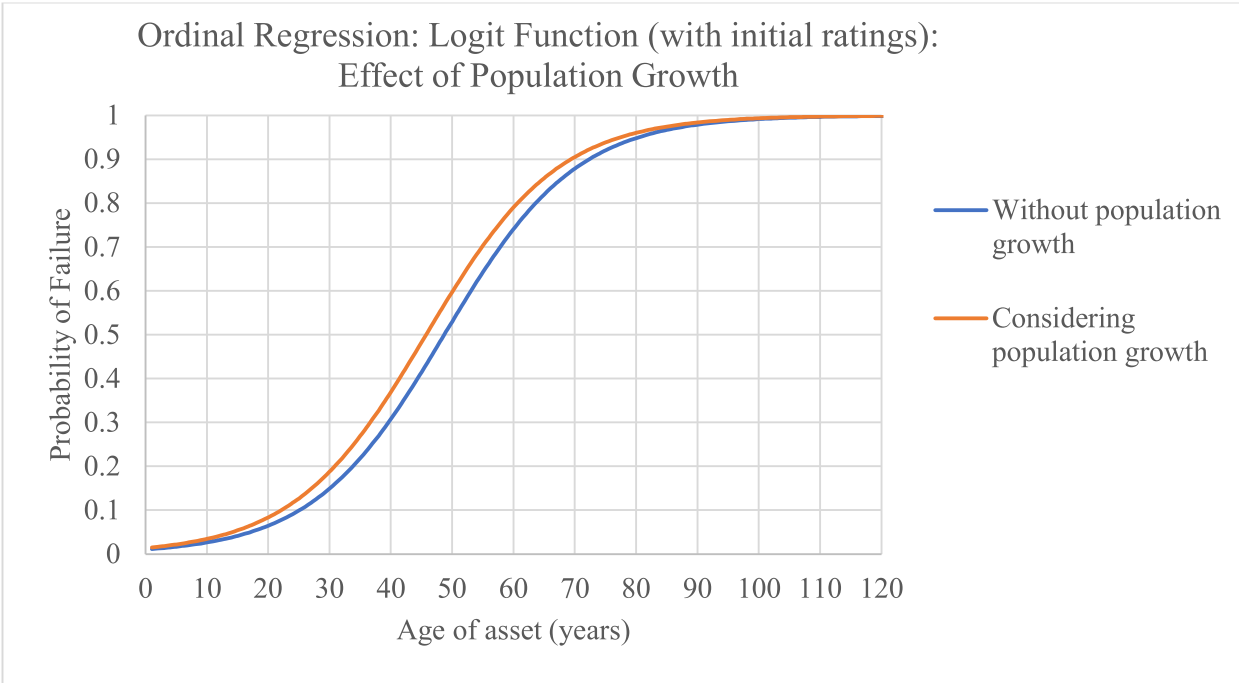

Figure 5.16: Effect of population growth on probability of failure with respect to age of asset (Ordinal Regression: Logit function and by considering initial ratings)

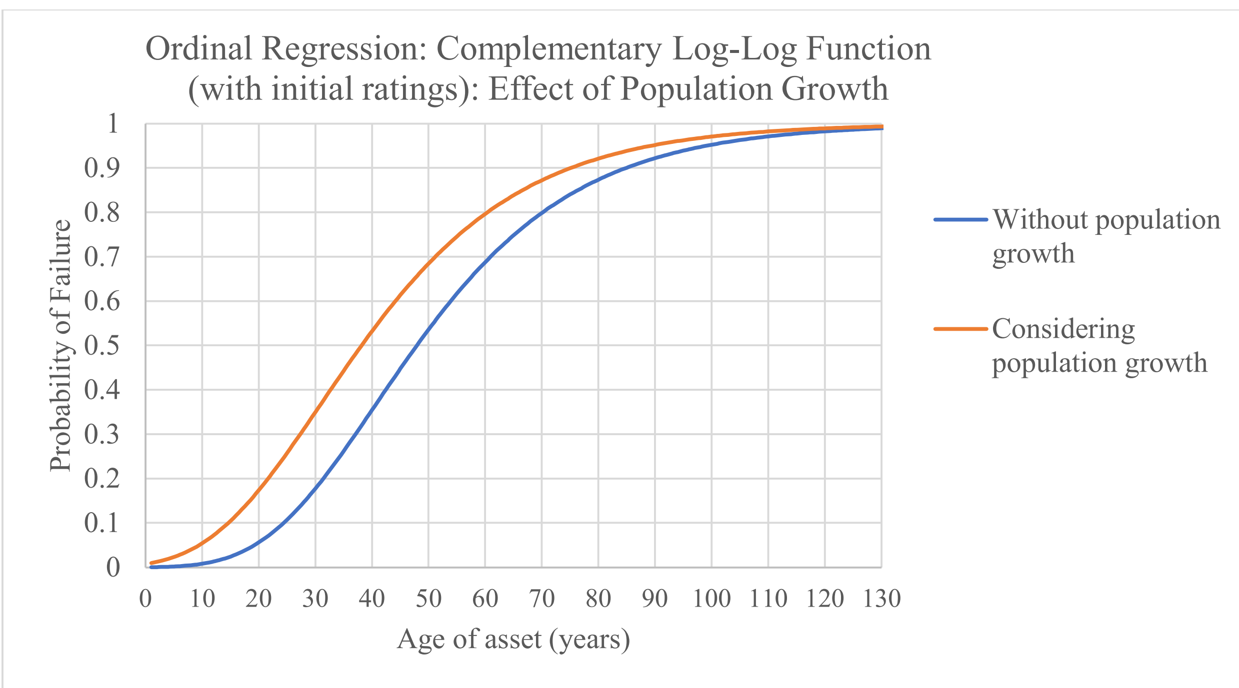

Figure 5.17: Influence of population growth on probability of failure (Ordinal Regression: Complementary Log-Log function and by considering initial ratings)

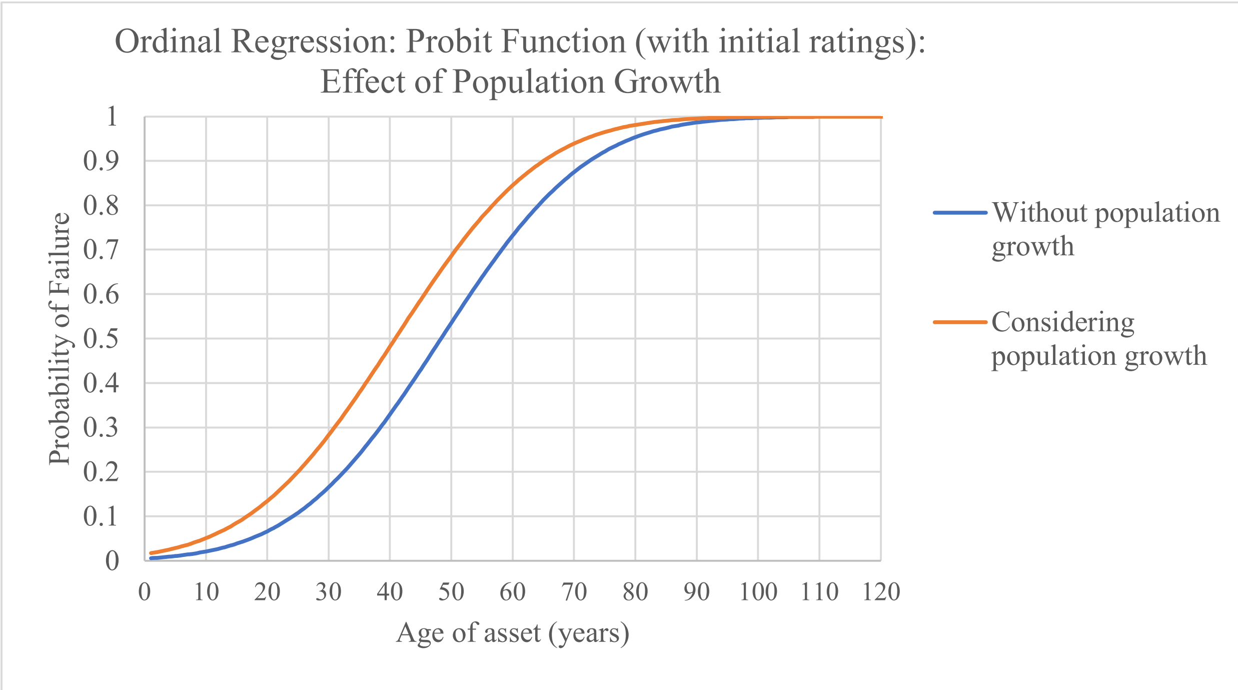

Figure 5.18: Effect of population growth on probability of failure with respect to age of asset (Ordinal Regression: Probit link function and by considering initial ratings)

5.8 Effect of independent variables on service life of gravity sewer pipes

In this section, based on the deterioration models developed using statistical approach and without considering the initial condition grading of sewer pipes (since these condition gradings were not available for assets), the influence of various independent variables on service life of sewer pipes as well as the probability of failure of assets are investigated. These independent variables are as follows: Age of pipe, Diameter of pipe, Pipe material, Average velocity of sewer flow, Average flow depth, Pipe slope, Length of pipe, Average flow in pipe (percent full). Moreover, it should be noted that proportional odds assumption was not validated through Cauchit link function and the results associated with this method is solely for illustration and comparison purposes.

5.8.1 Influence of pipe material on service life of gravity sewer pipes

Based on the results obtained through statistical approach, it was observed that for the sewer pipes considered in this study, when utilizing binomial logistic regression, multinomial logistic regression, ordinal regressions using various link functions (i.e. Logit, Probit, complementary log-log, negative log-log, and Cauchit functions), for the same sewer pipe, and without altering values of any other independent variables of the sewer pipe, the service lives associated with different pipe materials from highest to lowest value correspond to the following pipe materials:

- FRP
- PVC
- RCP
- VCP

Therefore, it is observed that the highest and lowest values of service life for these sewer pipes, considering values of all other independent variables remained unchanged, correspond to FRP and Vitrified Clay pipes, respectively. This observation may be due to the fact that, as illustrated in the data acquisition section, compared to FRP pipes, VCP pipes had significantly higher condition grading (in operational, structural, and overall categories). However, when considering these data, the ages of the pipes should also be accounted for. The lower service life obtained for vitrified clay pipes maybe due to the fact that vitrified clay pipes are in the brittle pipe categories, and are therefore more likely to gain higher condition grading (particularly considering structural condition grading) and moreover VCP pipes are more susceptible to sudden failure as well.

Furthermore, with regards to influence of pipe material on probability of failure of sewer pipes, it is observed that at any given point in time, the probability of failure obtained for different pipe materials from highest to lowest probability of failure correspond to the following pipe materials:

- VCP
- RCP
- PVC
- FRP

The same reasoning discussed for service life of sewer pipes with different pipe materials, can also be considered for probabilities of failure of sewer pipes as well.

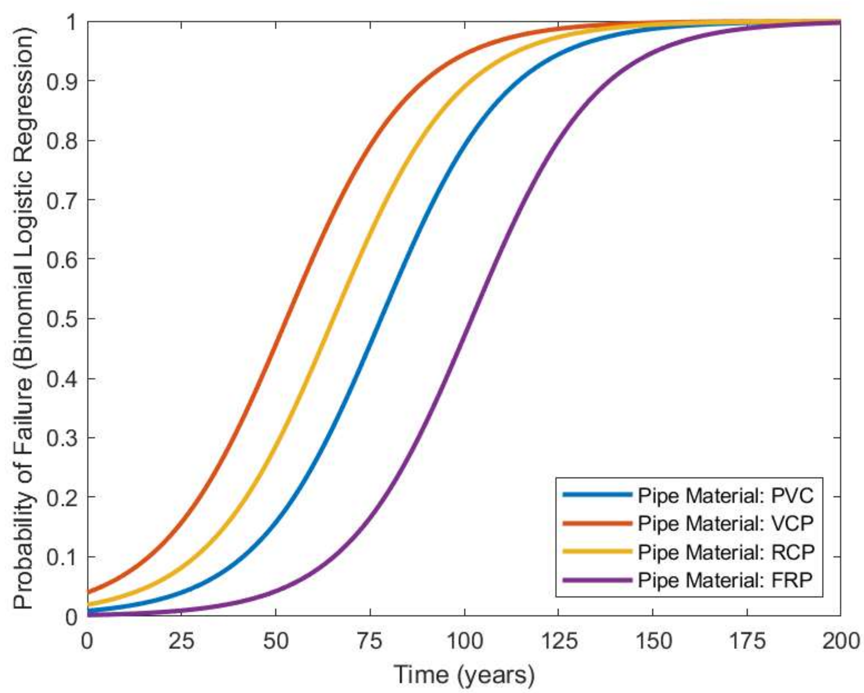

Figure 5.19: Probability of failure with respect to age for different pipe materials (Binomial Logistic Regression)

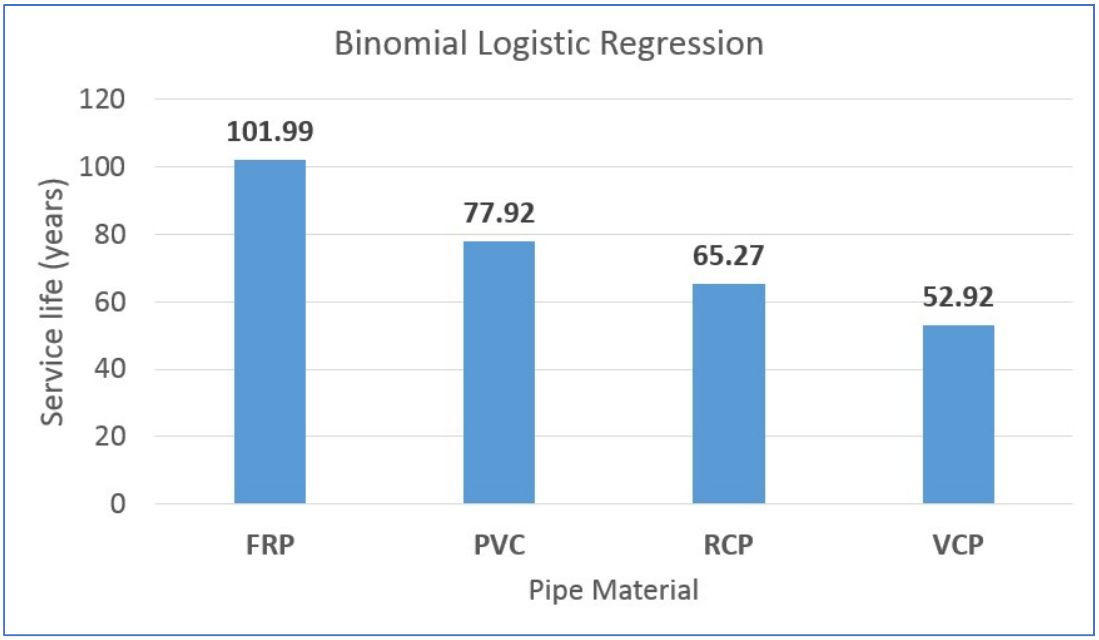

Figure 5.20: Service life of assets for different pipe materials (Binomial Logistic Regression)

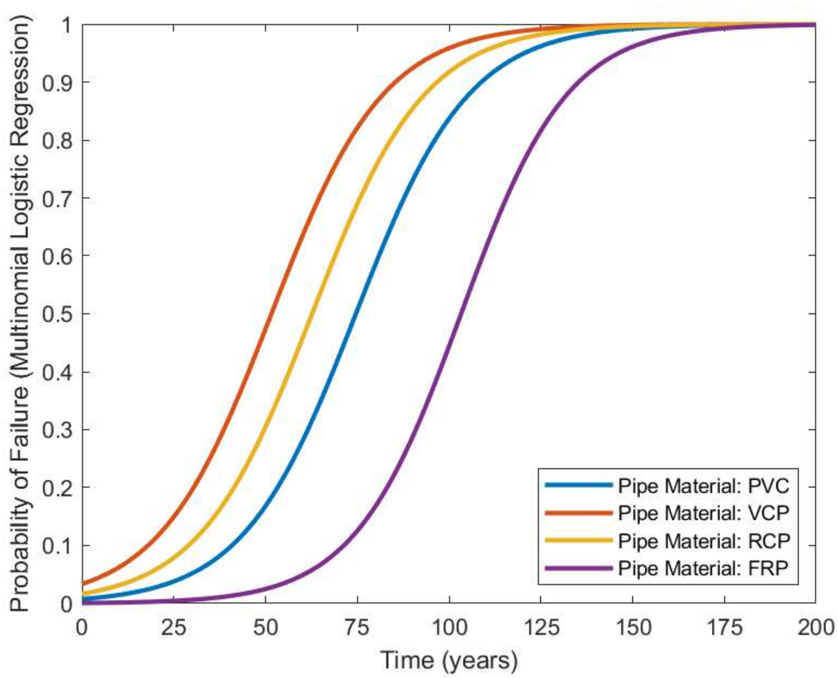

Figure 5.21: Probability of failure with respect to age for different pipe materials (Multinomial Logistic Regression)

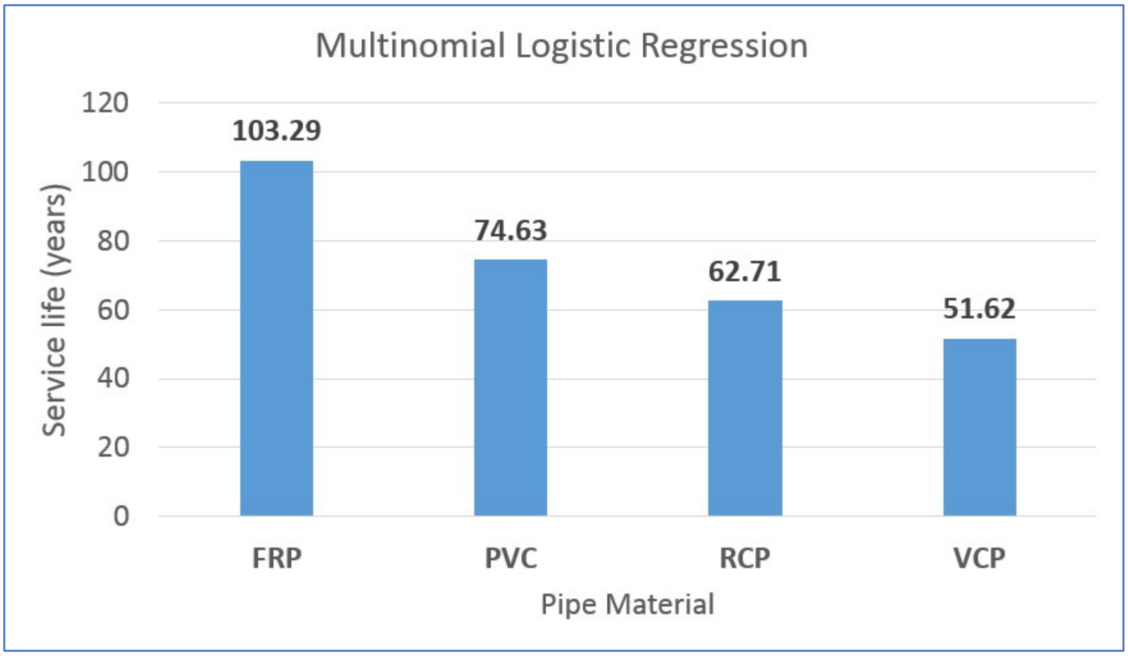

Figure 5.22: Service life of assets for different pipe materials (Multinomial Logistic Regression)

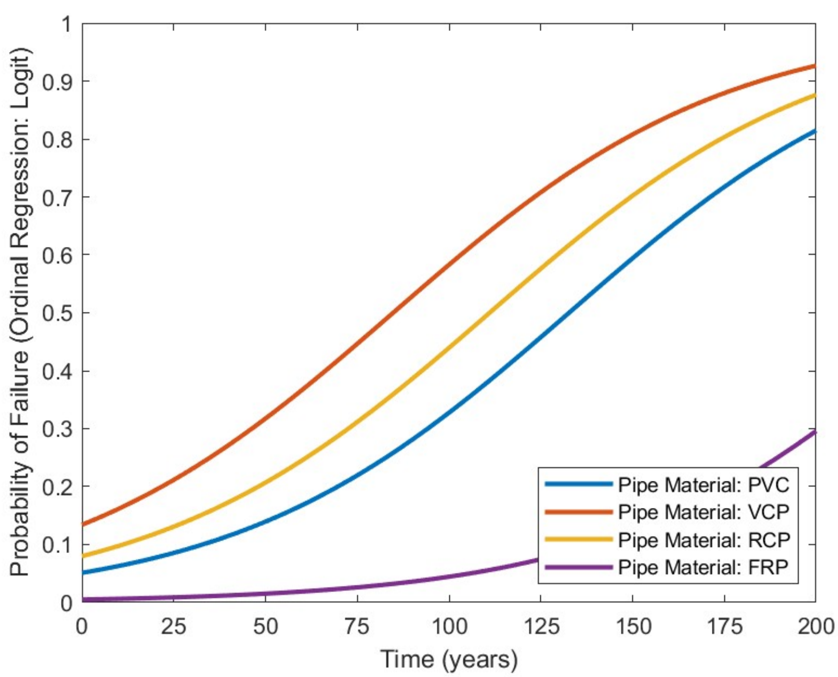

Figure 5.23: Probability of failure with respect to age for different pipe materials (Ordinal Regression: Logit Link Function)

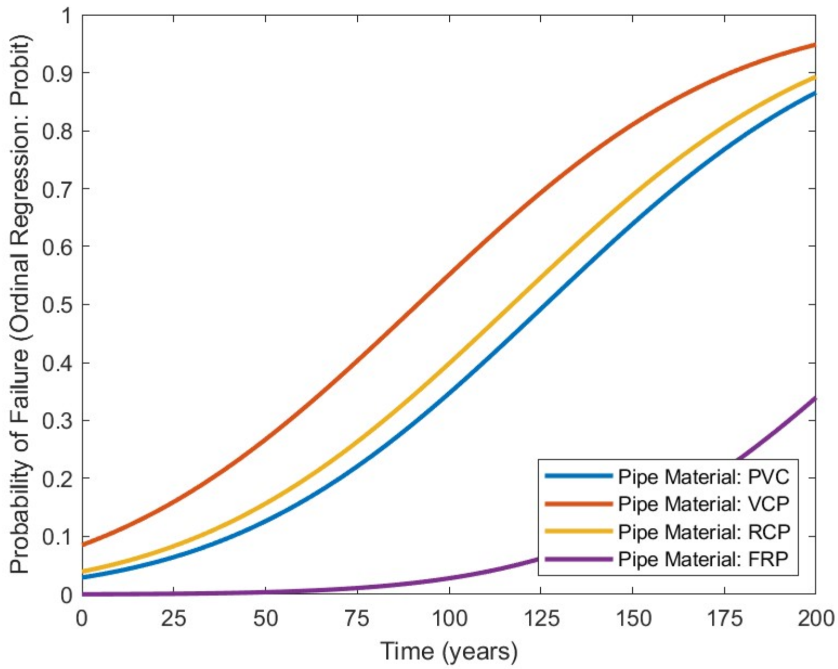

Figure 5.24: Probability of failure with respect to age for different pipe materials (Ordinal Regression: Probit Link Function)

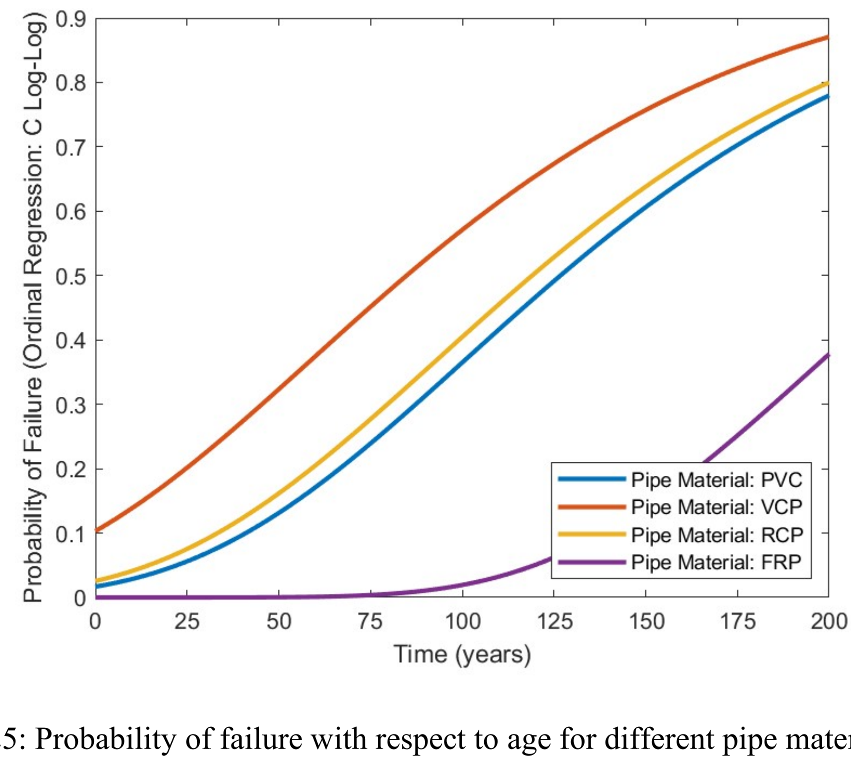

Figure 5.25: Probability of failure with respect to age for different pipe materials (Ordinal Regression: Complementary Log-Log Link Function)

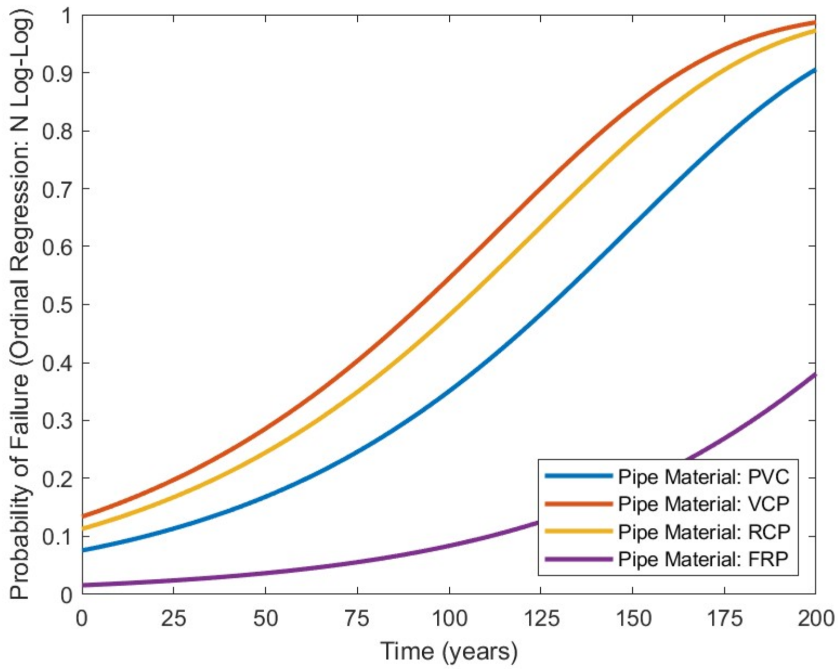

Figure 5.26: Probability of failure with respect to age for different pipe materials (Ordinal Regression: Negative Log-Log Link Function)

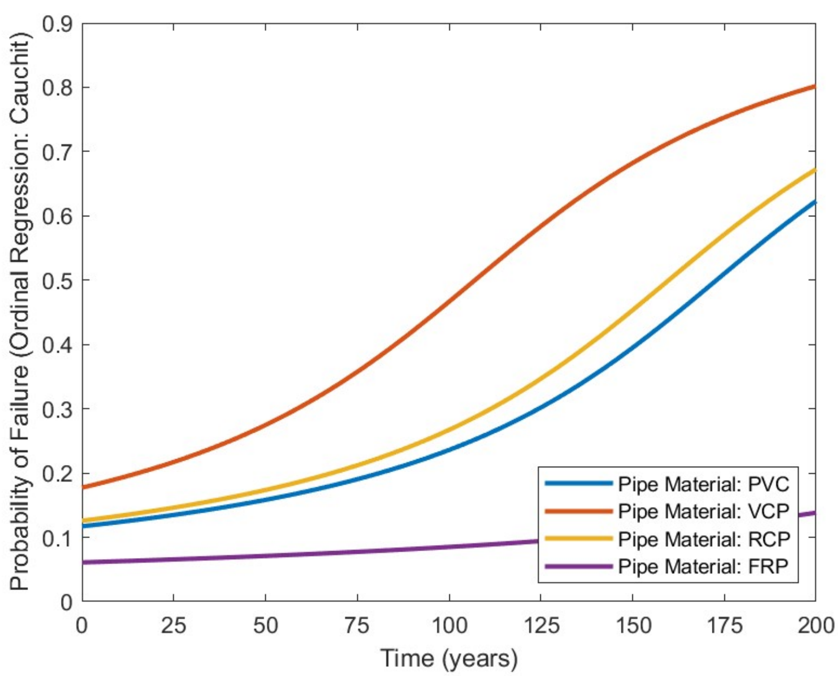

Figure 5.27: Probability of failure with respect to age for different pipe materials (Ordinal Regression: Cauchit Link Function)

5.8.2 Influence of pipe diameters on service life of gravity sewer pipes

For the sewer pipes considered in this study, the results obtained through statistical models illustrate that when modeling based upon binomial logistic regression, multinomial logistic regression, ordinal regressions using various link functions (i.e. Logit, Probit, complementary log-log, negative log-log, and Cauchit functions), for the same gravity sewer pipe, and while no changes are made in the values of other independent variables of the gravity sewer pipe, when the size of the pipe diameter increases, the service lives of sewer pipes are subjected to reduction. It is further observed that increase in the sewer pipe diameter results in increase in the probability of failure of sewer pipe too.

This observation may stem from variations in the installation of sewer pipes, bedding and backfill conditions, as well as depth of the sewer pipes (which can differ for different pipe diameters). The data associated with these potential factors were not available for processing.

With regards to significance of diameter of pipe, in the following methods, this parameter was found to be significant:

Binomial logistic regression

Multinomial logistic regression

Ordinal regression using Probit link function

Ordinal regression using Complementary Log-Log link function

It should be noted that when using all independent variables for Ordinal regression with Negative Log-Log and Logit link function, pipe diameter was determined as a significant variable as well, however, in their associated final models which some predictors were eliminated, based on these link functions, as well as Cauchit link function, pipe diameter was not significant.

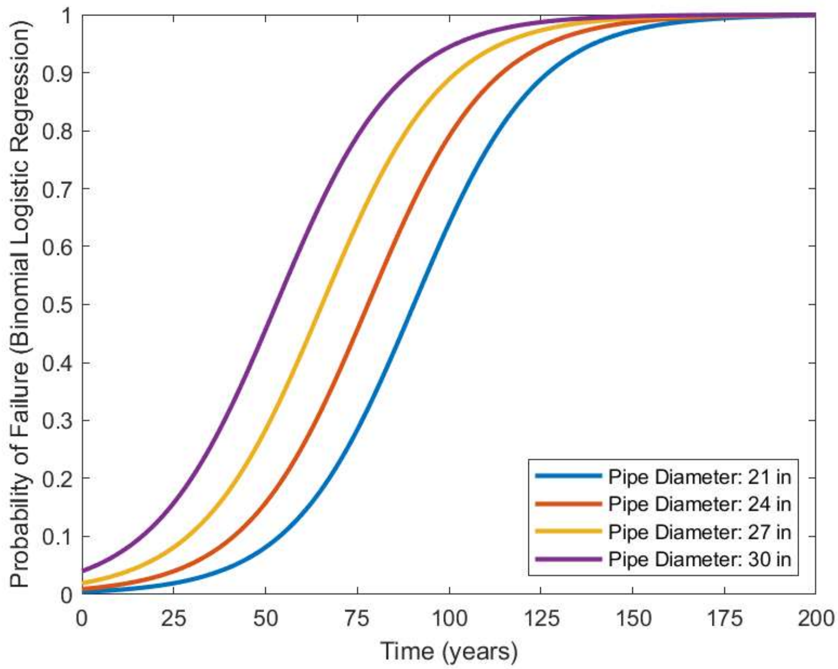

Figure 5.28: Probability of failure with respect to age for various pipe diameters (Binomial Logistic Regression)

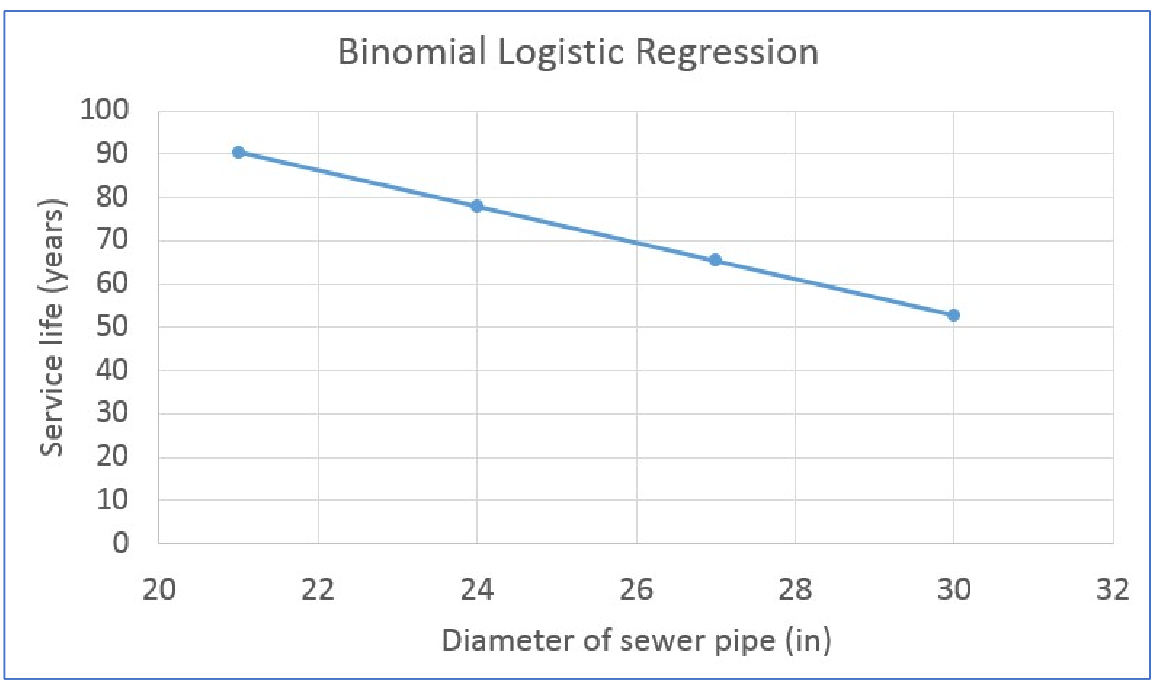

Figure 5.29: Service life of assets for various pipe diameters (Binomial Logistic Regression)

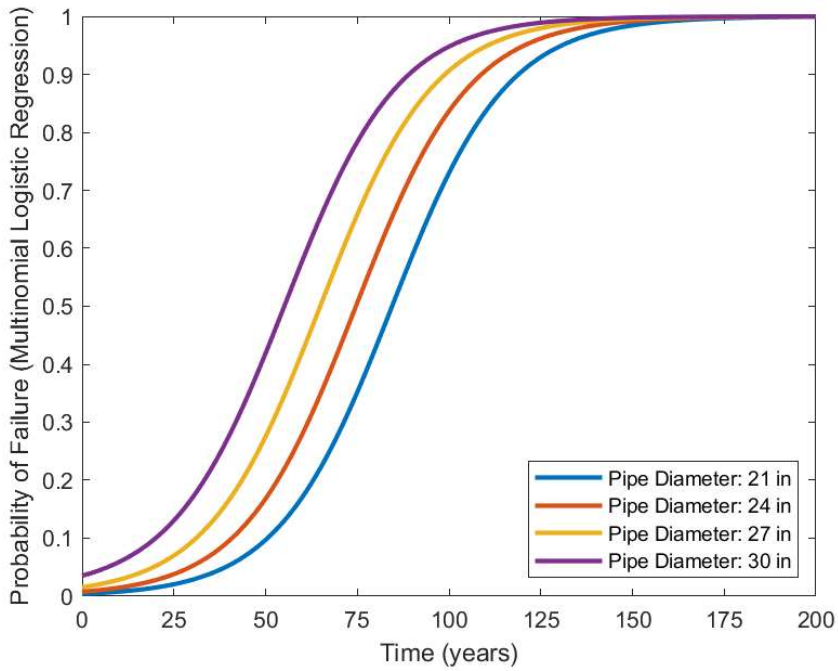

Figure 5.30: Probability of failure with respect to age for various pipe diameters (Multinomial Logistic Regression)

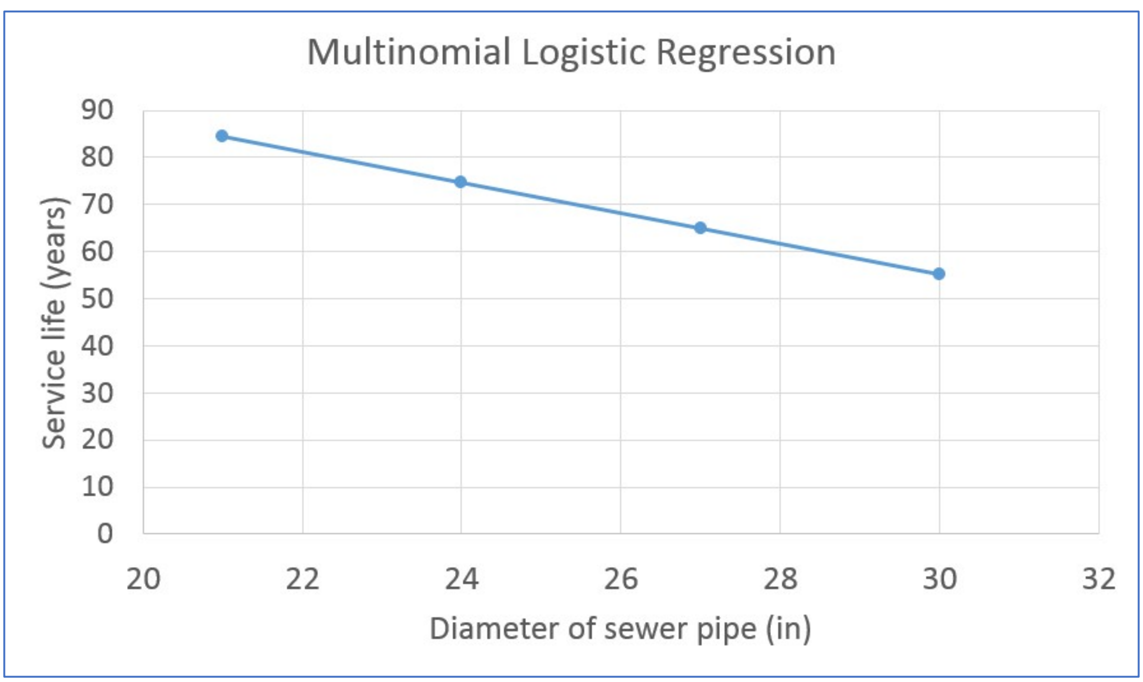

Figure 5.31: Service life of assets for various pipe diameters (Multinomial Logistic Regression)

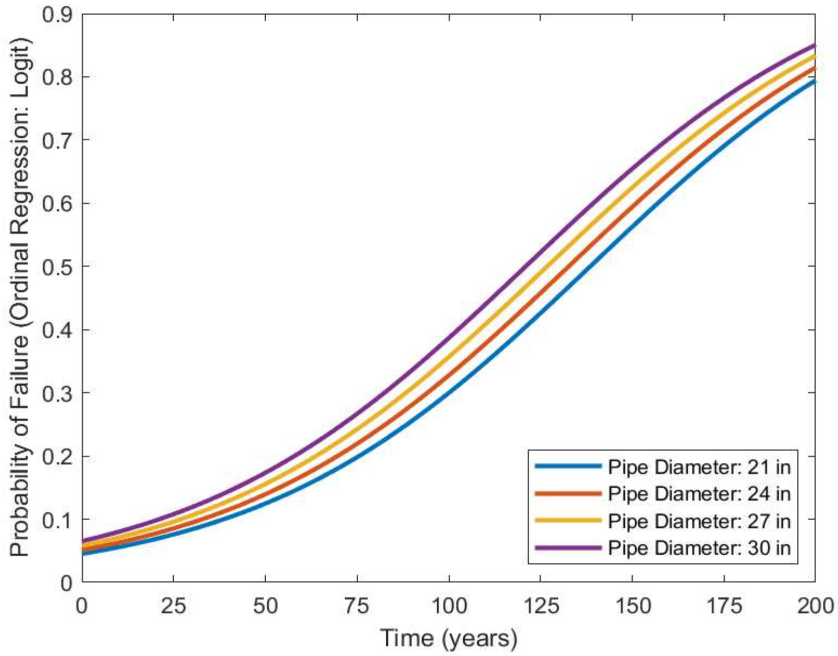

Figure 5.32: Probability of failure with respect to age for various pipe diameters (Ordinal Regression: Logit Link Function)

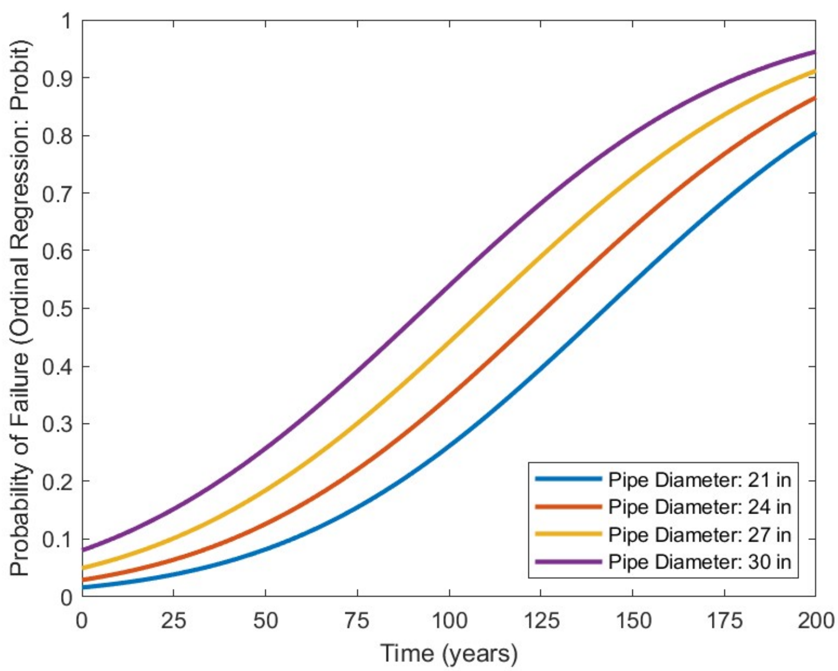

Figure 5.33: Probability of failure with respect to age for various pipe diameters (Ordinal Regression: Probit Link Function)

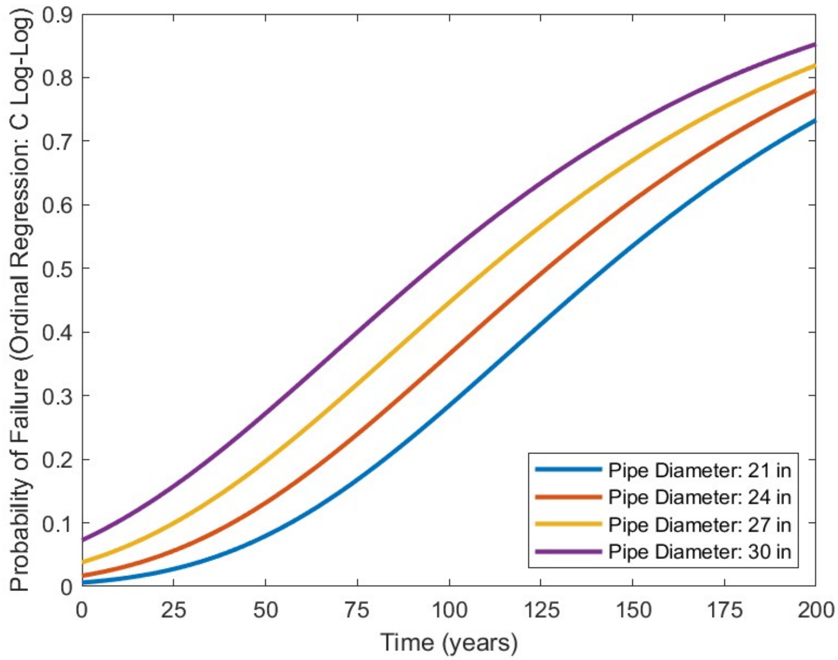

Figure 5.34: Probability of failure with respect to age for various pipe diameters (Ordinal Regression: Complementary Log-Log Link Function)

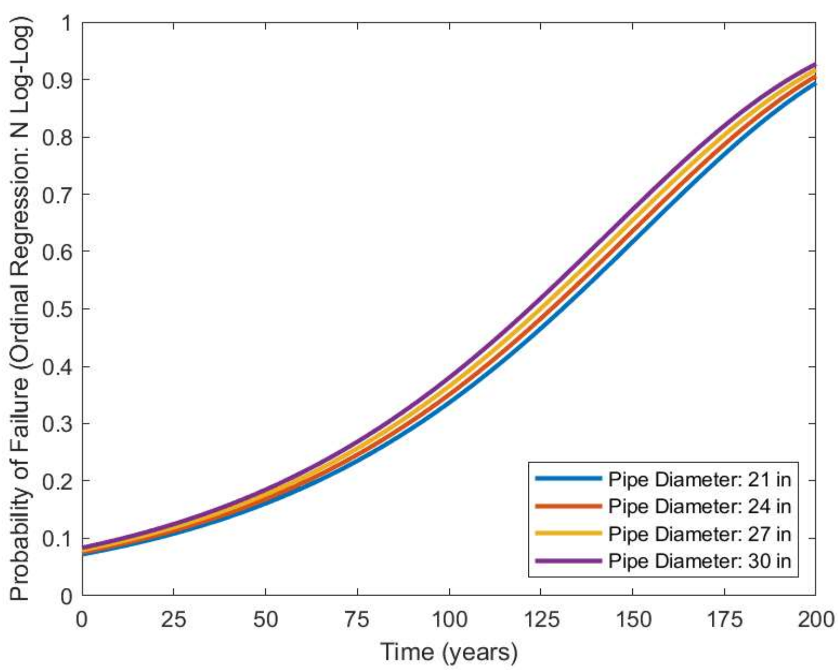

Figure 5.35: Probability of failure with respect to age for various pipe diameters (Ordinal Regression: Negative Log-Log Link Function)

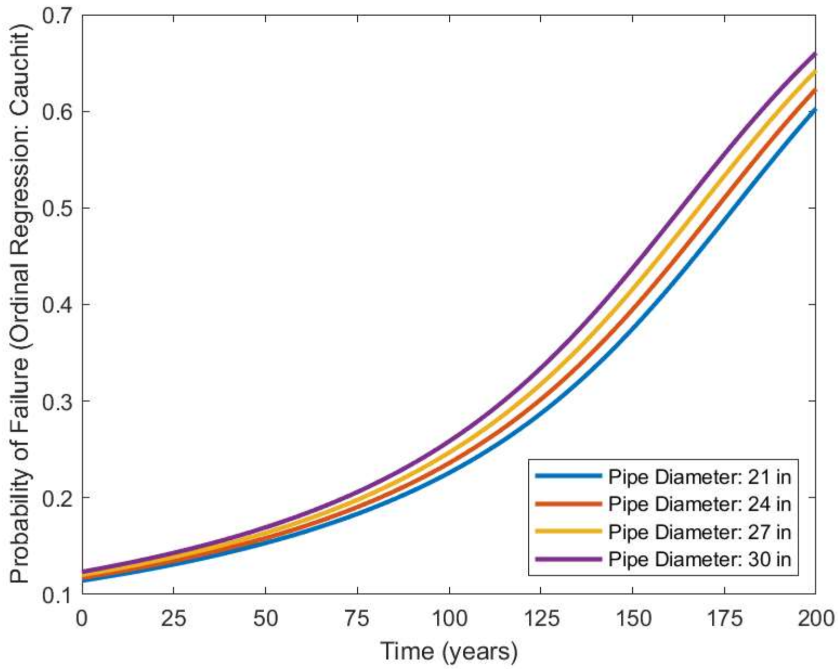

Figure 5.36: Probability of failure with respect to age for various pipe diameters (Ordinal Regression: Cauchit Link Function)

5.8.3 Influence of length of sewer pipes on service life of gravity sewer pipes

For the sewer pipes considered in this study, it was observed that through implementation of the following methods, both the service life and probability of failure of sewer pipes were not impacted by the length of the sewer pipe:

Binomial logistic regression

Ordinal regression with Logit link function

Ordinal regression with Probit link function

Ordinal regression with Complementary Log-Log link function

Ordinal regression with Cauchit link function

However, when utilizing multinomial logistic regression and ordinal regression with Negative Log-Log link function, it was observed that increasing the length of the sewer pipe slightly impacted the service life by decreasing its value and increasing the probability of failure of pipe with respect to age of pipe. Therefore, due to observations made from the majority of the methods, it can be concluded that based on the available data set and regression methods used herein, the length of the sewer pipes did not seem to have an impact on the service life of the sewer pipes considered in this study.

Length of sewer pipes was not found to be a significant independent variable in any of the statistical models developed herein. Hence, this verifies the observations made that this predictor does not seem to alter the service life of assets.

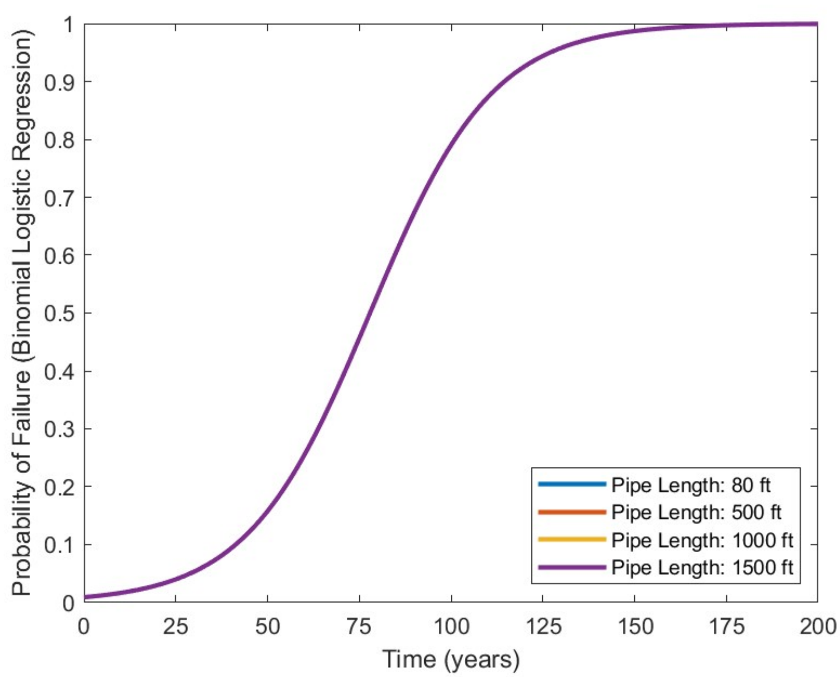

Figure 5.37: Probability of failure with respect to age for different pipe lengths (Binomial Logistic Regression)

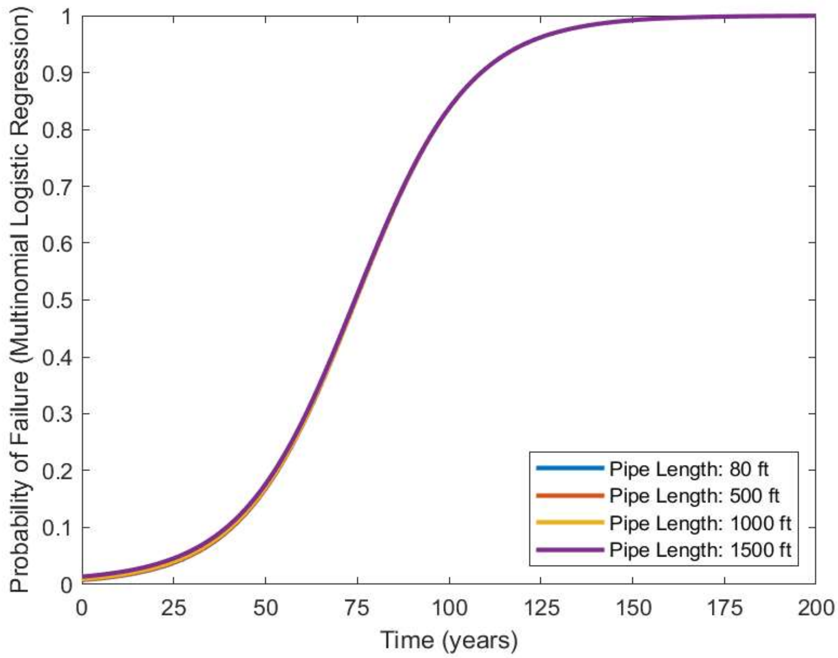

Figure 5.38: Probability of failure with respect to age for different pipe lengths (Multinomial Logistic Regression)

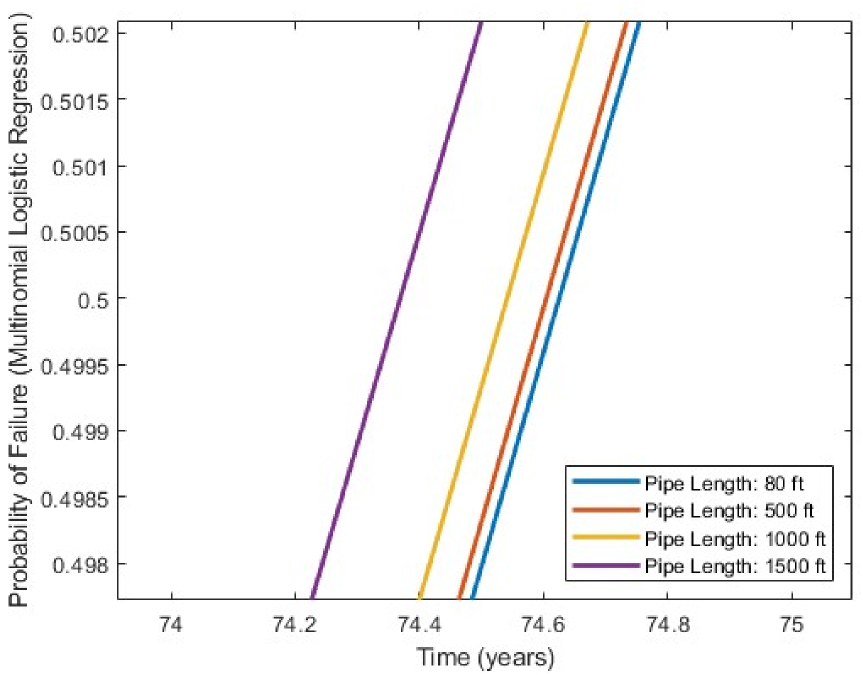

Figure 5.39: Comparing probabilities of failure for different pipe lengths (Multinomial Logistic Regression)

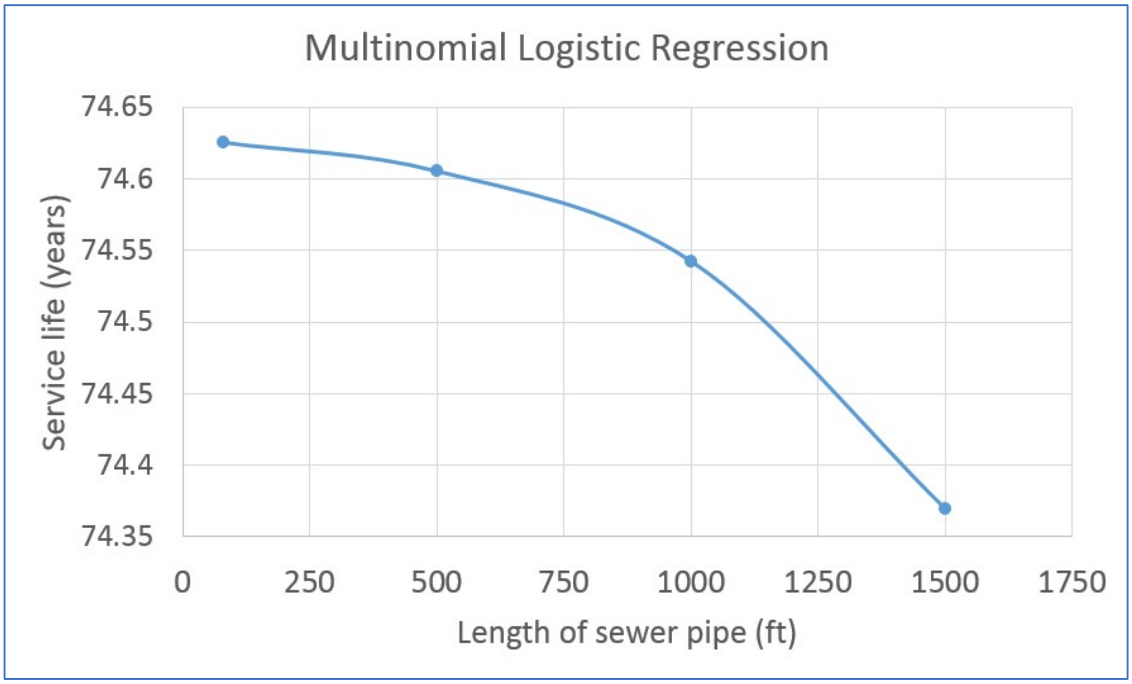

Figure 5.40: Service life of assets for different pipe lengths (Multinomial Logistic Regression)

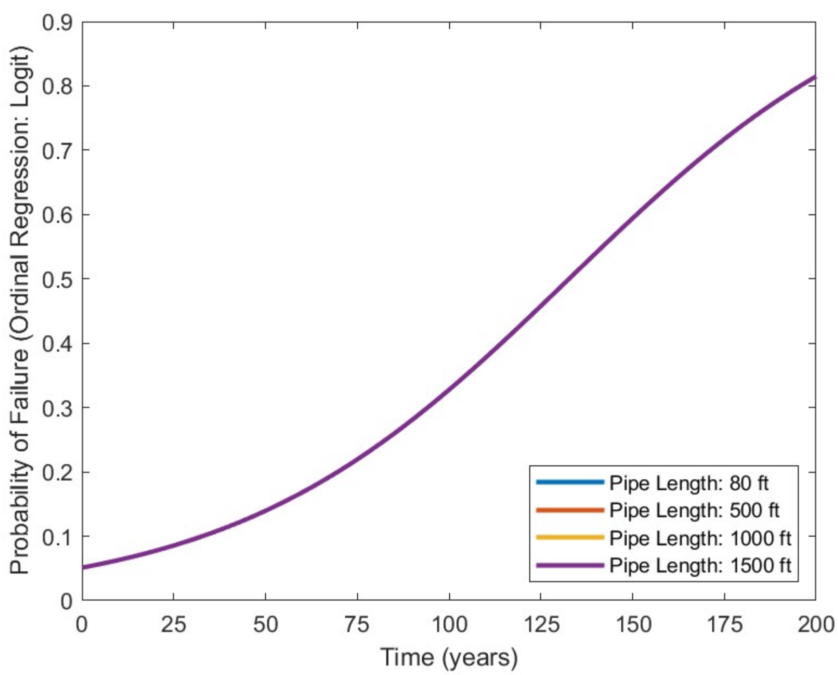

Figure 5.41: Probability of failure with respect to age for different pipe lengths (Ordinal Regression: Logit Link Function)

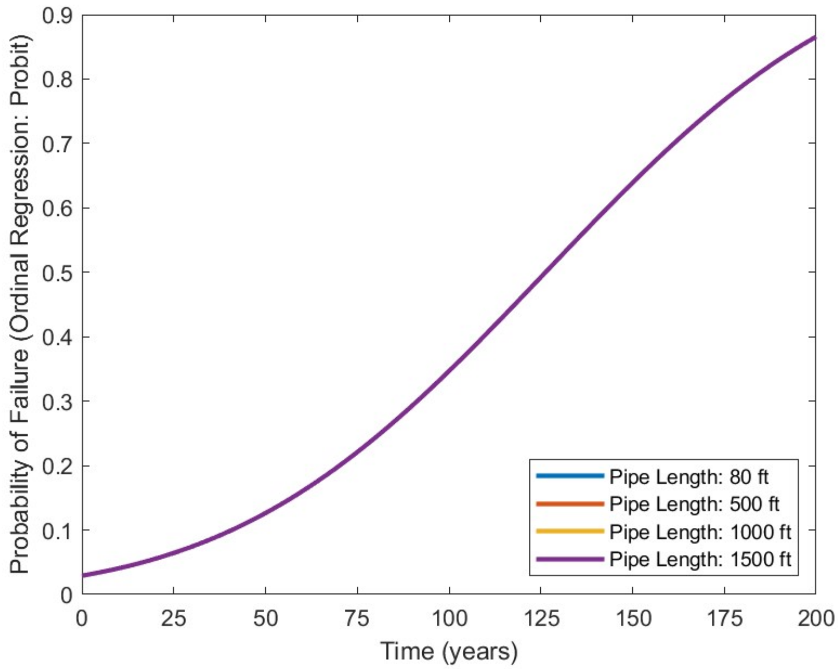

Figure 5.42: Probability of failure with respect to age for different pipe lengths (Ordinal Regression: Probit Link Function)

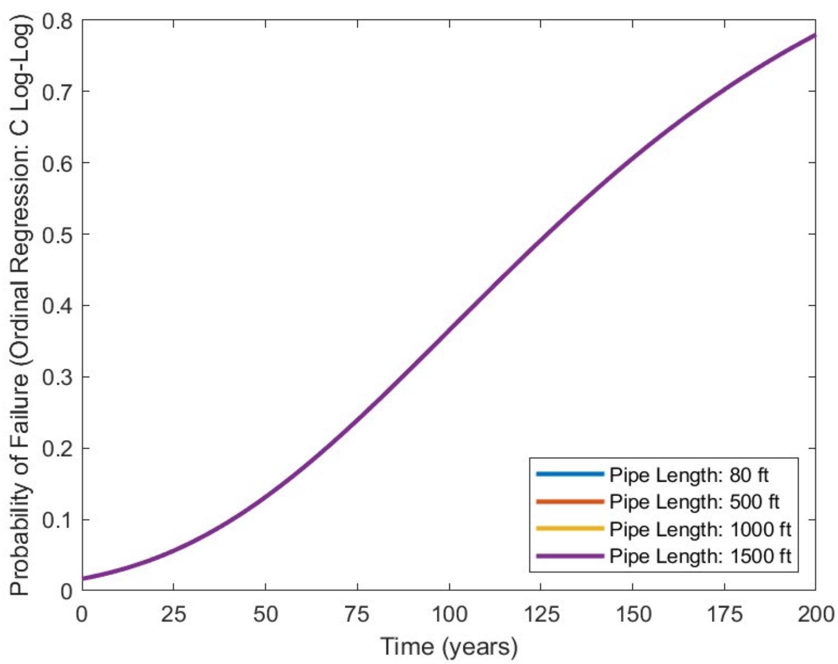

Figure 5.43: Probability of failure with respect to age for different pipe lengths (Ordinal Regression: Complementary Log-Log Link Function)

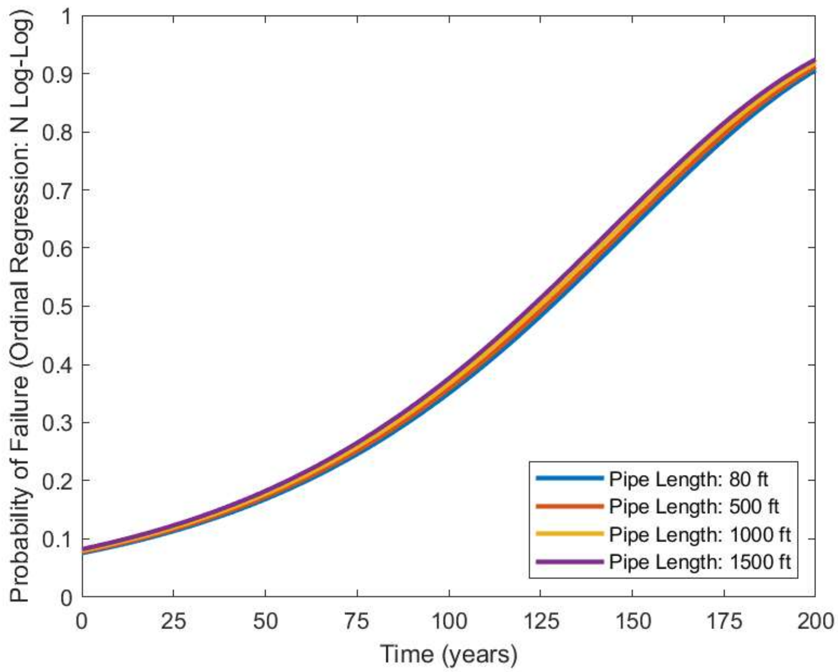

Figure 5.44: Probability of failure with respect to age for different pipe lengths (Ordinal Regression: Negative Log-Log Link Function)

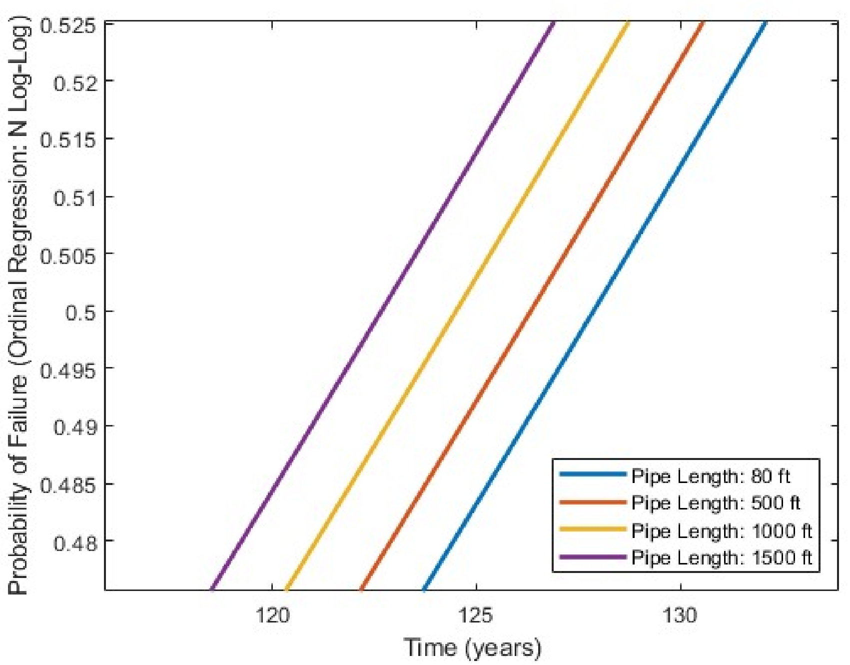

Figure 5.45: Comparing probabilities of failure for different pipe lengths (Ordinal Regression: Negative Log-Log Link Function)

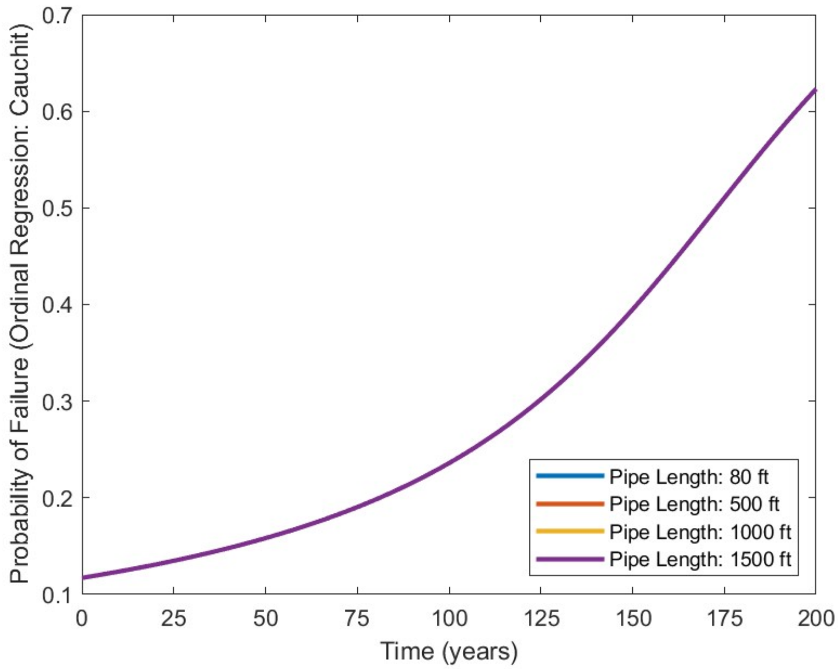

Figure 5.46: Probability of failure with respect to age for different pipe lengths (Ordinal Regression: Cauchit Link Function)

5.8.4 Influence of slope of sewer pipes on service life of gravity sewer pipes

Based on the developed models, it was observed that when utilizing the following modeling techniques, the values of service life of sewer pipes and the probability of failure of sewer pipes with respect to age were slightly decreased and increased, respectively due to increase in pipe slope:

Binomial logistic regression

Multinomial logistic regression

Ordinal regression with Probit link function

Ordinal regression with Complementary Log-Log link function

On the other hand, by using the following methods, by increasing the slope of sewer pipes, the values of service life of sewer pipes and the probability of failure of sewer pipes with respect to age were slightly increased and decreased, respectively:

Ordinal regression with Logit link function

Ordinal regression with Cauchit link function

Ordinal regression with Negative Log-Log link function

Considering the Cauchit link function was solely utilized for comparison purposes, as it did not satisfy the test of parallel lines and therefore, proportional odds assumption was not satisfied, it is concluded that the majority of the modeling techniques demonstrate that due to increase in the pipe slope, the values of service life of sewer pipes and the probability of failure of sewer pipes with respect to age were slightly decreased and increased, respectively.

However, based on the results obtained from all of the statistical deterioration models, the slope of sewer pipe was not determined as a significant predictor of these models. Therefore, this verifies the slight and varying impact of this predictor on service life and probability of failure of pipes.

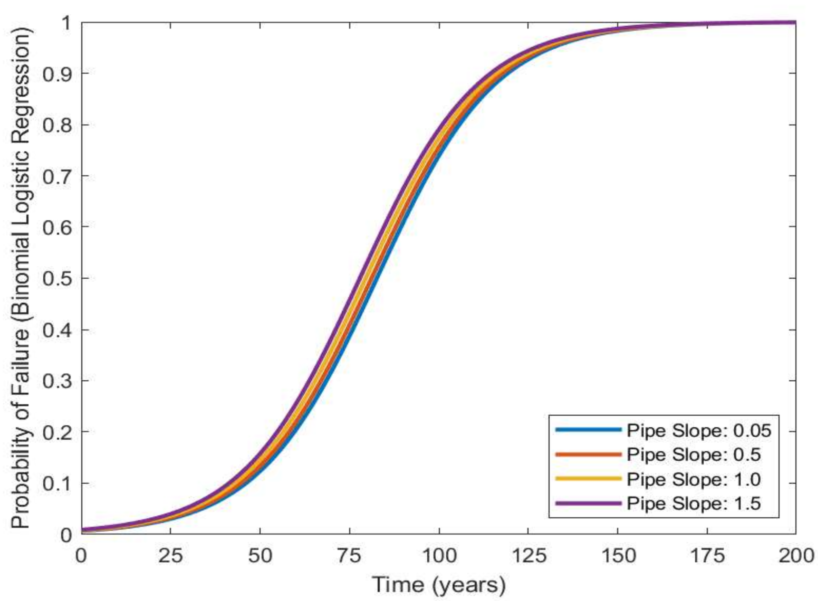

Figure 5.47: Probability of failure with respect to age for various pipe slopes (Binomial Logistic Regression)

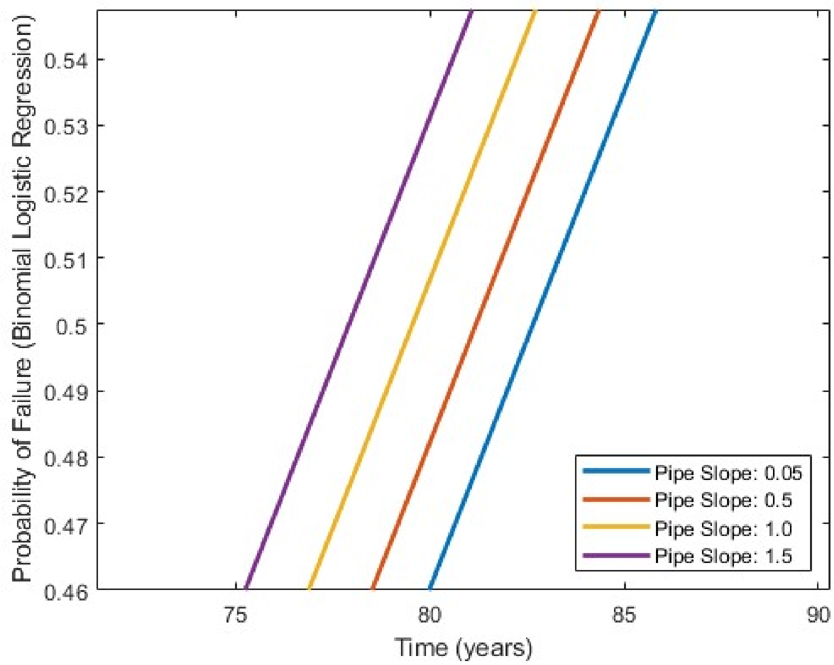

Figure 5.48: Comparing probabilities of failure for various pipe slopes (Binomial Logistic Regression)

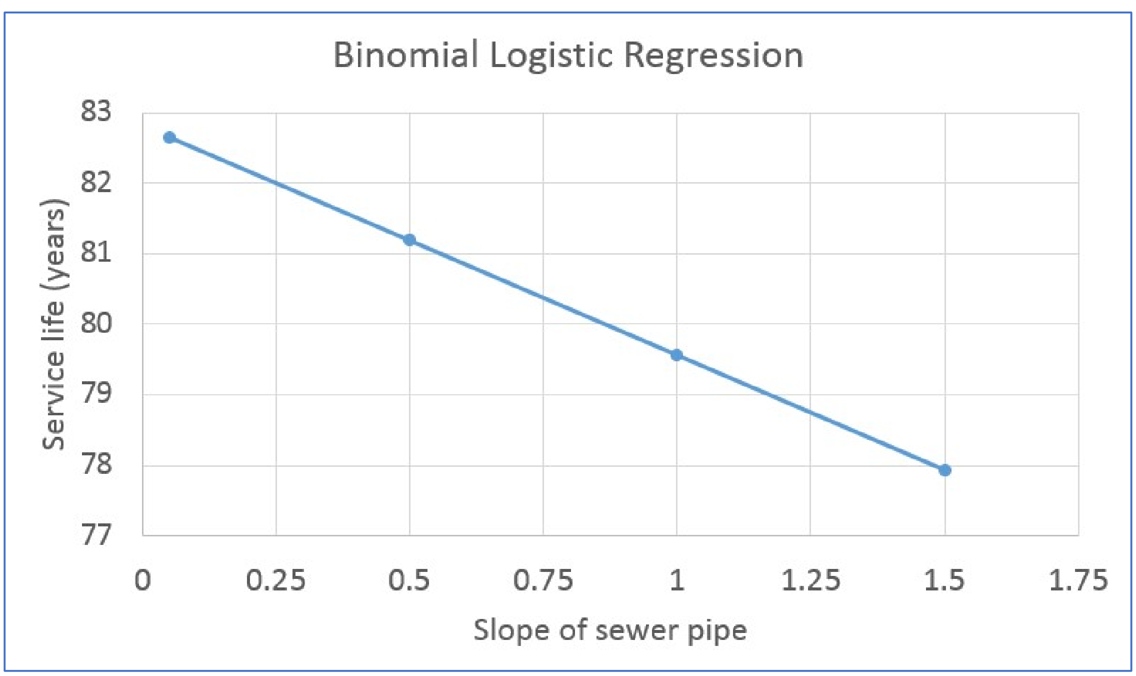

Figure 5.49: Service life of assets for various pipe slopes (Binomial Logistic Regression)

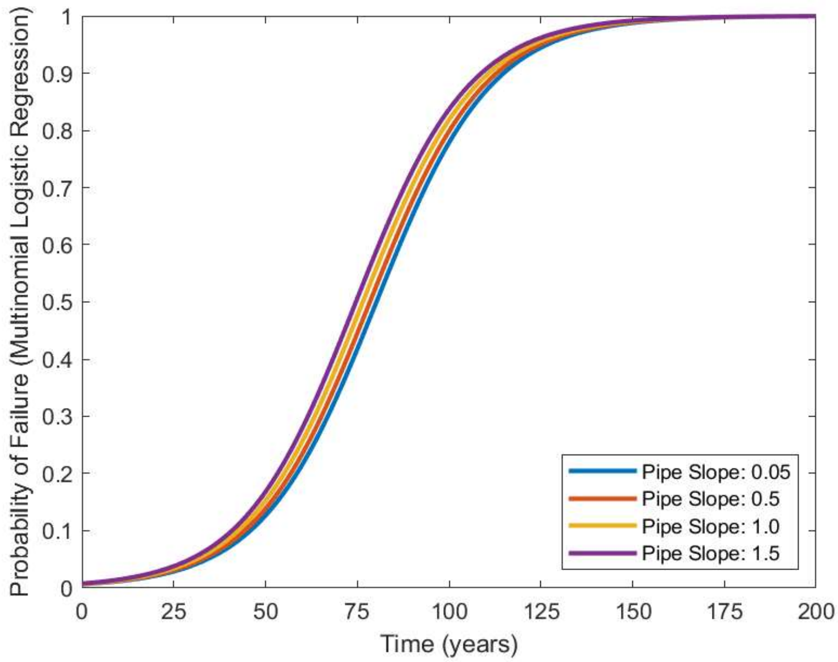

Figure 5.50: Probability of failure with respect to age for various pipe slopes (Multinomial Logistic Regression)

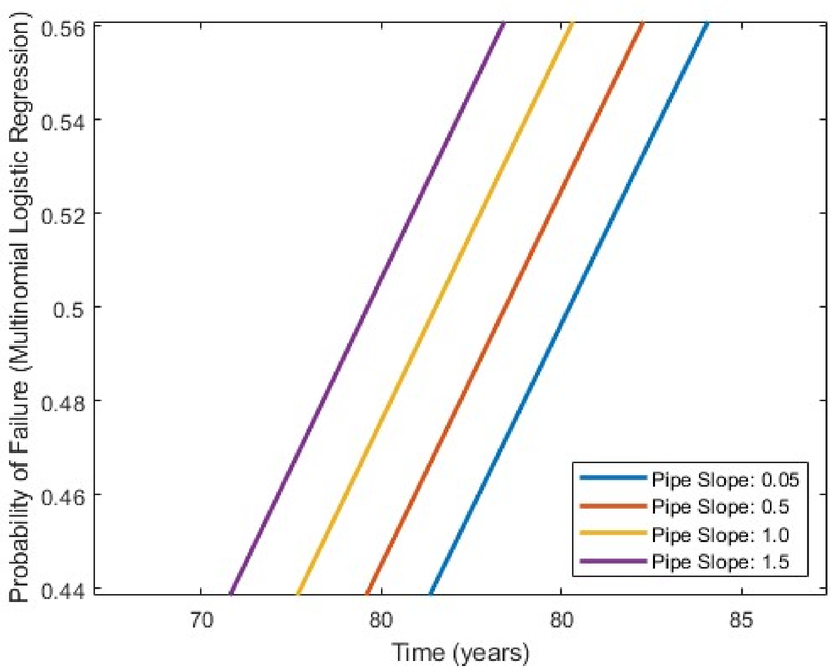

Figure 5.51: Comparing probabilities of failure for various pipe slopes (Multinomial Logistic Regression)

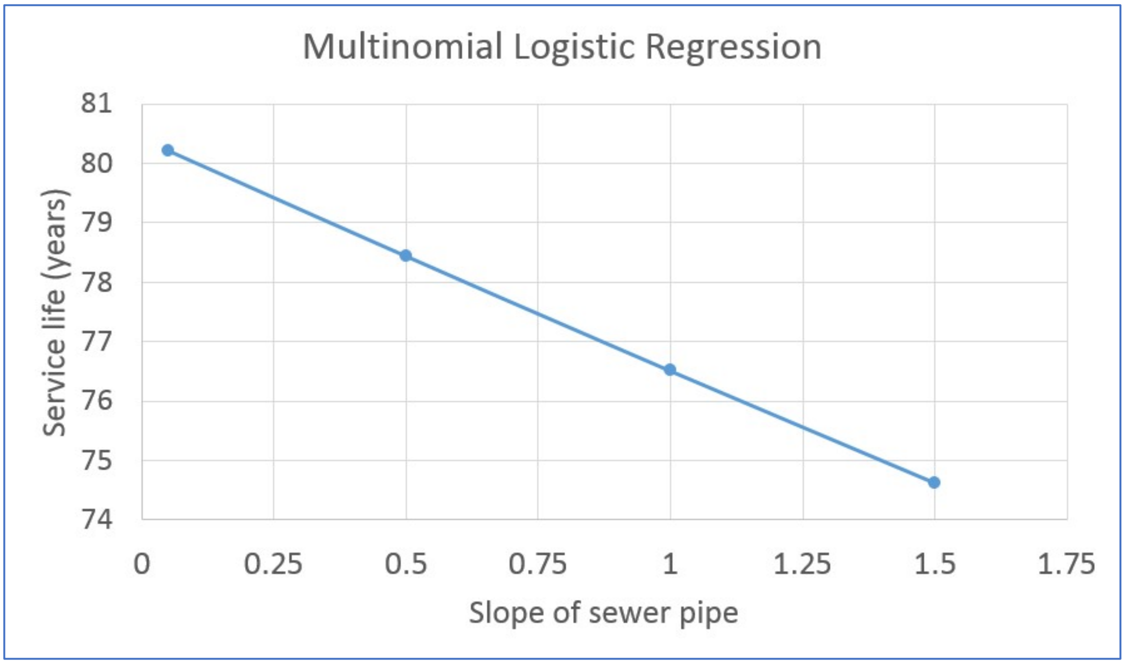

Figure 5.52: Service life of assets for various pipe slopes (Multinomial Logistic Regression)

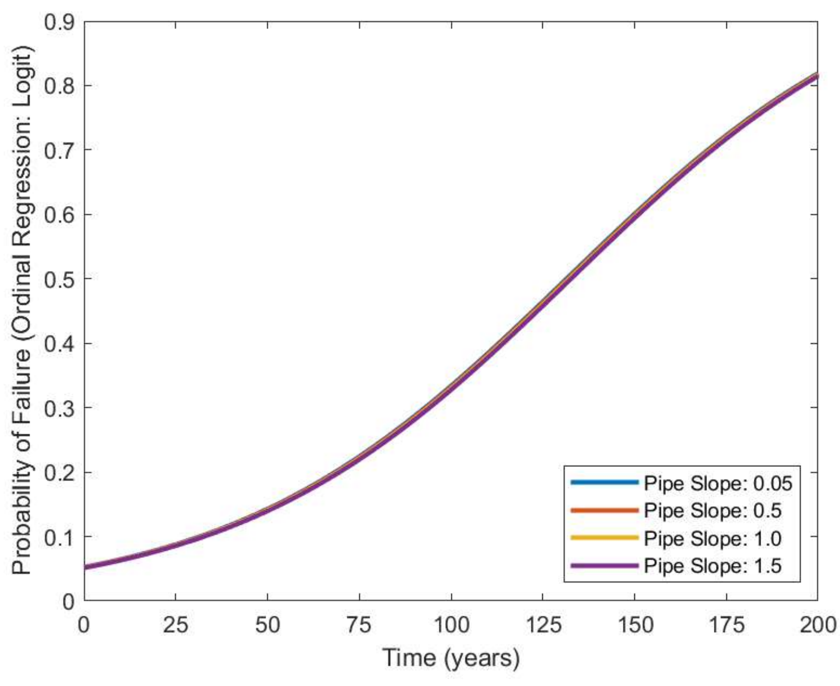

Figure 5.53: Probability of failure with respect to age for various pipe slopes (Ordinal Regression: Logit Link Function)

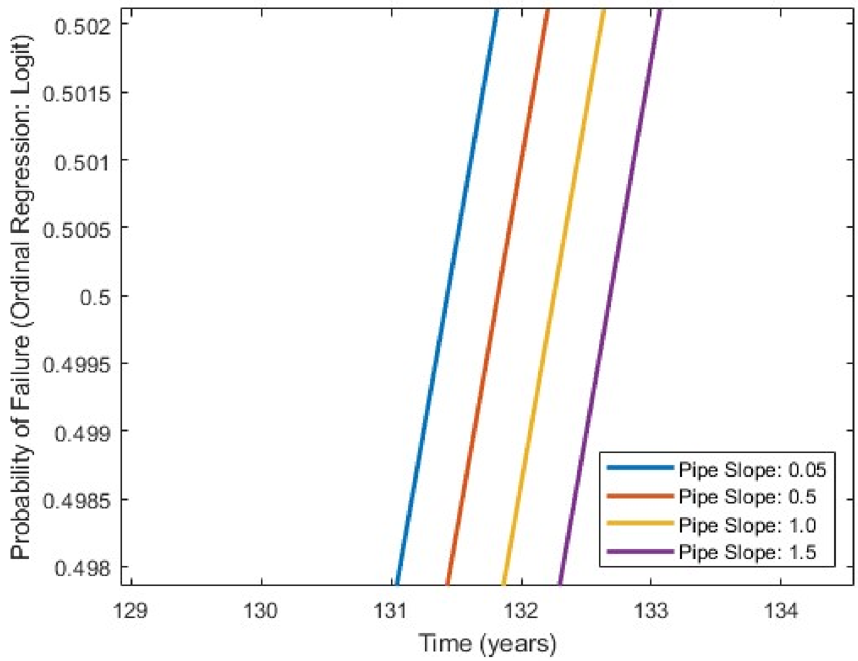

Figure 5.54: Comparing probabilities of failure for various pipe slopes (Ordinal Regression: Logit Link Function)

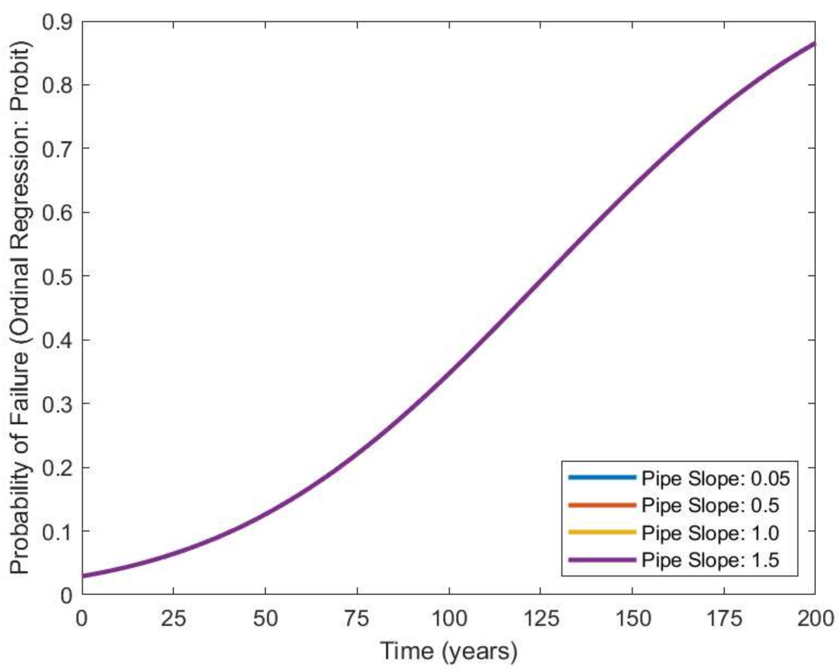

Figure 5.55: Probability of failure with respect to age for various pipe slopes (Ordinal Regression: Probit Link Function)

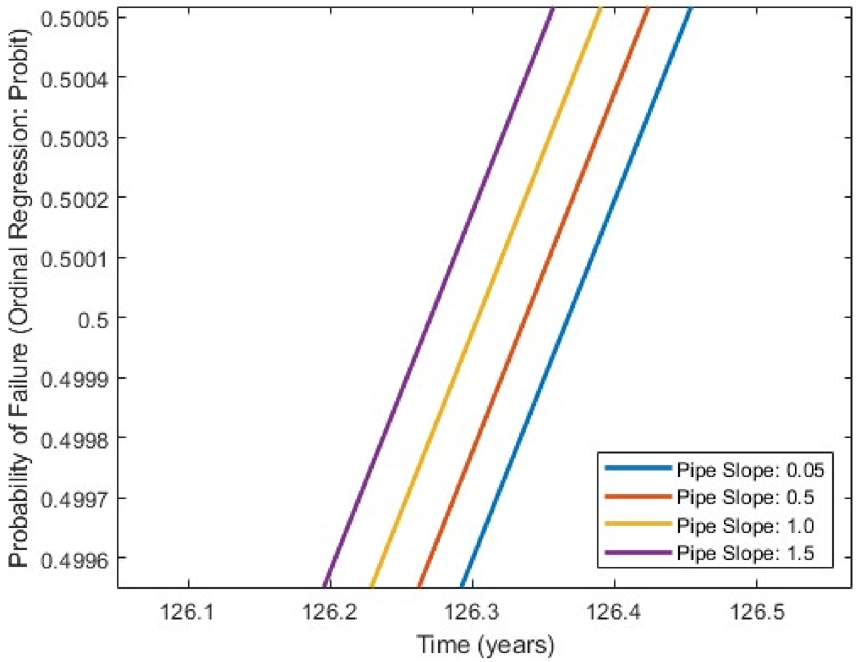

Figure 5.56: Comparing probabilities of failure for various pipe slopes (Ordinal Regression: Probit Link Function)

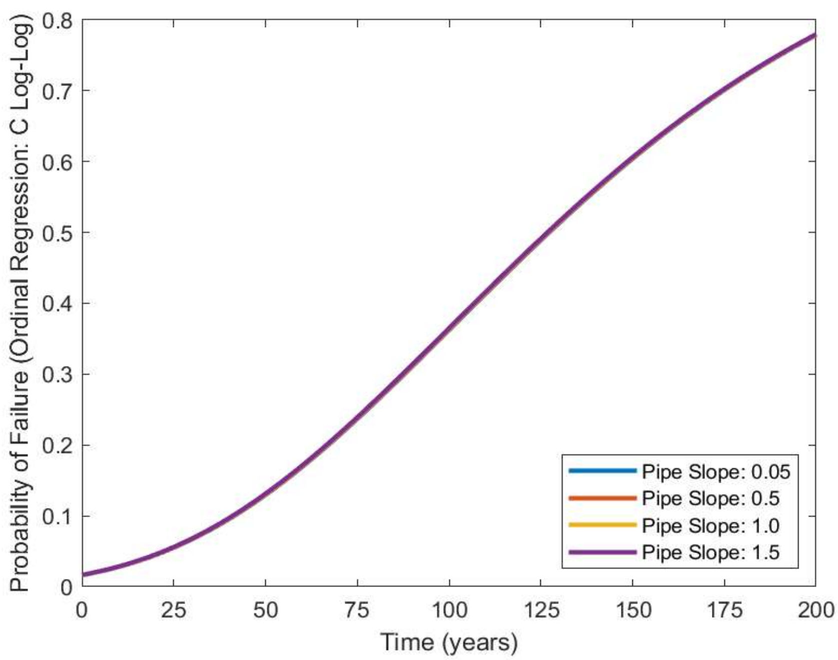

Figure 5.57: Probability of failure with respect to age for various pipe slopes (Ordinal Regression: Complementary Log-Log Link Function)

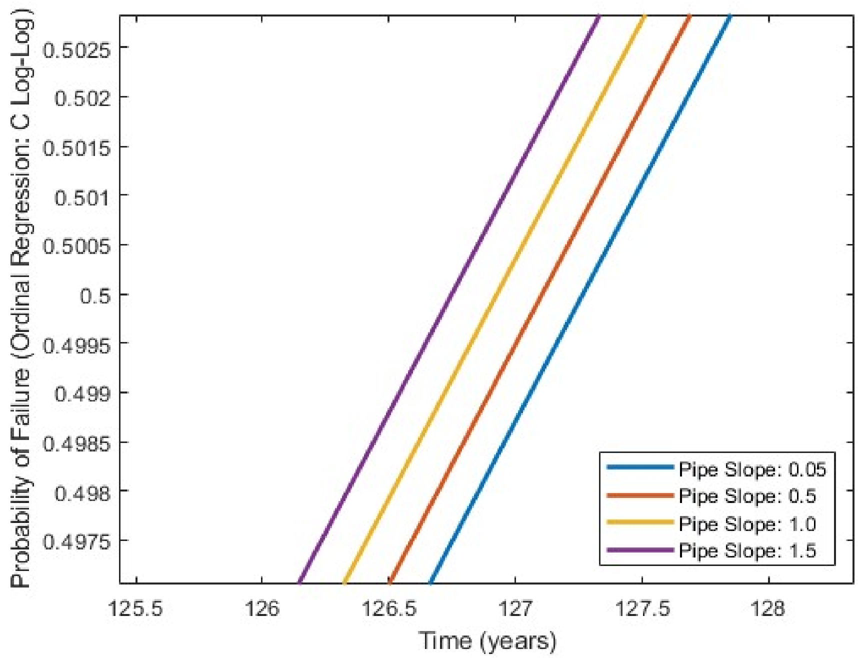

Figure 5.58: Comparing probabilities of failure for various pipe slopes (Ordinal Regression: Complementary Log-Log Link Function)

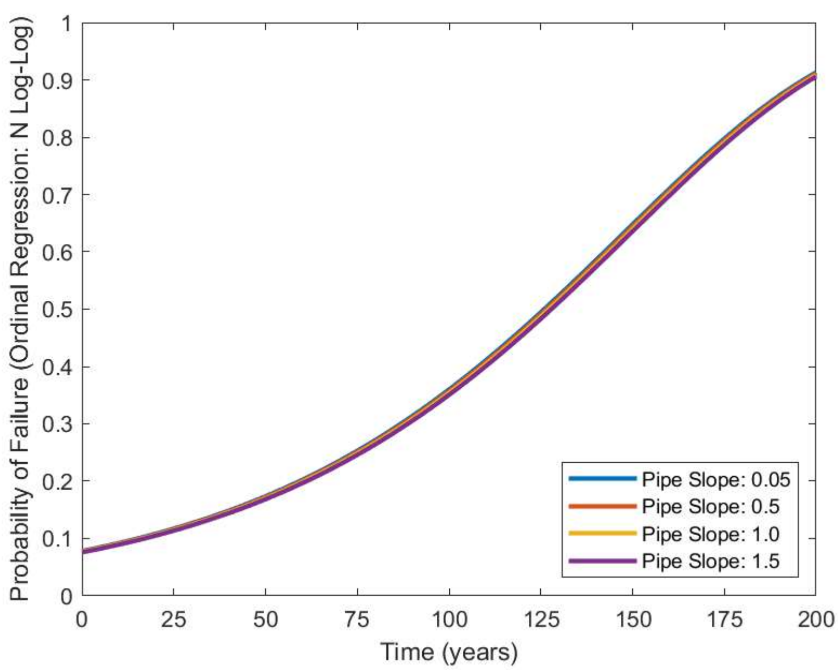

Figure 5.59: Probability of failure with respect to age for various pipe slopes (Ordinal Regression: Negative Log-Log Link Function)

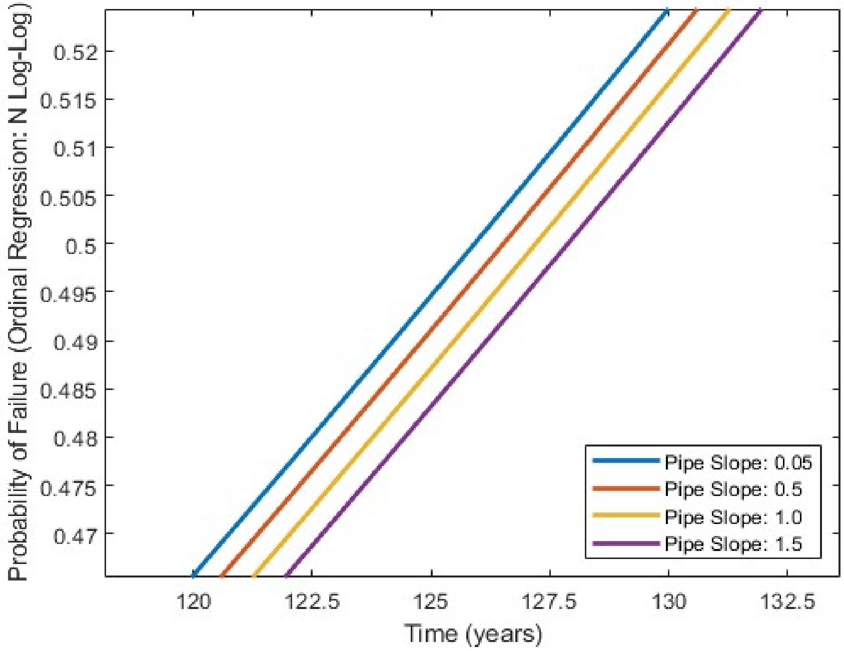

Figure 5.60: Comparing probabilities of failure for various pipe slopes (Ordinal Regression: Negative Log-Log Link Function)

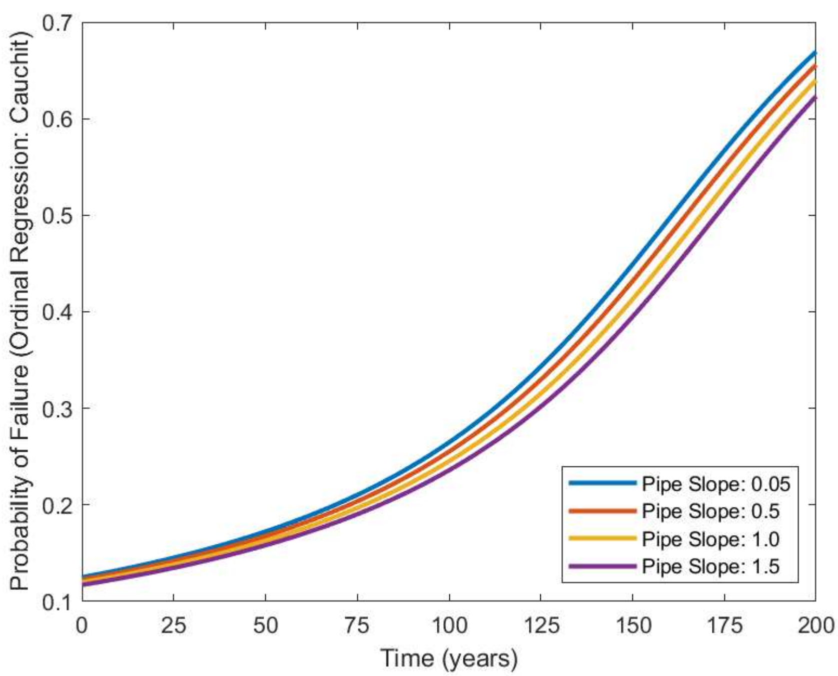

Figure 5.61: Probability of failure with respect to age for various pipe slopes (Ordinal Regression: Cauchit Link Function)

5.8.5 Influence of average velocity of sewer flow on service life of gravity sewer pipes

With regards to the influence of average velocity of sewer flow on service life of the gravity sewer pipes considered in this study, it is observed that the results obtained through statistical models illustrate that when deterioration models are constructed based upon binomial logistic regression, multinomial logistic regression, ordinal regressions using Logit, Complementary Log-Log, and Cauchit link functions, for the same gravity sewer pipe, and while other independent variables of the gravity sewer pipe remain unaltered, by increasing the value of average velocity of the sewer flow in the pipe, the service lives of sewer pipes are decreased.

Furthermore, it is observed that in these deterioration models, by increasing the average sewer velocity, the probability of failure of sewer pipe with respect to age of sewer pipe was subjected to increase. The properties of the sewer flow may have an impact on how the average velocity of the flow affects the service life and probability of failure of the sewer pipe, however, these properties were not available for the study at hand. These observations maybe due to the fact that by increasing the average velocity of the sewer flow, erosion is subjected to increase and therefore, the probability of failure of sewer pipe increases and the service life of sewer pipes are decreased.

On the other hand, when using ordinal regression with negative log-log link function, changes in the values of average velocity of the sewer flow do not change the service life of sewer pipes; this is due to the fact that this parameter was eliminated in this model so that the proportional odds assumption can be validated. Similar observations were made for probability of failure of pipes with respect to age of sewer pipe. Moreover, when ordinal regression with Probit link function is utilized for deterioration modeling of sewer pipes, it is observed that by increasing the average velocity of the sewer flow, the service life of sewer pipe is subjected to small increase in its value. However, based on the results from various methods, it can be concluded that the governing result would be decrease of service life and increase of probability of failure as average velocity of the sewer flow increases, since the majority of the methods have yielded this result.

Moreover, it should be noted that in all of the statistical deterioration models considered herein, the average velocity of sewer flow was not determined to be a significant independent variable of these models. Therefore, this explains why the changes in average flow velocity had different and inconsistent results using different methods.

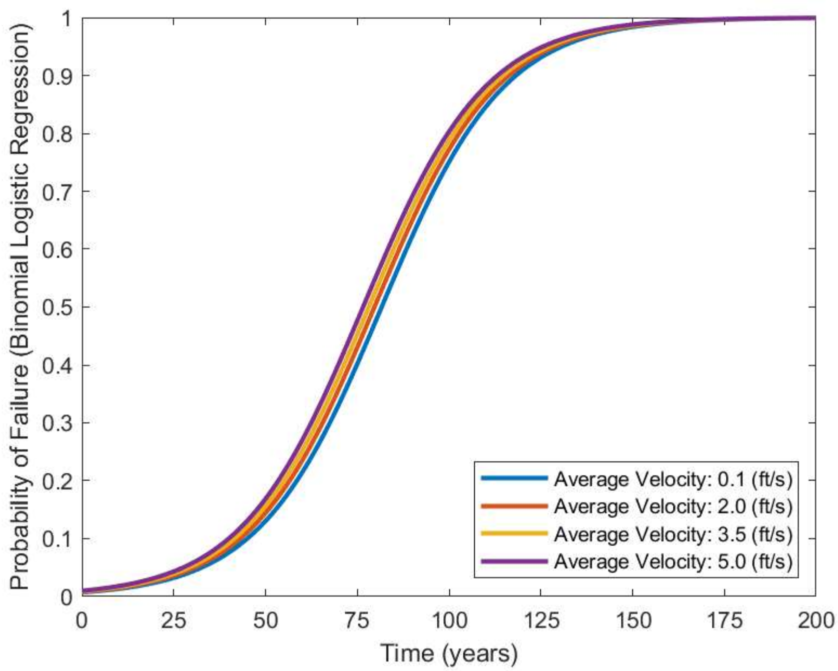

Figure 5.62: Probability of failure with respect to age for different average flow velocities (Binomial Logistic Regression)

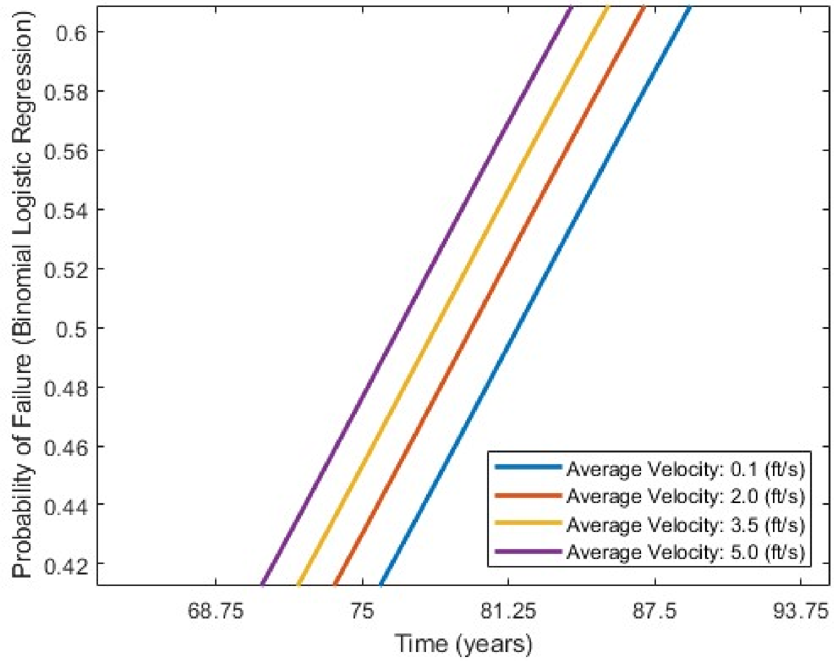

Figure 5.63: Comparing probabilities of failure for different average flow velocities (Binomial Logistic Regression)

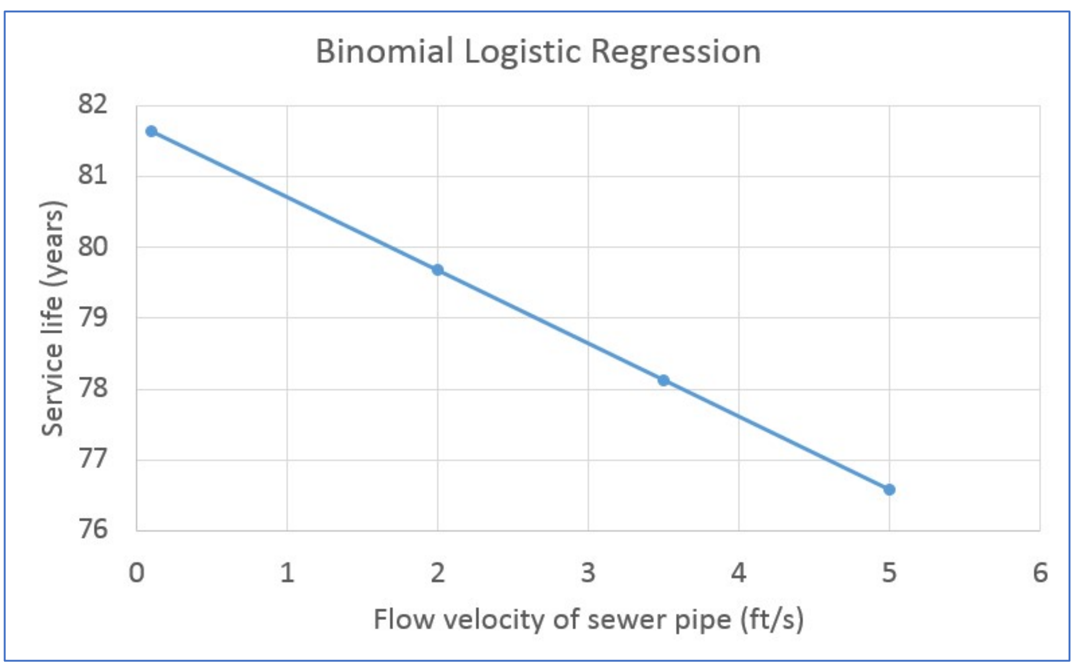

Figure 5.64: Service life of assets for different average flow velocities (Binomial Logistic Regression)

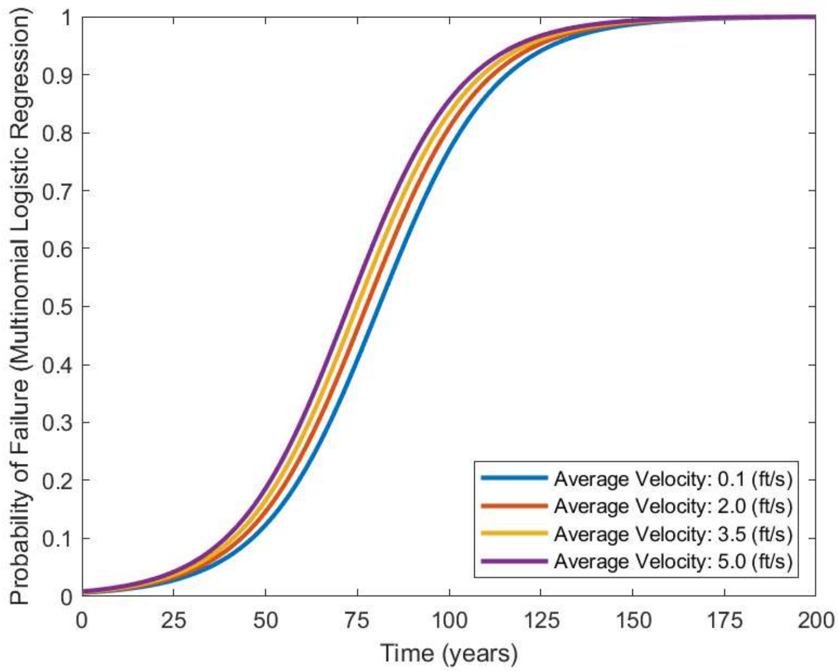

Figure 5.65: Probability of failure with respect to age for different average flow velocities (Multinomial Logistic Regression)

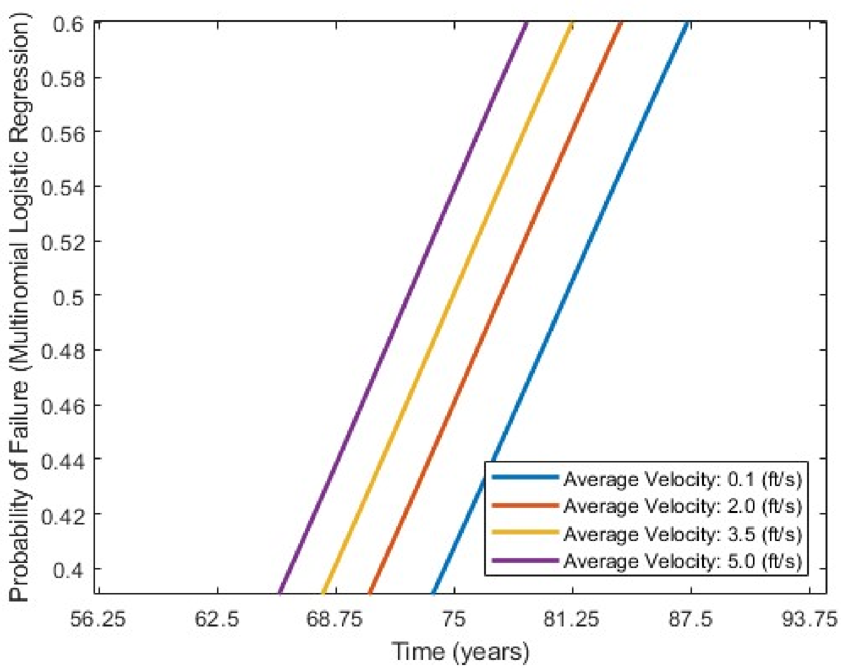

Figure 5.66: Comparing probabilities of failure for different average flow velocities (Multinomial Logistic Regression)

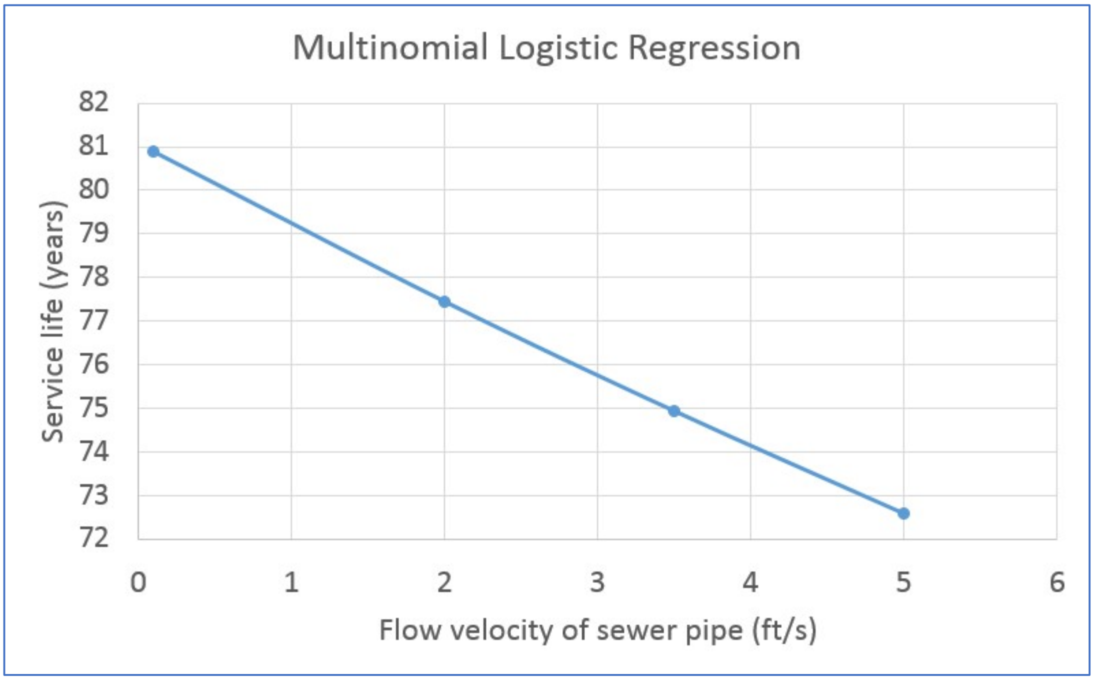

Figure 5.67: Service life of assets for different average flow velocities (Multinomial Logistic Regression)

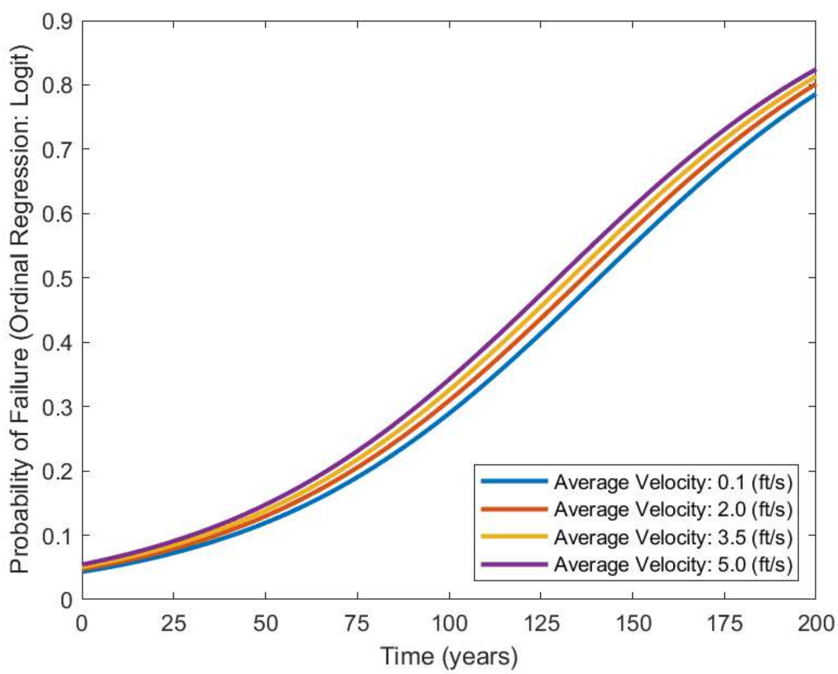

Figure 5.68: Probability of failure with respect to age for different average flow velocities (Ordinal Regression: Logit Link Function)

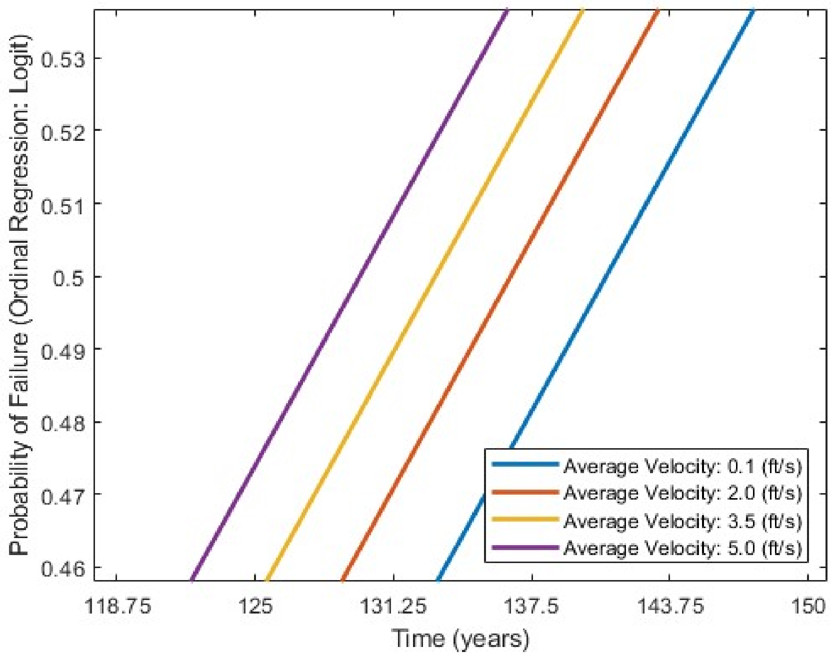

Figure 5.69: Comparing probabilities of failure for different average flow velocities (Ordinal Regression: Logit Link Function)

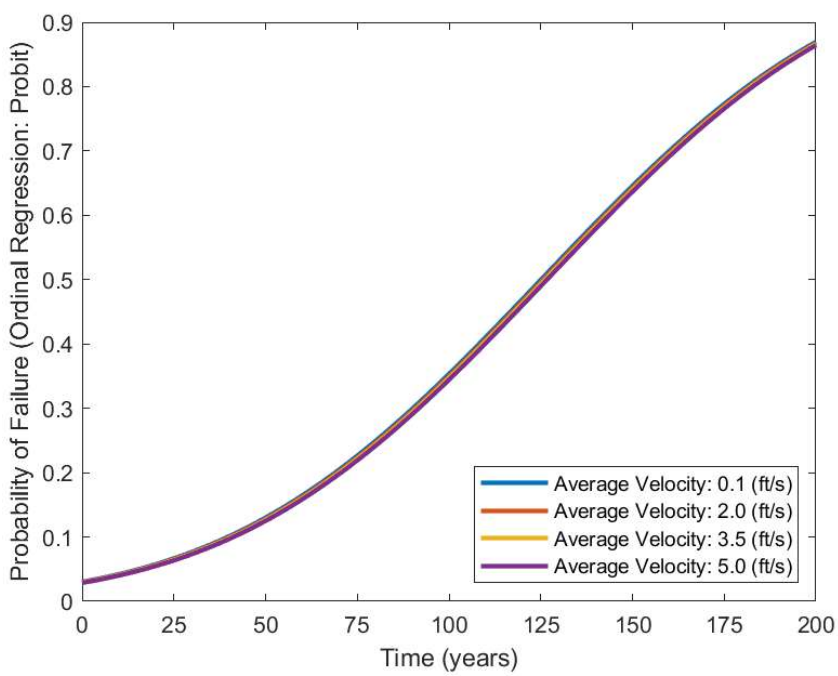

Figure 5.70: Probability of failure with respect to age for different average flow velocities (Ordinal Regression: Probit Link Function)

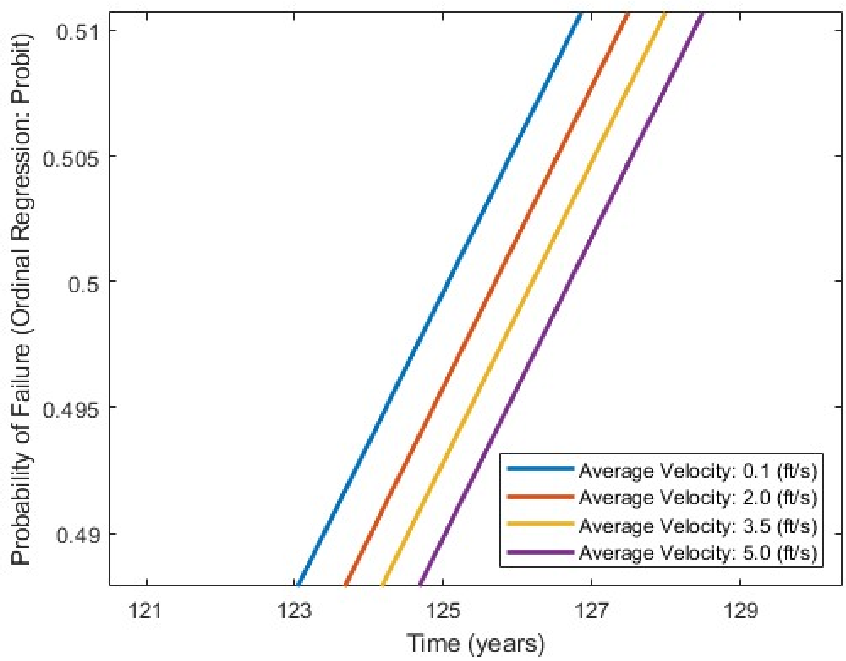

Figure 5.71: Comparing probabilities of failure for different average flow velocities (Ordinal Regression: Probit Link Function)

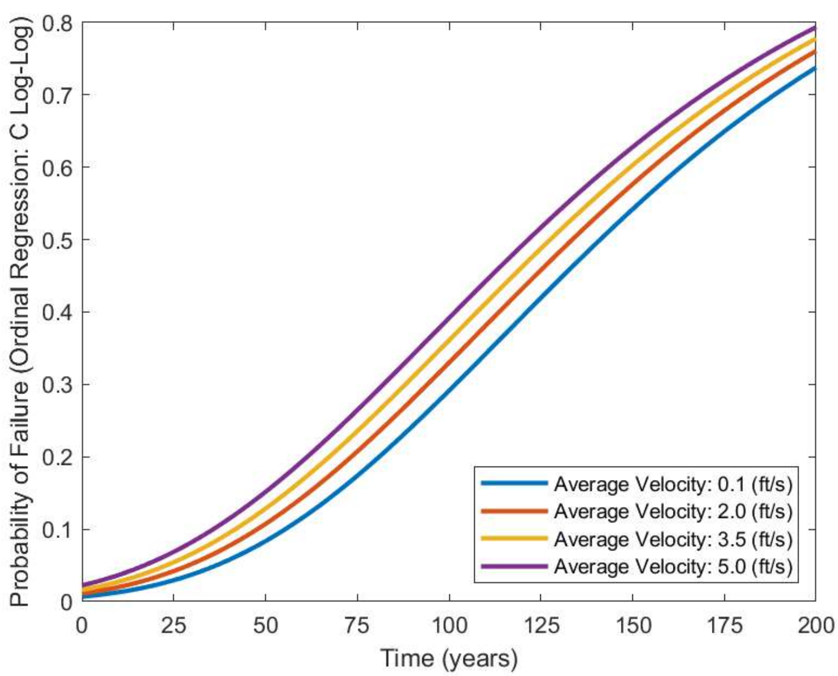

Figure 5.72: Probability of failure with respect to age for different average flow velocities (Ordinal Regression: Complementary Log-Log Link Function)

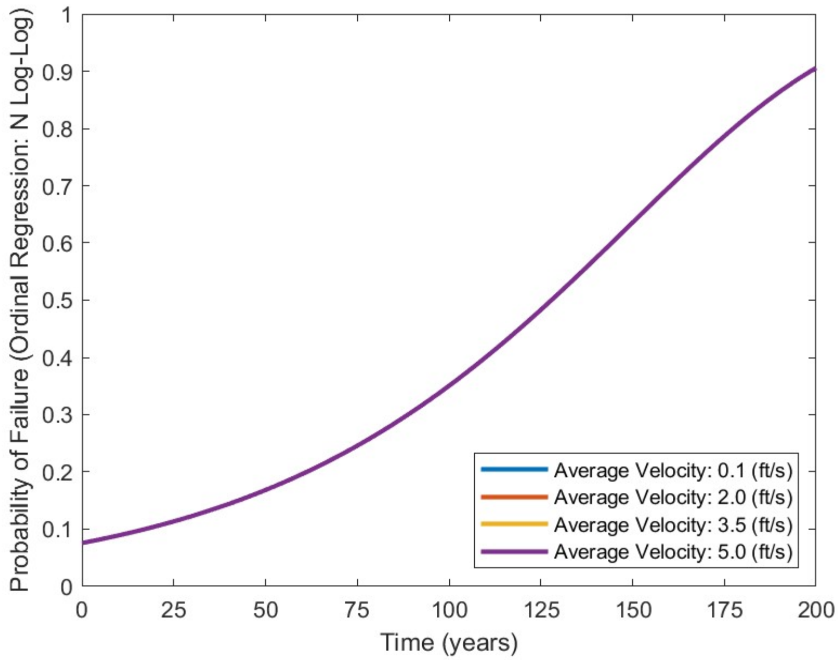

Figure 5.73: Probability of failure with respect to age for different average flow velocities (Ordinal Regression: Negative Log-Log Link Function)

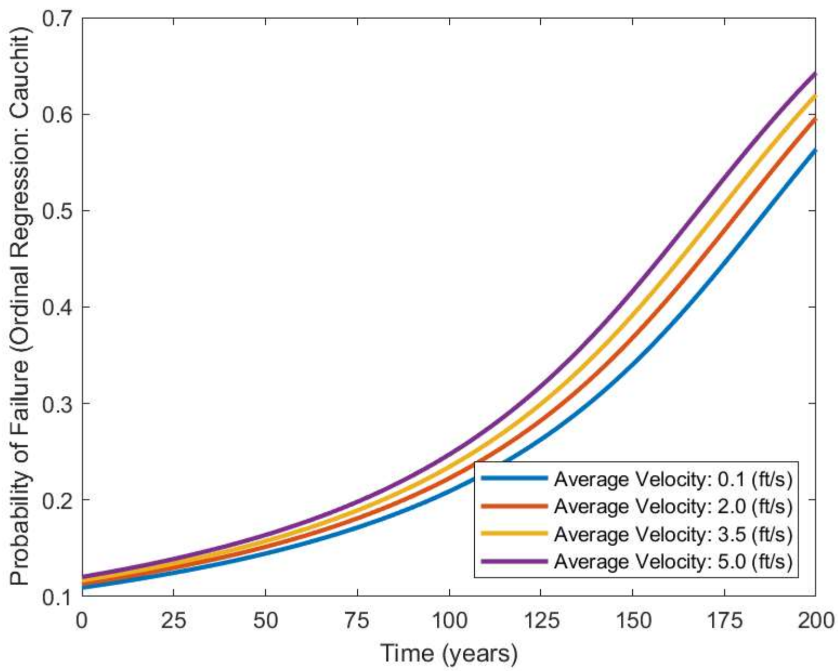

Figure 5.74: Probability of failure with respect to age for different average flow velocities (Ordinal Regression: Cauchit Link Function)

5.8.6 Influence of average flow depth on service life of gravity sewer pipes

Based on the results obtained through the following deterioration models, it is observed that when the average flow depth of pipes is increased, the service life of sewer pipe is subjected to increase and the probability of failure of sewer pipe is subjected to reduction:

Binomial logistic regression

Multinomial logistic regression

Ordinal regression with Probit link function

Ordinal regression with Complementary Log-Log link function

On the other hand, when the following deterioration techniques are utilized, increase in the value of average flow depth in pipe results in decrease in the service life of sewer pipe as well as increase in the associated probability of failure of pipe with respect to age:

Ordinal regression with Logit link function

Ordinal regression with Negative Log-Log link function

Ordinal regression with Cauchit link function

Considering the proportional odds assumption is not satisfied when using Cauchit link function in ordinal regression, it can be concluded that based on the majority of the modeling approaches, by increasing the value of average flow depth in pipe, the service life of sewer pipe will be subjected to increase and the associated probability of failure of pipe will be decreased.

Considering the deterioration models developed herein were based upon various diameters of pipes, therefore, the effect of average flow depth in pipes should be taken into consideration alongside the diameter of the sewer pipe; hence, investigating the effect of average flow in pipe (%full) on service life and probability of failure of sewer pipes, which is presented in the following section, can yield more comprehensive insights.

Additionally, based on the following statistical deterioration models, the average flow depth in pipes was determined to be a significant independent variable of the model:

Binomial logistic regression

Multinomial logistic regression

Ordinal regression using Probit link function

In ordinal regression using Logit link function, prior to elimination of average flow in pipe (%full), the average flow depth in pipe was determined to be significant, however, once the aforementioned variable was eliminated, this predictor was not determined to be a significant independent variable.

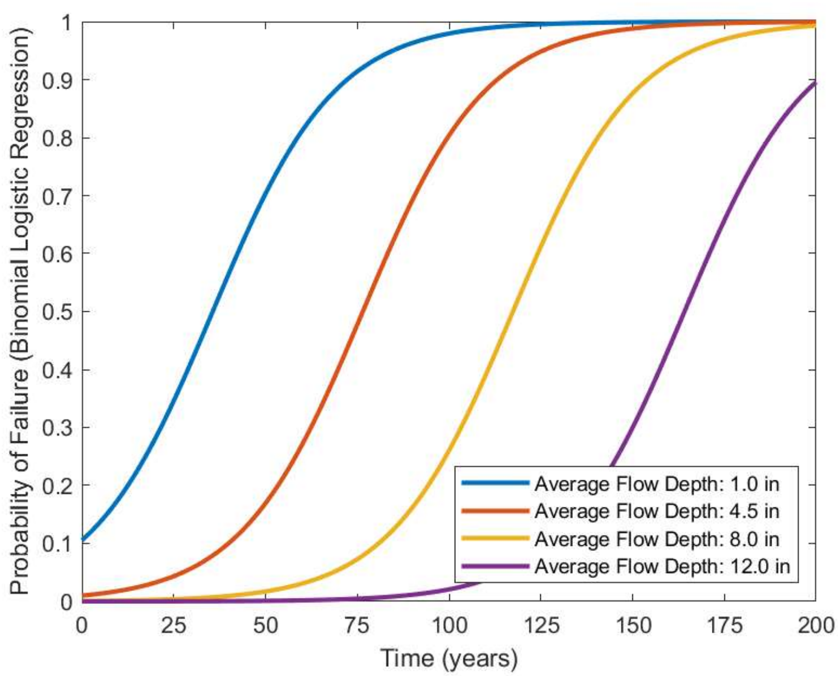

Figure 5.75: Probability of failure with respect to age for various average flow depths (Binomial Logistic Regression)

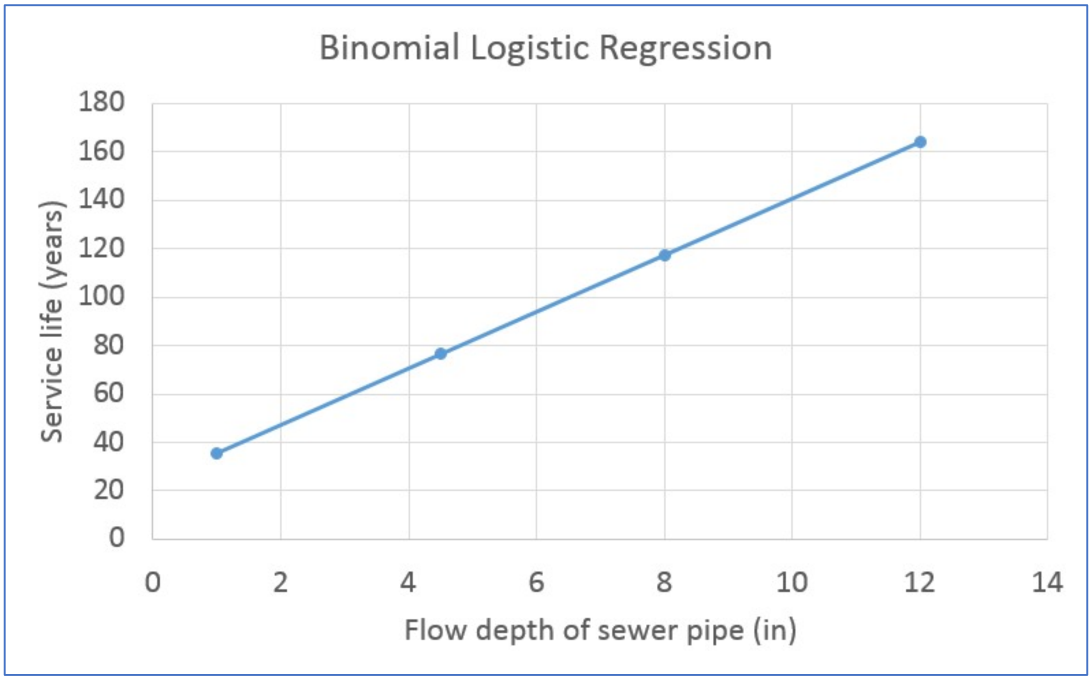

Figure 5.76: Service life of assets for various average flow depths (Binomial Logistic Regression)

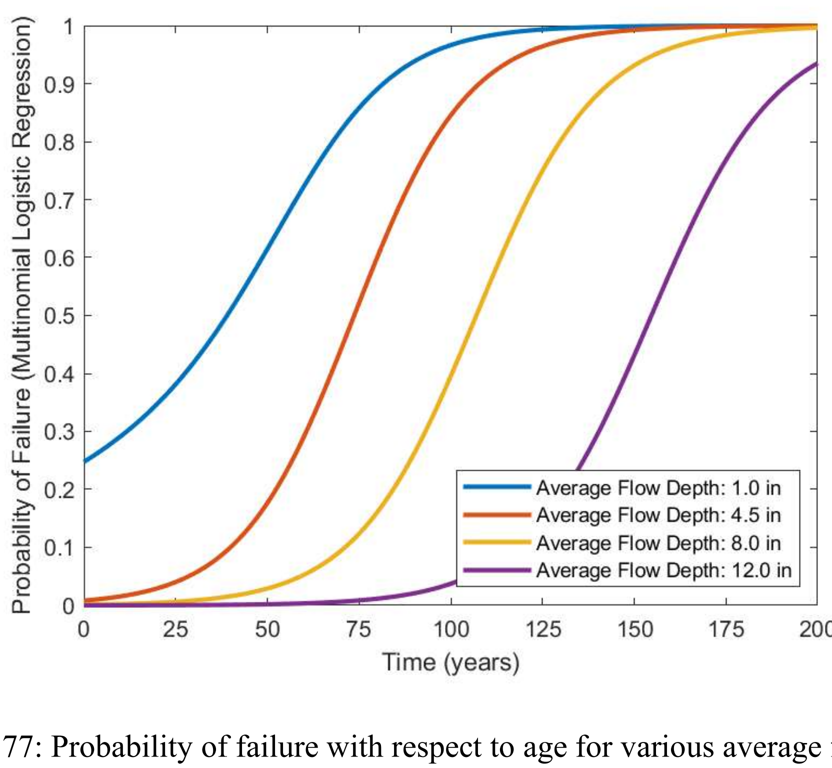

Figure 5.77: Probability of failure with respect to age for various average flow depths (Multinomial Logistic Regression)

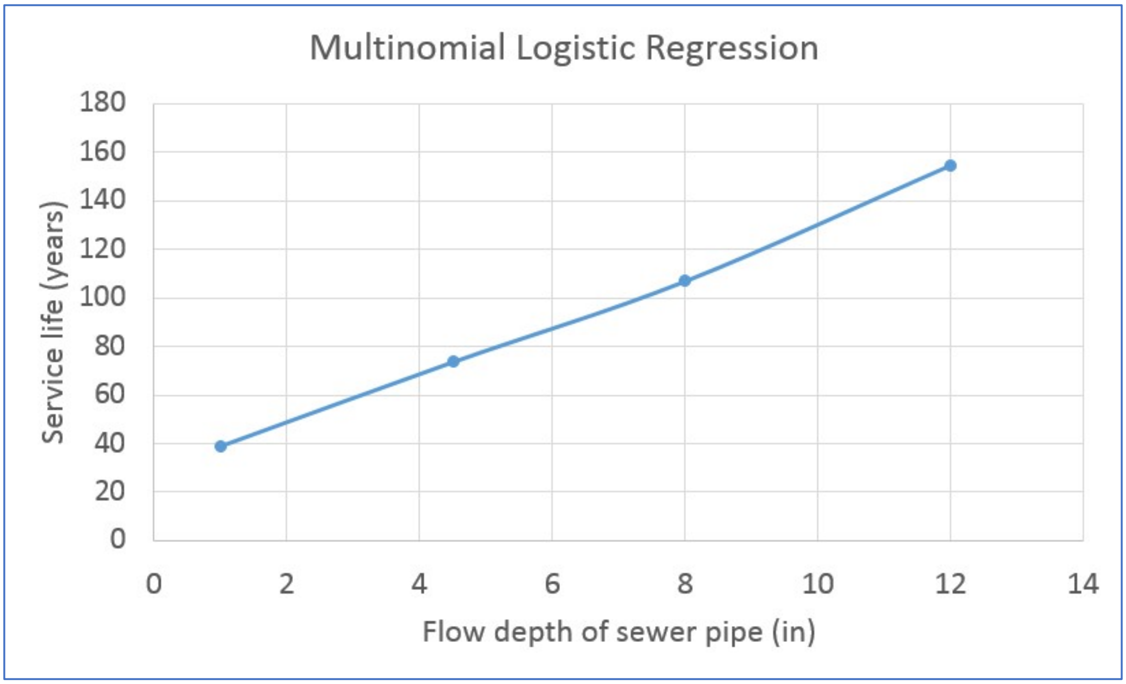

Figure 5.78: Service life of assets for various average flow depths (Multinomial Logistic Regression)

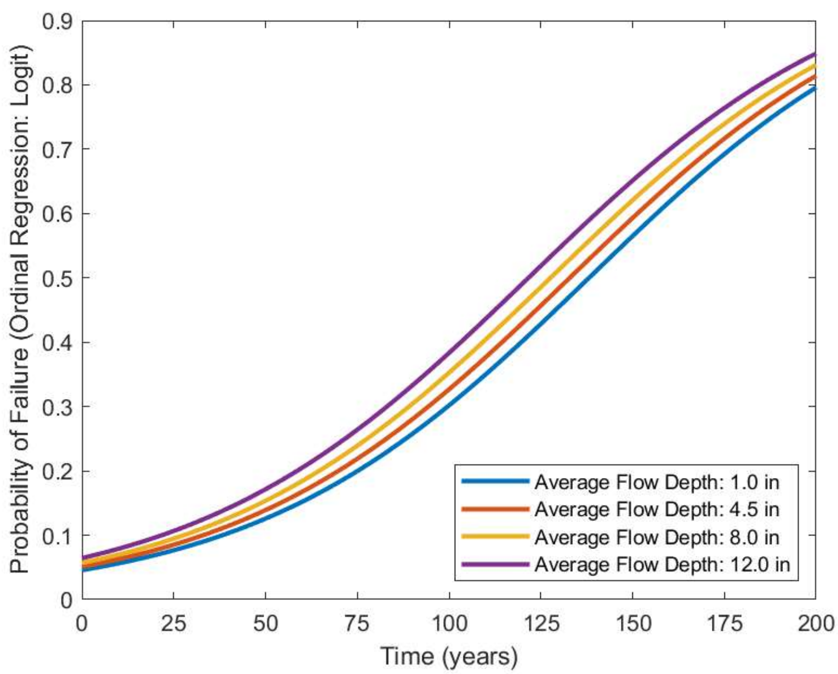

Figure 5.79: Probability of failure with respect to age for various average flow depths (Ordinal Regression: Logit Link Function)

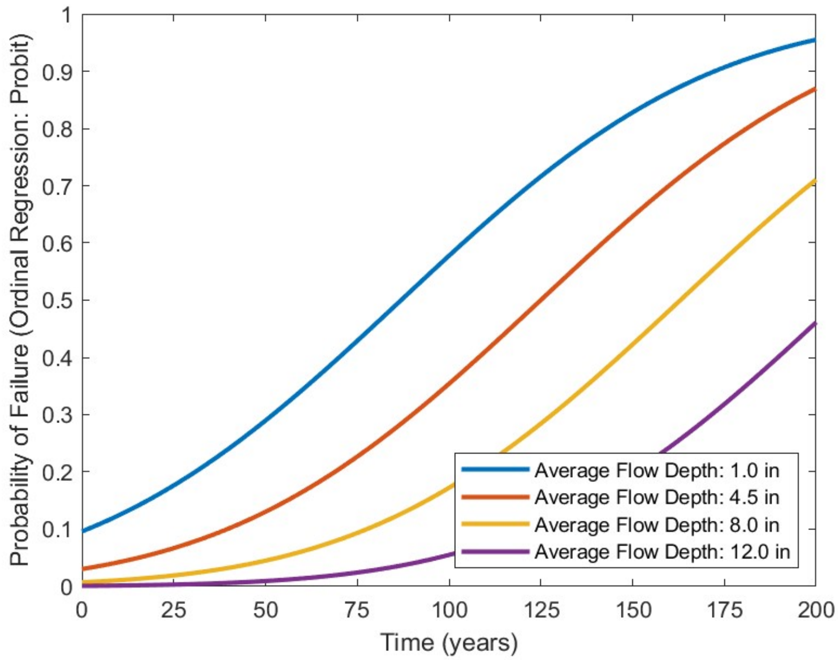

Figure 5.80: Probability of failure with respect to age for various average flow depths (Ordinal Regression: Probit Link Function)

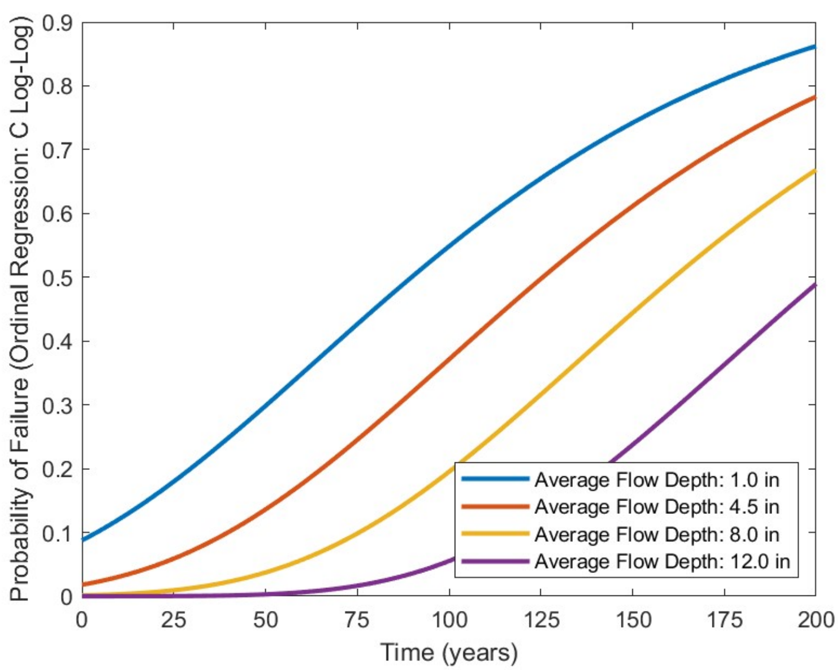

Figure 5.81: Probability of failure with respect to age for various average flow depths (Ordinal Regression: Complementary Log-Log Link Function)

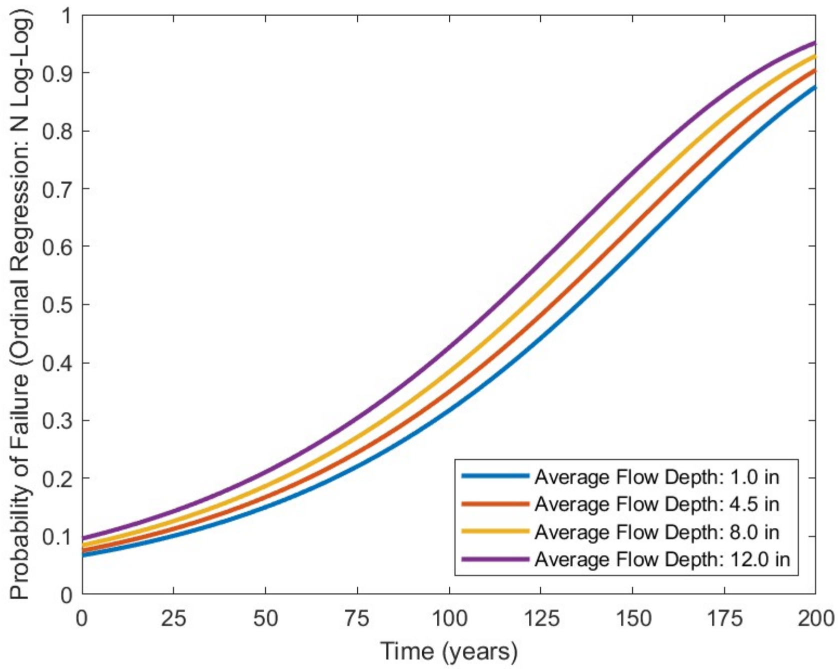

Figure 5.82: Probability of failure with respect to age for various average flow depths (Ordinal Regression: Negative Log-Log Link Function)

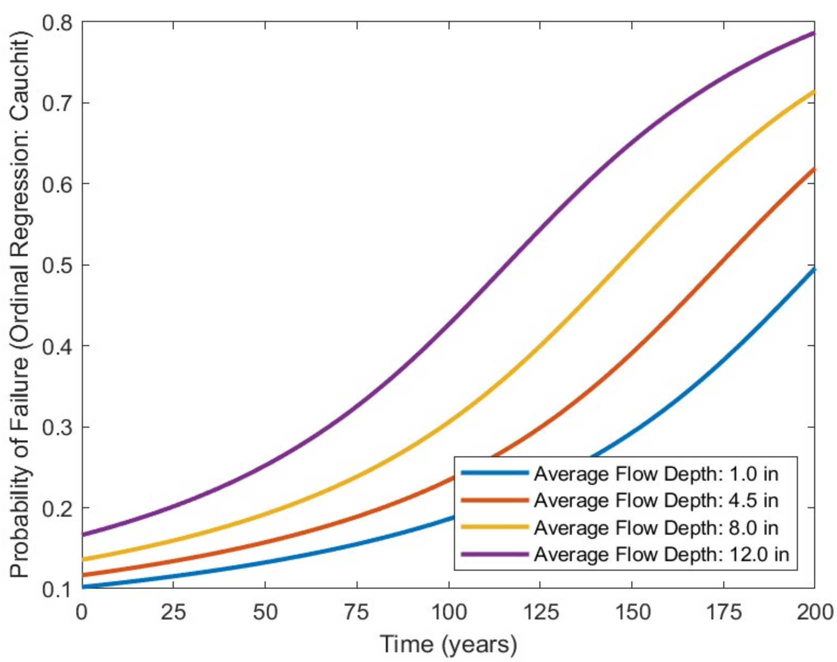

Figure 5.83: Probability of failure with respect to age for various average flow depths (Ordinal Regression: Cauchit Link Function)

5.8.7 Influence of average flow in pipe (%full) on service life of gravity sewer pipes

By investigating the effect of average flow in pipe (%full) on service life and probability of failure of sewer pipes with respect to age, it is observed that by increasing the value of average flow in pipe (%full), the following modeling approaches resulted in reduction of service life of sewer pipes as well as increase in the probability of failure of sewer pipes with respect to age:

Binomial logistic regression

Multinomial logistic regression

Ordinal regression with Probit link function

Ordinal regression with Complementary Log-Log link function

However, since average flow in pipe (%full) was eliminated from the following deterioration models (in order for proportional odds assumption to be validated) thus by using the following models, it was observed that probability of failure of sewer pipes with respect to age and the service life values were not impacted due to changes made in average flow in pipe (%full):

Ordinal regression with Logit link function

Ordinal regression with Negative Log-Log link function

Furthermore, using ordinal regression with Cauchit link function illustrated that by increasing the average flow in pipe (%full), service life of sewer pipes were increased and the probability of failure of sewer pipes with respect to age was decreased; however, as stated earlier, through test of parallel lines, it was observed that Cauchit function did not satisfy the proportional odds assumption and therefore, is solely used for illustration and comparison purposes.

Hence, based on the results from majority of the modeling techniques utilized herein, it can be concluded that increase in the value of average flow in pipe (%full) results in increase in the probability of failure of sewer pipes and therefore, decrease in their associated service life values. This observation maybe due to the fact that when average flow in pipe (%full) increases, therefore, depending on the properties of the flow (which its related data were not available in this study), occurrences such as corrosion, erosion, exfiltration etc. can be increased; moreover, increase in the value of average flow in pipe (%full) can be a sign of infiltration as well. Hence, these occurrences could lead to higher probabilities of failure in sewer pipe and reduction in their service life values.

Moreover, by using the following deterioration models, the average flow in pipe (%full) was found to be a significant independent variable:

Binomial logistic regression

Multinomial logistic regression

Ordinal regression with Probit link function

Ordinal regression with Complementary Log-Log link function

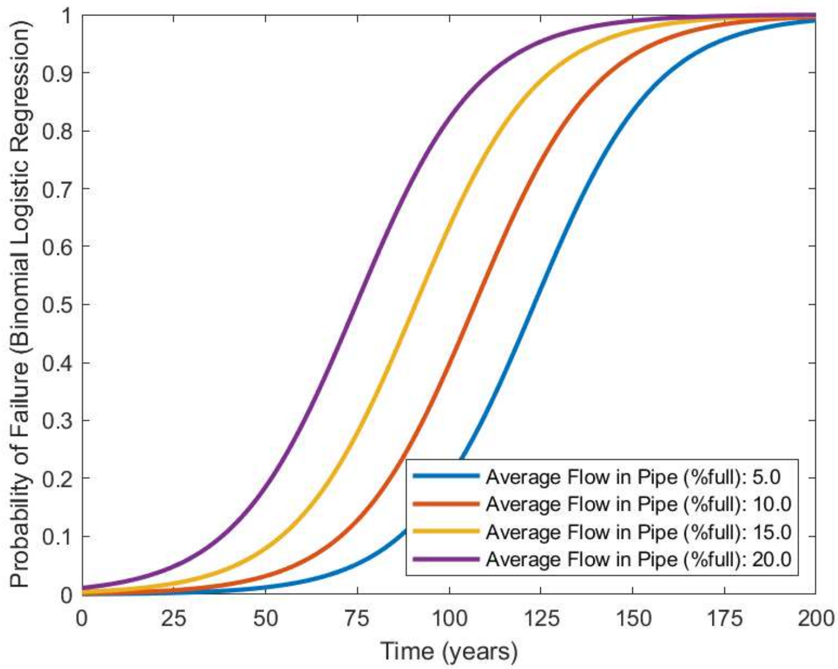

Figure 5.84: Probability of failure with respect to age for different values of average flow in pipes (%full) (Binomial Logistic Regression)

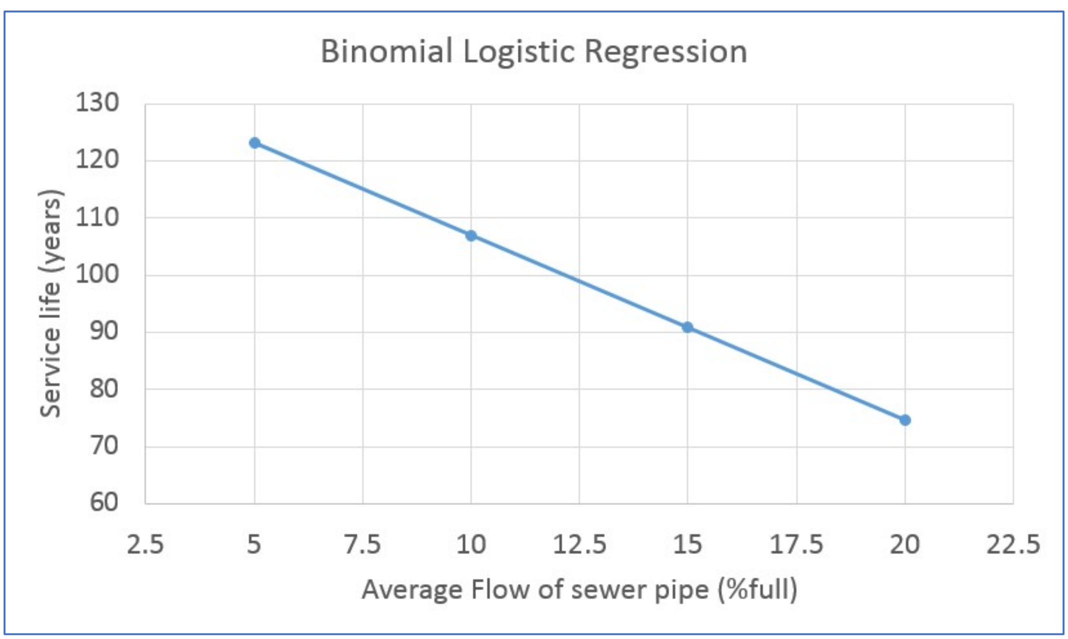

Figure 5.85: Service life of assets for different values of average flow in pipes (%full) (Binomial Logistic Regression)

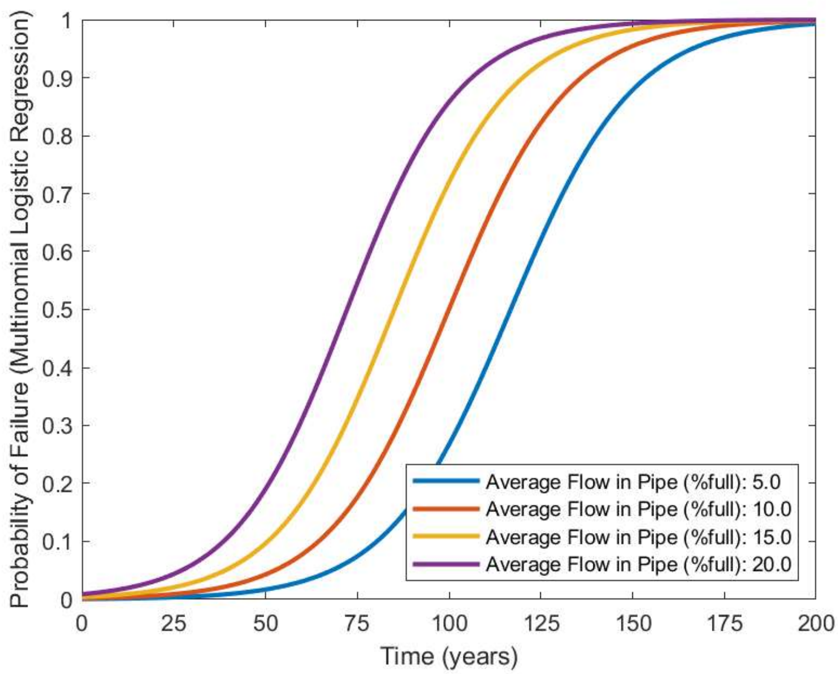

Figure 5.86: Probability of failure with respect to age for different values of average flow in pipes (%full) (Multinomial Logistic Regression)

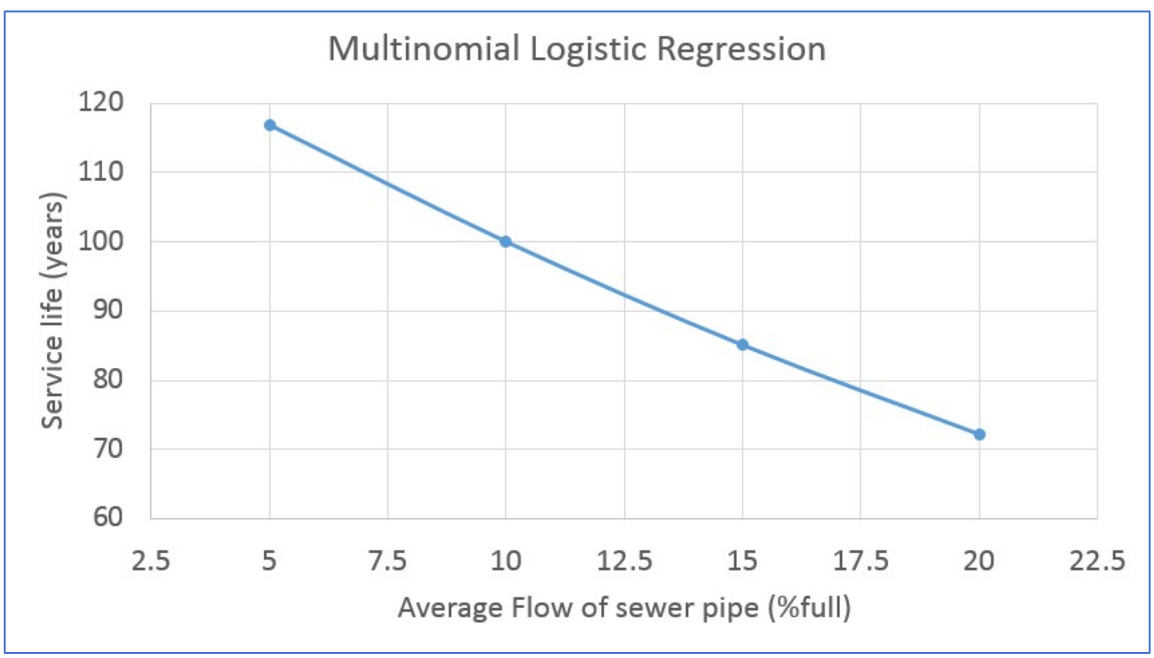

Figure 5.87: Service life of assets for different values of average flow in pipes (%full) (Multinomial Logistic Regression)

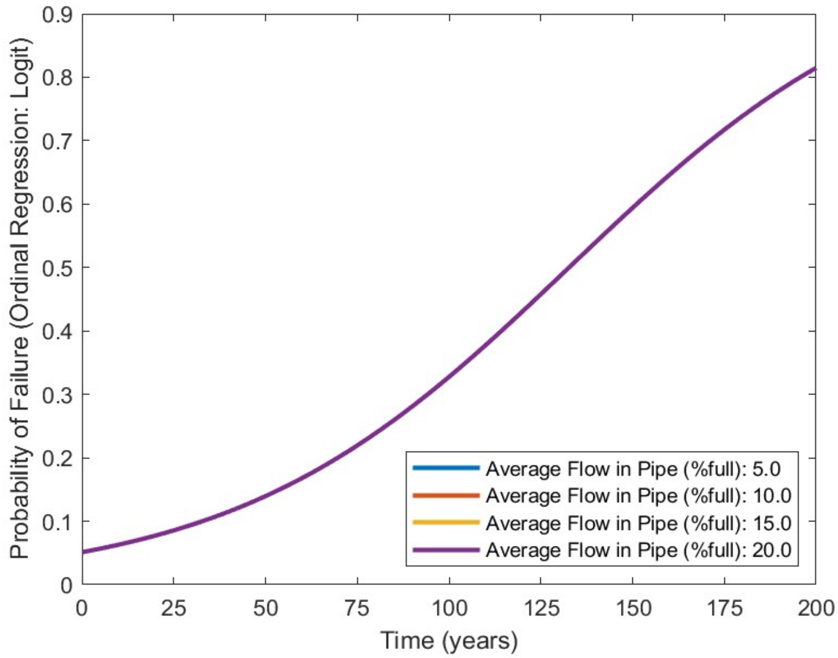

Figure 5.88: Probability of failure with respect to age for different values of average flow in pipes (%full) (Ordinal Regression: Logit Link Function)

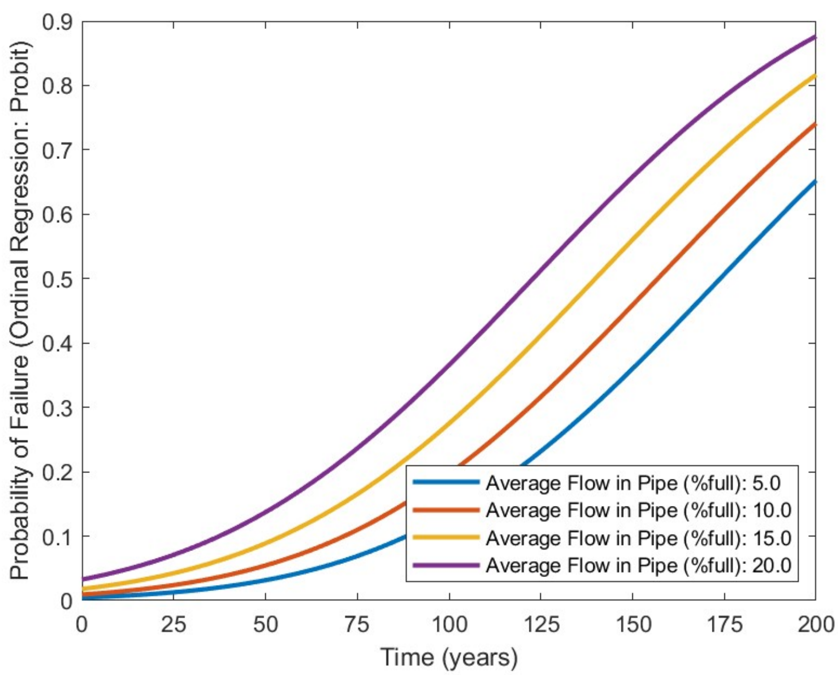

Figure 5.89: Probability of failure with respect to age for different values of average flow in pipes (%full) (Ordinal Regression: Probit Link Function)

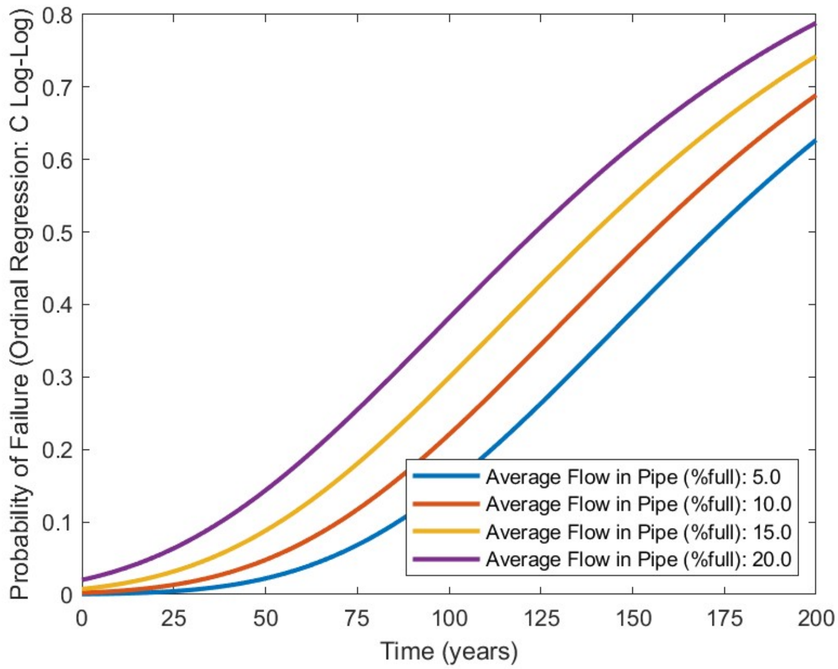

Figure 5.90: Probability of failure with respect to age for different values of average flow in pipes (%full) (Ordinal Regression: Complementary Log-Log Link Function)

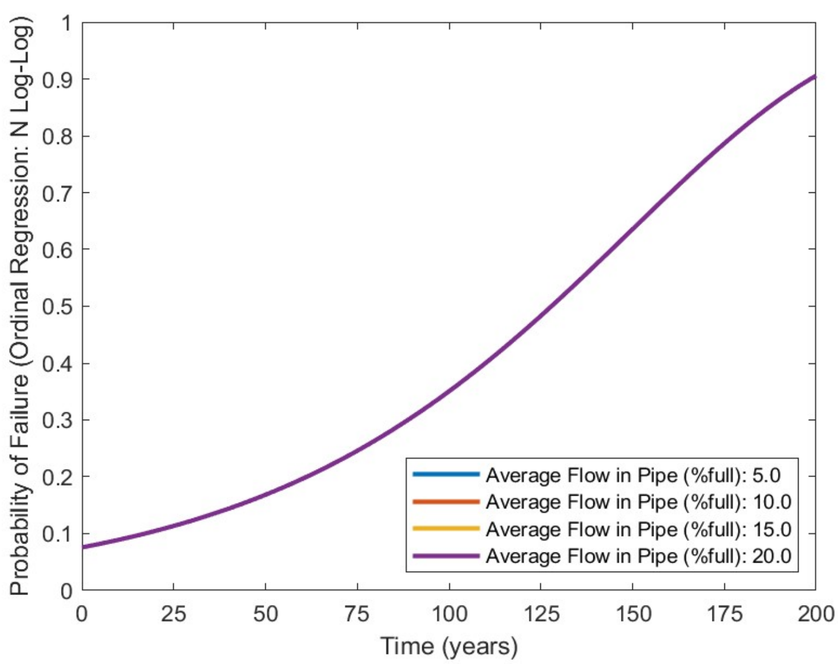

Figure 5.91: Probability of failure with respect to age for different values of average flow in pipes (%full) (Ordinal Regression: Negative Log-Log Link Function)

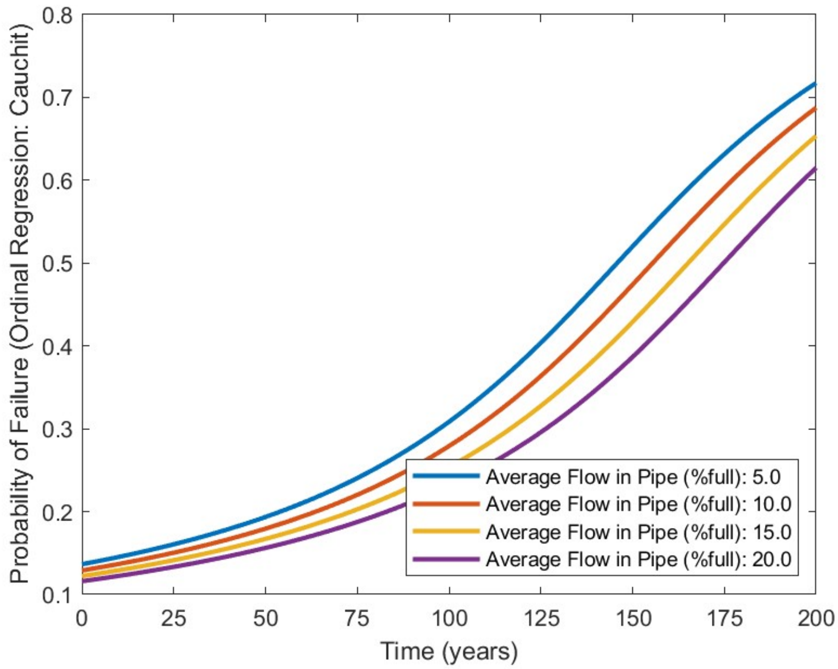

Figure 5.92: Probability of failure with respect to age for different values of average flow in pipes (%full) (Ordinal Regression: Cauchit Link Function)

Chapter 6 : Summary of Results and Concluding Remarks

In the study at hand, statistical and artificial intelligence based models were utilized to estimate the service life of sewer pipes. Statistical models applied for this purpose are as follows:

- Binomial logistic regression
- Multinomial logistic regression
- Ordinal regression (with various link function)
- LightGBM modeling
- CatBoost algorithm

With regards to ordinal regression, suitability of various available link functions were tested. These link functions are as follows:

- Logit function
- Probit function
- Complementary log-log function
- Negative log-log function
- Cauchit function

Based on the results obtained through test of parallel lines, it was realized that the proportional odds assumption cannot be validated for ordinal regression using Cauchit link function. However, for illustration and comparison purposes, the results associated with this link function are also included as well.

Furthermore, when utilizing statistical models, two different approaches were taken; in the first approach it was assumed that the initial condition grading of sewer pipes was not known; and in the second approach, it was assumed that the sewer pipes were in perfect condition at the start of their operation. In other words, in the second approach it was assumed that the installations of sewer pipes were conducted perfectly and no initial damages or inherent defects were present in sewer pipes.

Additionally, in this study, the effects of various independent variables for the inspected sewer pipes were investigated as well. The independent variables considered herein are as follows:

- Age of pipe
- Diameter of pipe
- Pipe material
- Average velocity of sewer flow
- Average flow depth
- Pipe slope
- Length of pipe
- Average flow in pipe (percent full)

In addition to the aforementioned independent variables, the effect of population growth on the estimated service life of sewer pipes was investigated as well. The annual rate of population growth utilized in this study was assumed to be 0.01. It should be noted that by using more pertinent independent variables the models may yield more accurate estimations. In this study however, only the data related to the aforementioned independent variables were available for developing deterioration models.

6.1 Influence of various predictors on service life of gravity sewer pipes

In this dissertation, based on the models developed without considering the initial condition of sewer pipes (as these condition gradings were not available), the influence of various predictors on service life of gravity sewer pipes as well as the probability of failure associated with them were investigated. Based on the observations made herein, the summary of results pertaining to effect of each available independent variable considered in this study are stated as follows:

6.1.1 Summary of effect of pipe material on service life of sewer pipes

The deterioration models developed using binomial logistic regression, multinomial logistic regression, ordinal regressions with various Logit, Probit, complementary log-log, negative log-log, and Cauchit link functions, demonstrate that for the sewer pipes considered in this study, and considering values of all other independent variables are the same, the highest to lowest values of service life associated with various pipe materials are as follows:

- FRP
- PVC
- RCP
- VCP

Additionally, the aforementioned pipe materials correspond to lowest to highest probability of failure of sewer pipes with respect to age of assets. Various factors which potentially play a role on how pipe material influences the service life are discussed in this dissertation. For instance, FRP and PVC are flexible pipes, whereas RCP and VCP are brittle and therefore, more susceptible to higher condition states especially structural defects.

6.1.2 Summary of effect of pipe diameters on service life of sewer pipes

The results obtained based on statistical models, i.e. binomial logistic regression, multinomial logistic regression, ordinal regressions using various link functions (Logit, Probit, complementary log-log, negative log-log, and Cauchit link functions), illustrated that by increasing the sewer pipe diameter, reduction in the values of service lives of sewer pipes are observed and moreover, the associated probability of failure of assets with respect to age of pipes will increase. Factors such as differences in bedding and backfill conditions, the installation process of sewer pipes, and depth of the assets can be influential herein; however, the data pertaining to these factors were not available.

Furthermore, diameter of sewer pipe was determined to be a significant independent variable when applying binomial logistic regression, multinomial logistic regression, ordinal regression with Probit and Complementary Log-Log link function. It was further realized that before eliminating predictors in ordinal regression with Negative Log-Log and Logit link function (in order for them to satisfy test of parallel lines), diameter of sewer pipe was a significant variable in these models as well, however, by removing predictors in those models, diameter of pipe was no longer significant.

6.1.3 Summary of effect of average velocity of sewer flow on service life of sewer pipes

When examined based on deterioration models constructed through binomial logistic regression, multinomial logistic regression, ordinal regressions using Logit, Complementary Log-Log, and Cauchit link functions, it was observed that when the value of average velocity of the sewer flow in the pipe is increased, the associated service lives of pipes are then decreased and this also resulted in probabilities of failure of sewer pipes with respect to age of assets to increase. The decline in the service life of assets due to increased flow velocity maybe due to increased occurrences of erosion, and flow properties.

Average velocity of flow was removed from ordinal regression with negative log-log link function in order to satisfy the proportional odds assumption and thus the service life and probability of failure of pipes do not vary due to changes in average flow velocity. Furthermore, based on the model obtained from ordinal regression with Probit link function increase in the average velocity of the sewer flow, results in small decline in the service life of sewer pipe. Due to the results obtained from the majority of the deterioration models, it can be concluded that the governing result is increase in probability of failure and decline in service life due to increase in average velocity of the sewer flow.

Based on the deterioration models considered herein, the average velocity of sewer flow was not found to be a significant predictor in these models, hence, explaining different and inconsistent results associated with effects of this variable in different methods.

6.1.4 Influence of population growth on service life of gravity sewer pipes

Considering an annual population growth of 0.01, it was observed that when using artificial intelligence based models, i.e. LightGBM and CatBoost models, and by taking into account the effect of population growth, the service life of sewer pipes are subjected to reduction. Furthermore, when population growth is taken into consideration, the probability of failure of gravity sewer pipes are also increased. The same observation was made when using statistical deterioration models, as illustrated in previous sections. The reason behind reduction of service life of sewer pipes due to population growth maybe due to the fact that when population increases, the volume of the sewer flow is therefore increased as well; this subsequently results in more occurrences of blockages, overflows, corrosions, erosions, etc. that may occur due to the sewer and its properties (such as amount of debris carried by the sewer, acidity, alkalinity, etc. of the sewer flow). These factors can also contribute to the increase in proximities of failure of sewer pipes with respect to age of pipes.

6.1.5 Summary of effect of length of sewer pipes on service life of sewer pipes

Based upon deterioration models obtained using binomial logistic regression, ordinal regression with Logit, Probit, Complementary Log-Log, and Cauchit link functions, it was realized that changes in the length of sewer pipe resulted in no variation in the service life and probability of failure of sewer pipes. Additionally, multinomial logistic regression and ordinal regression with Negative Log-Log link function illustrated that the service life and probability of failure of assets were slightly affected by increase and decrease in their values, respectively. Hence, based on the majority of models, it can be stated that the service life and probability of failure of assets are not notably impacted due to variations made in the length of sewer pipes.

Based on the statistical models considered herein, length of assets was not found to be a significant predictor. Therefore, this verifies the aforementioned results that this independent variable does not seem to influence the service life of sewer pipe.

6.1.6 Summary of effect of slope of sewer pipes on service life of sewer pipes

By increasing the slope of sewer pipe, models based on binomial logistic regression, multinomial logistic regression, and ordinal regressions using Probit and Complementary Log-Log link functions, demonstrated slight increase and decrease in the probability of failure and service life of sewer pipes, respectively. However, utilizing ordinal regressions with Logit, Cauchit, and Negative Log-Log link functions showed that due to increase in the pipe slope, the probability of failure was subjected to slight reduction and therefore, service life of assets increased. Taking into account that Cauchit link function does not satisfy proportional odds assumption and is solely presented for comparison purposes, it can thus be stated that based on the majority of the models, when the slope of pipe is increased, service life and the probability of failure of sewer pipes are slightly subjected to decline and increase, respectively.

However, the results from statistical models illustrated that the slope of pipe was not a significant predictor in these models; thus, verifying the slight and varying impact of this independent variable on service life and probability of failure of assets.

6.1.7 Summary of effect of average flow depth on service life of sewer pipes

The outcome of deterioration models based on binomial logistic regression, multinomial logistic regression, and ordinal regressions using Probit and Complementary Log-Log link functions show that by increasing the average flow depth in assets, the values of service life and probability of failure of sewer pipes are increased and decreased, respectively; moreover, the results obtained utilizing ordinal regressions with Logit, Negative Log-Log, and Cauchit link functions illustrate that when the average flow depth in pipe is increased, the service life and the associated probability of failure of assets will be subjected to reduction and increase, respectively. Therefore, as Cauchit link function is only used for comparison purposes, the majority of techniques demonstrate that effect of increasing the average flow depth in pipes will result in an increase in service life and decrease in probability of failure of sewer pipes. However, it should be noted that since the models were developed using different pipe sizes, the average flow depth of the pipe should be considered alongside the diameter of the pipe, hence, more accurate insights can be achieved by considering the influence of average flow in pipe (%full) as well.

Additionally, average flow depth in pipe was found to be significant using binomial logistic regression, multinomial logistic regression, and ordinal regression with Probit link function; prior to elimination of average flow in pipe (%full) in ordinal regression using Logit link function, average flow in pipe (%full) was also a significant predictor, however, by removing the aforementioned variable, average flow depth was not determined to be a significant predictor.

6.1.8 Summary of effect of average flow in pipe (%full) on service life of sewer pipes

Considering models developed through binomial logistic regression, multinomial logistic regression, ordinal regressions using Probit and Complementary Log-Log link functions, it is observed that when the average flow in pipe (%full) increases, service life of assets are reduced and the probability of failure of pipes are increased. Furthermore, as average flow in pipe (%full) was eliminated from ordinal regressions with Logit and Negative Log-Log link functions so that these models satisfy proportional odds assumption, therefore, when using these models functions, this predictor did not affect the service life and probability of failure of sewer pipes. However, results from Cauchit link function showed that increase in average flow in pipe (%full) resulted in increase in service life and decrease in probability of failure of sewer pipes. It should be noted that Cauchit function is only used for comparison purposes (it didn't validate proportional odds assumption).

Therefore, the majority of the models illustrate decrease in service life and increase in the probability of failure of sewer pipes when average flow in pipe (%full) is increased. The factors resulting increase in the probability of failure can be associated with the flow properties (its data were not available), which can increase erosion, corrosion, exfiltration, etc.; additionally, increase in the average flow in pipe (%full) maybe due to infiltration as well, which can further increase the probability of failure of asset and reduce the service life.

Furthermore, based on models obtained using binomial logistic regression, multinomial logistic regression, ordinal regression with Probit and Complementary Log-Log link function, the average flow in pipe (%full) was determined to be a significant predictor.

6.2 LightGBM and CatBoost models

A summary of comparison between models developed based on LightGBM and CatBoost algorithms are presented herein.

6.2.1 Receiver Operating Characteristic Curves Associated with LightGBM and CatBoost models

It was observed that the area underneath receiver operating characteristic curve (AUC) associated with CatBoost algorithm was greater than AUC associated with LightGBM model. Based on the obtained results, it is observed that the area under the receiver operating characteristic curves associated with LightGBM model is 0.85, whereas AUC for CatBoost algorithm is computed to be 0.88. The greater area under the receiver operating characteristic curve associated with CatBoost algorithm, illustrates a more suitable model for the available sewer pipe data set.

Based on the receiver operating characteristic curves for both LightGBM and CatBoost, it is observed that for the sewer pipe data set considered in this study, CatBoost seemingly results in a more suitable model compared to LightGBM (except for the beginning part of the graphs and a small portion after that). Based on the receiver operating characteristic curves, except for the cases wherein the false positive rate is less than 0.0256 or the false positive rate is between 0.1731 and 0.1795 (in which LightGBM seems to be a more suitable model), in all other instances, CatBoost algorithm will deliver a more suitable model for the available sewer pipe data set.

6.2.2 Feature Importance Associated with LightGBM and CatBoost models

Based on the results obtained from LightGBM and CatBoost models, the independent variables were ranked from the highest to the lowest feature importance. It is observed that the age of the sewer pipe has the highest feature importance in CatBoost model, whereas in the LightGBM model, the highest feature importance is associated with the pipe length. Furthermore, pipe slope, pipe dimeter, PVC, and VCP have the same ranks in both methods.

6.3 Influence of initial condition grading of sewer pipes on service life

When initial condition grading of sewer pipes is taken into consideration in developing the deterioration models, and given the sewer pipes are initially at perfect condition, the deterioration models will be more conservative. In other words, the deterioration rate of sewer pipes will be potentially greater compared to the rate of deterioration in models developed without considering the initial condition grading of sewer pipes.

For instance, in the case of a model developed through binomial logistic regression, when the initial condition grading of sewer pipes are not accounted for, the coefficient of the age of the pipe is equal to 0.060; however, when the initial condition grading of sewer pipes are considered, the coefficient of the binomial logistic regression model associated with age of pipe is equal to 0.100. In other words, when the initial condition grading of sewer pipes are not taken into consideration, when age of pipe increases by one year, the odds ratio increases by 6.18%; however, by considering the initial condition grading of sewer pipes, based on binomial logistic regression model, by increasing the age of pipe by one year, the odds ratio will be subjected to 10.52% increase. Based on this observation, it is concluded that for the sewer pipes considered in this study, by considering the initial condition grading of pipe, the results will be more conservative compared to the case wherein the initial condition grading of sewer pipes are not accounted for.

6.4 Summary of Results Based on Binomial Logistic Regression:

When modeling deterioration of the sewer pipes considered in this study, it was observed that binomial logistic regression was a suitable modeling approach (based on Ombinus tests of model coefficients as well as Hosmer and Lemeshow test). Furthermore, the classification table based on this deterioration model illustrated 83.2% overall accuracy. However, the percent of correct predictions associated with survival and failure of sewer pipes (condition gradings of 0 and 1 in binomial logistic regression) were 95.7% and 16.9%, respectively. Therefore, it is observed that the percent correct predictions of sewer pipes which did not meet failure criterion was much higher compared to sewer pipes which were considered to be in failure condition.

When initial condition gradings of sewer pipes are also taken into consideration, it is observed that both Ombinus tests of model coefficients as well as Hosmer and Lemeshow test are satisfied. Furthermore, based on the classification table it was observed that the percent of correct predictions associated with survival and failure of sewer pipes were 97.6% and 16.9%, respectively. Moreover, the overall percent of correct predictions was computed to be 91.2%, which is greater than the case wherein the initial condition grading of sewer pipes were not accounted for. However, similar to the model without initial condition grading of sewer pipes, it is observed that the correct predictions obtained with regards to sewer pipes which did not meet failure criterion was much more accurate compared to sewer pipes which were considered to be in failure condition.

6.5 Summary of Results Based on Multinomial Logistic Regression:

The deterioration model developed based on multinomial logistic regression was found to be suitable based on Goodness-of-Fit test as well as likelihood ratio test available within model fitting information table. Additionally, based on the classification table for the developed model, it was observed that condition gradings 1 and 4 had 88.7% and 72.9% accuracy in predictions, respectively; whereas condition gradings 2, 3, and 5 had only 1.8%, 11.5%, and 33.3% correct predictions, respectively. The overall percent of correct predictions was found to be 59.3%. Therefore, based on these results, it seems that for the sewer pipes considered in this study, multinomial logistic regression model is best suited for predicting condition gradings of 1 and 4, and condition gradings 2, 3, and 5 were poorly predicted.

When initial condition states of sewer pipes are accounted for, it is observed that multinomial logistic regression once again satisfies Goodness-of-Fit test test as well as likelihood ratio test available within model fitting information table. With regards to classification table, it is observed that the overall percent of correct predictions is equal to 79.8%; which is greater than the overall correct predictions obtained without considering the initial condition grading of sewer pipes (20.5% higher). Once again, the highest percent of correct predictions are associated with condition gradings 1 and 4 with 96.8% and 72.9%, respectively. For condition gradings 2, 3, and 5, the percent of correct predictions are respectively 9.1%, 3.8%, and 16.7%. Even though by considering the initial condition grading of sewer pipes the percent of correct predictions are increased for the overall condition grading prediction (from 59.3% to 79.8%) as well as condition gradings 1 (from 88.7% to 96.8%) and 2 (from 1.8% to 9.1%), however, for condition gradings 3 and 5, the percent of correct predictions have declined from 11.5% to 3.8% and from 33.3% to 16.7%, respectively; and for condition grading 4, this percentage remains constant at 72.9%.

6.6 Summary of Results Based on Ordinal Regressions:

The summary of results associated with each of link functions utilized in this study are as follows:

6.6.1 Link function: Logit

When ordinal regression with Logit link function was used to estimate the service life of sewer pipes, it was observed that even though the obtained model satisfied Goodness-of-Fit test and likelihood ratio test available within model fitting information table, however, considering the result pertaining to test of parallel lines, the proportional odds assumption was not satisfied for the sewer pipes considered in this study. By further investigation, it was observed that upon removal of the average flow in pipe (%full) from the model, the Goodness-of-Fit test as well as likelihood ratio test available within model fitting information table were still satisfied and additionally, based on the results obtained from test of parallel lines, it was observed that the proportional odds assumption was also satisfied.

By considering the initial condition grading of sewer pipes, it is observed that by using the full model and without the need to eliminate any of the independent variables of the model (unlike the case wherein the initial condition states are not accounted for), not only were the Goodness-of-Fit test as well as likelihood ratio test available within model fitting information table satisfied, but also test of parallel lines illustrated that the proportional odds assumption was satisfied too.

6.6.2 Link functions: Probit

Based on the obtained results, it was observed that ordinal regression utilizing Probit link function satisfied the Goodness-of-Fit test as well as likelihood ratio test available within model fitting information table. Furthermore, it was observed that based on the outcome of test of parallel lines, the proportional odds assumption was satisfied as well.

When initial condition states of sewer pipes were used with Probit link function, it was observed that similar to Logit link function wherein the initial condition states of sewer pipes were not accounted for, even though the developed model through Probit link function satisfied both Goodness-of-Fit test and likelihood ratio test available within model fitting information table, nevertheless, based on test of parallel lines, it was concluded that the proportional odds assumption was not satisfied. It was, however, observed that when the average flow in pipe (%full) was removed from the model, besides satisfying the Goodness-of-Fit test and likelihood ratio test available within model fitting information table, and based upon the test of parallel lines, the proportional odds assumption was satisfied as well.

6.6.3 Link function: Negative Log-Log

Even though the model developed based on ordinal regression with Negative Log-Log link function did satisfy Goodness-of-Fit test as well as likelihood ratio test available within model fitting information table, test of parallel lines did not yield desirable result and it was realized that proportional odds assumption is not validated. However when the average flow in pipe (%full) and the average velocity of flow in sewer pipe were removed, it was observed that the new model satisfied the Goodness-of-Fit test (Pearson and Deviance), the likelihood ratio test available within model fitting information table, and test of parallel lines. Therefore, proportional odds assumption is validated.

When the initial condition states of sewer pipes were taken into account, the results based on Negative Log-Log link function illustrated that the Goodness-of-Fit test results as well as likelihood ratio test available within model fitting information table were satisfied. Based on the test of parallel lines it was realized that the proportional odds assumption was satisfied as well.

6.6.4 Link functions: Complementary Log-Log

By using ordinal regression with Complementary Log-Log link function, it was observed that only one the Goodness-of-Fit test results was satisfactory (i.e. deviance). Furthermore, likelihood ratio test available within model fitting information table was desirable as well. Additionally, the proportional odds assumption was found to be satisfied through test of parallel lines.

Similarly, when the initial condition grading of sewer pipes were taken into account, the results based on Complementary Log-Log link function illustrated that only one the Goodness-of-Fit test results was satisfactory (i.e. deviance) and likelihood ratio test available within model fitting information table was also satisfied. The outcome of test of parallel lines showed that the proportional odds assumption was satisfied as well.

6.6.5 Link functions: Cauchit

Ordinal regression utilizing Cauchit link function did not yield satisfactory results through test of parallel lines, and elimination of independent variables (similar to the procedure described for Negative Log-Log and Logit link functions) did not help with this test either; therefore, the proportional odds assumption was not satisfied with Cauchit link function. However, it was observed that the Goodness-of-Fit test result as well as likelihood ratio test available within model fitting information table yielded satisfactory results. Even though the proportional odds assumption was not satisfied, however, for purposes of illustration and comparison the results of this method was also discussed herein.

Similarly, by considering the initial condition states of sewer pipes, the same observation was made for Cauchit link function. Even though the Goodness-of-Fit test result as well as likelihood ratio test available within model fitting information table yielded satisfactory results, however, based on the test of parallel lines, it was observed that the proportional odds assumption was not satisfied.

6.7 Summary of research

In this dissertation an effort has been made to investigate the deterioration of sewer pipes utilizing various artificial intelligence-based as well as statistical models. Furthermore, by assuming failure criteria, as specified in the dissertation, service lives associated with sewer pipes can be estimated based on the aforementioned deterioration models. However, it should be noted that for various sewer pipes and based on the availability of suitable data, and due to different failure modes that could occur in different sewer pipes, the results will be subjected to uncertainties and variation. In other words, the estimated service lives and deterioration curves could change for different sewer pipes and based on the decision-makers' priorities and failure criteria as well as the available data. Therefore, for various projects and sewer pipes, the suitable modeling approach may differ; i.e. a model which yields suitable results for one project may not necessarily yield suitable and reliable results in another project. This stems from the assumptions and uncertainties associated with the aforementioned approaches.

References

Aamondt, A. & Plaza, E. (1994). "Case Based Reasonings: Foundational Issues, Methodological Variations and System Approaches", *AICOM*, Vol. 7, No. 1, pp. 39-59.

Ablin, R. L., and Kinshella, P. (2004). "Dude, where's my pipe – accelerated corrosion rate threatens Phoenix sewers." *Conf. Proc., Pipeline Engineering and Construction: What's on the Horizon?,* ASCE, Reston, VA., 1 – 8.

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. (2002). *Categorical data analysis,* 2nd Ed., Wiley, Hoboken, New Jersey.

Aldrich, H. J. & Nelson, D. F. (1990). *Linear Probability, Logit, and Probit Models,* John Wiley, New York.

Anderson, D., and Cullen, N. (1982). "Sewer failures 1981, the full year." *WRC External Report No73E.*

Appel, R., Fuchs, T.J., Dollár, P., and Perona, P. (2013). Quickly boosting decision trees-pruning underachieving features early. In ICML (3), pages 594–602.

Ariaratnam, S., El-Assaly, A., and Yang, Y. (2001). "Assessment of infrastructure inspection needs using logistic models." *J. Infrastruct. Syst.*, 7(4), 160 -165.

ASCE, (1994). *Existing Sewer Evaluation and Rehabilitation,* ASCE and Water Environment Federation, New York.

ASCE. (2017). Report Card for America's Infrastructure. ASCE, Reston, VA, USA. Retrieved January 2020 from https://www.infrastructurereportcard.org

Baik, H. (2003) "Development of an asset valuation model for waste water infrastructure assets." Ph.D. thesis, School of Civil Engineering, Purdue University.

Baik, H. S., Jeong, H. S. & 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, ASCE*, Vol. 132, No. 1, pp. 15-24.

Balmer, R., and Meers, K. (1981). “The sewer renewals project of the severn trent water authority.” *Public Health Engineer*, 9(1).

Balmer, R., and Meers, K. G., (1982). “Surveying severn trent’s sewers.” Restoration of sewerage systems, Thomas Telford, London.

Ben-Akiva, M. and Lerman, S.R. (1985). Discrete choice analysis: theory and application to travel demand. MIT Press. Cambridge, Mass.

Ben-Akiva, M. and Ramaswamy, R. (1993). “An approach for predicting latent infrastructure facility deterioration.” Transportation Science, Operations Research Society of America, 27(2), 174-193.

Boden, J. B., Nath, P., Trott, J. J., and Farrar, D. M. (1975). “Research on underground pipelines at the transport and road research laboratory.” *Symposium on R&D sewerage and drainage design, Institution of Public Health Engineers*.

Bottou, L. and Cun, Y. L. (2004). Large scale online learning. In Advances in neural information processing systems, pages 217–224.

Burges, C. JC. (2010). From ranknet to lambdarank to lambdamart: An overview. Learning, 11(23-581):81.

Butt, A. A., Feighan, K. J. and Shahin, M. Y. (1987). “Pavement performance prediction model using the Markov process.” Transportation Research Record 1123, TRB, National Research Council, Washington, D.C., 12-19.

Carnahan, J.V., Davis, W.J., Shahin, M.Y., Keane, P.L., and Wu, M.I. (1987). “Optimal maintenance decisions for pavement management.” Journal of transportation engineering, ASCE, 113(5), pp 554-572.

Cemagref, (2003). *Degradation Models for Drain and Sewer Pipelines Parameterised with CCTV Inspection Data,* Internal Report, Bordeaux.

Cestnik, B. et al. (1990). Estimating probabilities: a crucial task in machine learning. In ECAI, volume 90, pages 147–149.

Chapelle, O., Manavoglu, E., and Rosales, R. (2015). Simple and scalable response prediction for display advertising. ACM Transactions on Intelligent Systems and Technology (TIST), 5(4):61.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785–794.

Cowe Falls, L., Haas, R., McNeil, S., and Tighe, S. (2001). “Asset management and pavement management using common elements to maximize overall benefits.” *Transportation Research Record 1769,* TRB, National Research Council, Washington, D.C., 1-10.

Cramer, D. (1994). *Introducing statistics for social research: step-by-step calculations and computer techniques using SPSS*, Routledge, London; New York.

Dasu, T. & Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning,* John Wiley and Sons Inc., New York.

Davies, J. P., Clarke, B. A., Whiter, J. T., and Cunningham, R. J. (2001a). "Factors Influencing the Structural Deterioration and Collapse of Rigid Sewer Pipes", *Urban Water*, Vol. 3, pp. 73-89.

Davies, J. P., Clarke, B. A., Whiter, J. T., Cunningham, R. J., and Leidi, A. (2001b). “The structural condition of rigid sewer pipes: a statistical investigation.” *Urban Water*, 3(4), 277-286.

Dubout, C. and Fleuret, F. (2011). Boosting with maximum adaptive sampling. In Advances in Neural Information Processing Systems, pages 1332–1340.

EPA. (2004). “Report to Congress; Impacts and Control of CSOs and SSOs.”, United States Environmental Protection Agency, Office of Water, Washington, DC, August 2004

EPA. (2009), “Condition Assessment of Wastewater Collection Systems.”, United States Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory.

EPA. (2015). “Condition Assessment of Underground Pipes”, U.S. EPA, Water Infrastructure Outreach, April 2015.

Fenner, R. A., and Sweeting, L. (1999). “A decision support model for the rehabilitation of noncritical sewers.” *Water Science & Technology*, 39(9), 193 – 200

Fenner, R. A., Sweeting, L., and Marriott, M. (2000). “A new approach for directing pro-active sewer maintenance.” *Proceedings of the institution of civil engineers, water and maritime Engineering Journal*, 142(2), 67 – 78

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. The annals of statistics, 28(2):337–407.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, pages 1189–1232.

Galletti, G. C., Bocchini, P., Pinelli, F. & Pozzi, R. (2003). "Rapid Quality Checking of Polyethylene Water/Gas-Pipelines: A Model Based on Pyrolysis-Gas Chromatography/Mass Spectrometry and Discriminant Analysis", *Journal of Analytical and Applied Pyrolysis*, Vol. 70, No. 2, pp. 291-302.

Gravetter, F., J., and Wallnau, L. B. (2004). *Statistics for the behavioral sciences*, 6th Ed., Thomson/Wadsworth, Australia; Belmont, CA.

Greene, W.H. (2003). Econometric analysis. 5th Edition. Prentice Hall. Upper Saddle River, NJ.

Grigg, N. S. (2003). *Water, wastewater, and stormwater infrastructure management,* Lewis Publishers, Boca Raton, Florida.

Hahn, M. A., Palmer, R. N., Merrill, M. S. & Lukas, A. B. (2002). "Expert System for Prioritizing the Inspection of Sewers: Knowledge Base Formulation and Evaluation", *Journal of Water Resources Planning and Management, ASCE*, Vol. 128, No. 2, pp. 121-129.

Henley, J. E. & Kumamoto, H. (1992). *Probabilistic Risk Assessment,* IEEE Press, Tokyo.

Hopke, P. K., Lamb, R. E. & Natusch, D. F. S. (1980). "Multi-Elemental Characterization of Urban Roadway Dust", *Environmental Science and Technology*, Vol. 14, No. 2, pp. 164-172.

Huberty, C. J. (1994). *Applied Discriminant Analysis,* John Wiley & Sons, New York.

Hudson, W. R., Haas, R., and Uddin, W. (1997). *Infrastructure management: integrating design, construction, maintenance, rehabilitation, and renovation*, McGraw-Hill, New York. International Society for Trenchless Technology (ISTT). (1990). *Trenchless technology guidelines*, ISTT, London, UK.

Jiang Y. , Saito, M., and Sinha, K.C. (1988). “Bridge performance prediction modeling using the Markov chain.” Transportation Research Record 1180, TRB, pp 25-32.

Jiang, Y. and Sinha, K. C. (1989). “Bridge service life prediction model using the Markov chain.” *Transportation Research Record 1223,* TRB, National Research Council, Washington, D.C., 24-30.

Jimenez, L. O. and Landgrebe, D. A. (1999). Hyperspectral data analysis and supervised feature reduction via projection pursuit. IEEE Transactions on Geoscience and Remote Sensing, 37(6):2653–2667.

Jin, R. and Agrawal, G. (2003). Communication and memory efficient parallel decision tree construction. In Proceedings of the 2003 SIAM International Conference on Data Mining, pages 119–129.

Johnson, R. A. & Wichern, D. W. (2002). *Applied Multivariate Analysis,* Prentice Hall, New York.

Johnson, V. E. & Albert, J. H. (1999). *Ordinal Data Modeling,* New York, NY, Springer-Verlag.

Jolliffe, I. (2002). Principal component analysis. Wiley Online Library.

Kathula, V. S. (2001). *Structural Distress Condition Modeling for Sanitary Sewers*, PhD Thesis, Civil of Engineering, Louisiana Tech University, pp.293.

Kawabata, T., Uchida, K. , Ariyoshi, M., Nakase, H., Mohri, Y., and Ling, H.I (2003): "D.E.M. Analyses on behavior of shallowly buried pipe subject to traffic loads", *New Pipeline Technologies, Security, and Safety*, *ASCE*, 1218-1227.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, pp. 3149–3157.

Kearns, M., and Valiant, L. (1994). Cryptographic limitations on learning Boolean formulae and finite automata. Journal of the ACM (JACM), 41(1):67–95.

Kleiner, Y. & Rajani, B. (2001). "Comprehensive Review of Structural Deterioration of Water Mains: Statistical Models", *Urban Water Journal*, Vol. 3, pp. 131-150.

Kleiner, Y. (2001). "Scheduling Inspection and Renewal of Large Infrastructure Assets", *Journal of Infrastructure Systems, ASCE*, Vol. 7, No. 4, pp. 136-143.

Kleiner, Y., Rajani, B. & Wang, S. (2007). "Consideration of Static and Dynamic Effects to Plan Water Main Renewal", *Middle East Water 2007, 4th International Exhibition and Conference for Water Technology*, Manama, Bahrain.

Koksal, A. and Arditi, D. (2004). "Predicting construction company decline." *Journal of Construction Engineering and Management*, 130(6), 799 – 807.

Koo, D.-H. & Ariaratnam, S. T. (2006). "Innovative Method for Assessment of Underground Sewer Pipe Condition", *Automation in Construction*, Vol. 15, pp. 479-488.

Kuzin, S. A. & Adams, B. J. (2005). "Probabilistic Approach to the Estimation of Urban Stormwater Pollution Loads on Receiving Waters", *World Water Congress 2005*, ASCE, pp.143-156.

Laitinen, E. K. (2007). "Classification Accuracy and Correlation: Failure Prediction", *European Journal of Operational Research*, Vol. 183, pp. 210-225.

Langford, J., Li, L., and Zhang, T. (2009). Sparse online learning via truncated gradient. Journal of Machine Learning Research, 10(Mar):777–801.

Lee, Y. and Chang, L. (2003). "Econometric model for predicting deterioration of bridge deck expansion joints." Proceedings o f 9th International Bridge Management Conference, Transportation Research Board, April 28-30, Orlando, Florida, 255-265.

Lei, L., Luo X., and Jiang, X. (2009). "Access mode choice model for highway passenger station." *Proc., International Conference on Transportation Engineering ICTE*, ASCE/T&DI, Chengdu, China, pp 1939 – 1944.

Lester, J. and Farrar, D. M. (1979). "An examination of the defects observed in 6 km of sewers "*TRRL Supplementary Report 531*.

Leung, P. & Tran, L. T. (2000). "Predicting Shrimp Disease Occurrence: Artificial Neural Networks Vs. Logistic Regression", *Aquaculture*, Vol. 187, No. 1-2, pp. 35-49.

Li, D., Xiao, M. & Zeng, Q. (2003). "Impact of Deep Excavations on Adjacent Buried Pipelines", *New Pipeline Technologies, Security, and Safety*, *ASCE*, 1116-1125

Li, M. (2003). "Construction Methods for the Installation of Transmission Pipelines in Western China", *New Pipeline Technologies, Security, and Safety*, *ASCE*, 1080-1085

Li, P. (2012). Robust logitboost and adaptive base class (abc) logitboost. arXiv preprint arXiv:1203.3491.

Li, P., Burges, C. JC., Wu, Q., Platt, JC., Koller, D., Singer, Y., and Roweis, S. (2007). Mcrank: Learning to rank using multiple classification and gradient boosting. In NIPS, volume 7, pages 845–852.

Li, W., Zhang, X., and Papacostas, C. S. (2008). "Analysis of parking fee effect on travel behavior in a downtown district." *Proc., Transportation and Development Innovative Best Practices*, ASCE/T&DI, Reston, VA., pp 69 -74.

Ling, X., Deng, W., Gu, C., Zhou, H., Li, C., and Sun, F. (2017). Model ensemble for click prediction in bing search ads. In Proceedings of the 26th International Conference on World Wide Web Companion, pages 689–698. International World Wide Web Conferences Steering Committee.

Madanat, S. & Ibrahim, W. H. W. (1995). "Poisson Regression Models of Infrastructure Transition Probabilities", *Journal of Transportation Engineering, ASCE*, Vol. 121, No. 3, pp. 267-272.

Madanat, S. M., Karlaftis, M. G. & McCarthy, P. S. (1997). "Probabilistic Infrastructure Deterioration Models with Panel Data", *Journal of Infrastructure Systems, ASCE*, Vol. 3, No. 1, pp. 4-9.

Madanat, S., Mishalani, R. & Ibrahim, W. H. W. (1995). "Estimation of Infrastructure Transition Probabilities from Condition Rating Data", *Journal of Infrastructure Systems, ASCE*, Vol. 1, No. 2, pp. 120-125.

Makar, J. M. & Kleiner, Y. (2000). "Maintaining Water Pipeline Integrity", *AWWA Infrastructure Conference and Exhibition*, Baltimore, Maryland, pp.78-93.

Mason, L., Baxter, J., Bartlett, P. L., and Frean, M. R. (2000). Boosting algorithms as gradient descent. In Advances in neural information processing systems, pages 512–518.

McCullagh, P. (1980). "Regression models for ordinal data." *Journal of the Royal Statistics Society, Series B (Methodological)*, 42(2), 109 – 142.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, 2nd Ed., Chapman and Hall, London; New York

McDonald, S. E. & Zhao, J. Q. (2001). "Condition Assessment and Rehabilitation of Large Sewers", *International Conference on Undreground Infrastructure Research*, Waterloo, Canada, pp.361-369.

Mehta, M., Agrawal, R., and Rissanen, J. (1996). Sliq: A fast scalable classifier for data mining. In International Conference on Extending Database Technology, pages 18–32. Springer.

Menard, S. (2002). "Applied logistic regression analysis." *Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-106*, Thousand Oaks, CA.

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

Micci-Barreca, D. (2001). A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. ACM SIGKDD Explorations Newsletter, 3(1):27–32.

Micevski, T., Kuczera, G. & Coombes, P. (2002). "Markov Model for Storm Water Pipe Deterioration", *Journal of Infrastructure Systems, ASCE*, Vol. 8, No. 2, pp. 49-56.

Mishalani, R. G. & Madanat, S. M. (2002). "Computation of Infrastructure Transition Probabilities Using Stochastic Duration Models", *Journal of Infrastructure Systems, ASCE*, Vol. 8, No. 4, pp. 139-148.

Morcous, G., Rivard, H. & Hanna, A. M. (2002a). "Case-Based Reasoning System for Modeling Infrastructure Deterioration", *Journal of Computing in Civil Engineering, ASCE*, Vol. 16, No. 2, pp. 104-114.

Morcous, G., Rivard, H. & Hanna, A. M. (2002b). "Modeling Bridge Deterioration Using Case-Based Reasoning", *Journal of Infrastructure Systems, ASCE*, Vol. 8, No. 3, pp. 86-95.

Morrison, R. S. & Thomson, J. C. (2003). "Innovative Inspection Methodologies for Wastewater Systems", *New Pipeline Technologies, Security, and Safety*, *ASCE*, 1563-1579

Moselhi, O. & Shehab-Eldeen, T. (2000). "Classification of Defects in Sewer Pipes Using Neural Networks", *Journal of Infrastructure Systems, ASCE*, Vol. 6, No. 3, pp. 97- 104.

Moser, P. A. (2001). *Buried Pipe Design,* McGraw-Hill, New York.

NASSCO. (2015). "Pipeline Assessment Certification Manual", National Association of Sewer Service Companies, Marriottsville, MD, USA.

Norusis, M. J., (2008). *SPSS statistics 17.0: Advanced statistical procedures companion,* Prentice Hall, Upper Saddle River, New Jersey.

O'Reilly, M. P., Rosbrook, R. B., Cox, G. C., and McCloskey, A. (1989). "Analysis of defects in 180 km of pipe sewers in southern water authority." *TRRL Research Report 172*.

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

Park, T. (2009). A Comprehensive Asset Management System for Sewer Infrastructures, Civil and Environmental Engineering. State College, PA, Pennsylvania State University. Ph.D.

Penn, M, Vargas, F. and Chimba, D. (2008). “Multinomial modeling of purpose driven trip.” *Proc., 2007 Transportation, Land Use Planning, and Air Quality Congress*, ASCE/T&DI, Reston, VA., pp. 64-78.

Pohls, O. (2001). *The Analysis of Tree Root Blockages in Sewer Lines & Their Prevention Methods*, Master Thesis, Institute of Land and Food Resources, University of Melbourne.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). CatBoost: Unbiased Boosting with Categorical Features. In Advances in Neural Information Processing Systems. 6639-6649

Ranka, S. and Singh, V. (1998). Clouds: A decision tree classifier for large datasets. In Proceedings of the 4th Knowledge Discovery and Data Mining Conference, pages 2–8.

Ratliff, A. (2003). "An Overview of Current and Developing Technologies for Pipe Condition Assessment", *New Pipeline Technologies, Security, and Safety*, *ASCE*, 103-114

Rauch, W., and Stegner, T. (1994). “The colmation of leaks in sewers systems during dry weather flow.” *Water Science & Technology*, 30(1), 205 – 210.

Read, G. F. and Vickridge, I. (1995). *Sewers: rehabilitation and new construction repair and renovation*, Arnold, London; New York.

Richardson, M., Dominowska, E., and Ragno, R. (2007). Predicting clicks: estimating the click-through rate for new ads. In Proceedings of the 16th international conference on World Wide Web, pages 521–530. ACM.

Riesbeck, C. K. & Schank, R. C. (1989). *Inside Case-Based Reasoning,* Lawrence Earlbaum Associates, New Jersey.

Roe, B. P., Yang, H.-J., Zhu, J., Liu, Y., Stancu, I., and McGregor, G. (2005). Boosted decision trees as an alternative to artificial neural networks for particle identification. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 543(2):577-584.

Ross, S. M. (2000). Introduction to probability models, 7th Ed., Academic, San Diego.

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

Seo, S., Aramaki, T., Hwang, Y. & Hanaki, K. (2004). "Fuzzy Decision-Making Tool for Environmental Sustainable Buildings", *Journal of Construction Engineering and Management, ASCE*, Vol. 130, No. 3, pp. 415-423.

Shafer, J., Agrawal, R., and Mehta, M. (1996). “Sprint: A scalable parallel classifier for data mining”, *Proc. of the 22nd International Conference on Very Large Data Bases*, pages 544–555.

Shan, Y., Zhao, R., Xu, G., Liebich, H. M. & Zhang, Y. (2002). "Application of Probabilistic Neural Network in the Clinical Diagnosis of Cancers Based on Clinical Chemistry Data", *Analytica Chimica Acta*, Vol. 471, No. 1, pp. 77-86.

Sikora, E. J. (1979). “Load factors and non-destructive testing of clay pipe.” *Journal Water Pollution Control Federation*, 52(12), 2964 - 2976.

Singh, D. & Tiong, R. L. K. (2005). "A Fuzzy Decision Framework for Contractor Selection", *Journal of Construction Engineering and Management, ASCE*, Vol. 131, No. 1, pp. 62-70.

Soulie, F. F. & Gallinari , P. (1998). *Industrial Applications of Neural Networks,* World Scientific, Singapore.

Sutherland, R. A. & Tolosa, C. A. (2000). "Multi-Element Analysis of Road Deposited Sediment in an Urban Drainage Basin," *Environmental Pollution*, Honolulu, Hawaii, pp.483-495.

Tabachnick, B. G. & Fidell, L. S. (2001). *Using Multivariate Statistics,* Allyn and Bacon, New York.

Tai, Y.-L. (1991). *Physical and Chemical Characterization of Street Dust and Dirt from Urban Areas*, Pennsylvania State University, U.S.

Taylor, G. J. (1993). *The Promise of Neural Networks,* Springer-Verlag.

Taylor, G. J. (1996). *Neural Networks and Their Applications,* John Wiley and Sons, London.

Terano, T., Asai, K. & Sugeno, M. (1991). *Fuzzy Systems Theory and Its Applications,* Academic Press, Tokyo.

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

Tran, H. D., Perera, B. J. C., and Ng, A. W. M. (2009). "Predicting structural deterioration condition of individual storm-water pipes using probabilistic neural networks and multiple logistic regression models." *Journal of Water Resources Planning and Management*, 135(6), 553-557.

Tsai, C.-Y. (2006). "On Detecting Nonlinear Patterns in Discriminant Problems", *Information Sciences*, Vol. 176, No. 7, pp. 772-798.

US Environmental Protection Agency (USEPA) and American Society of Civil Engineers (ASCE). (1999). "Optimization of collection system maintenance frequencies and system performance". *USEPA Office of Water, Washington. DC, February, EPA Cooperative Agreement No. CX824902- 01- 0.*

USEPA. (2002). "Fact sheet: Asset management for sewer collection systems." http://www.epa.gov/npdes/pubs/assetmanagement.pdf

Washington, S.P., Karlaftis, M.G., and Mannering, F.L. (2003). Statistical and econometric methods for transportation data analysis, Chapmand & Hall/CRC, Boca Raton, Florida.

Water Environment Federation (WEF) and American Society of Civil Engineers (ASCE). (1992). *Design and Construction of Urban Stormwater Management Systems*, ASCE, New York, NY; WEF, Alexandria, VA.

Water Research Center (WRC). (2001). *Sewerage rehabilitation manual,* Wiltshire, U.K

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.

Wilmot, C. G. & Mei, B. (2005). "Neural Network Modeling of Highway Construction Costs", *Journal of Construction Engineering and Management, ASCE*, Vol. 131, No. 7, pp. 765-771.

Windle, R., and Dresner, M. (1995). “Airport choice in multi-airport regions.”*Journal of Transportation Engineering*, 121(4), 332 – 337.

Winston, W.L. (1994). Operations research applications and algorithms. 3nd Edition. Duxbury Press. Belmont, CA.

Wirahadikusumah, R., Abraham, M. D. & Iseley, T. (2001). "Challenging Issues in Modeling Deterioration of Combined Sewers", *Journal of Infrastructure Systems, ASCE*, Vol. 7, No. 2, pp. 77-84.

Wirahadikusumah, R., Abraham, M. D., Iseley, T. & Prasanth, R. K. (1998). "Assessment Technologies for Sewer System Rehabilitation", *Automation in Construction*, Vol. 7, pp. 259-270.

WRC, (1983). *Sewerage Rehabilitation Manual,* Water Research Center, London, U.K.

WRC, (1986). *Sewerage Rehabilitation Manual,* Water Research Center, London, U.K.

Wright, L., and Dent, S. (2007). “Risk management for planning and decision making of pipeline projects.” *Proc., Pipelines 2007 International Conference: Advances and Experiences with Trenchless Pipeline Projects*, ASCE, Reston, VA., 1 – 14.

WSAA, (2002). *Sewer Inspection Reporting Code of Australia,* Water Service Association of Australia (WSAA), Melbourne.

WSAA, (2006). *Conduit Inspection Reporting Code of Australia,* Water Service Association of Australia (WSAA), Melbourne.

Wu, Q., Burges, C. J., Svore, K. M., and Gao, J. (2010). Adapting boosting for information retrieval measures. Information Retrieval, 13(3):254–270.

Yang, Z. R., Platt, M. B. & Platt, H. D. (1999). "Probabilistic Neural Networks in Bankruptcy Prediction", *Journal of Business Research*, Vol. 44, No. 2, pp. 67-74.

Yoshizaki, K. & Sakanoue, T. (2003). "Experimental Study on Soil-Pipeline Interaction Using Eps Backfill", *New Pipeline Technologies, Security, and Safety*, *ASCE*, 1126-1134.

Young, O. C. & O'Reilly, M. P. (1983). *A Guide to Design Loading for Buried Pipes,* Prentice Hall, London.

Zhang, Y. and Haghani, A. (2015). A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 58:308-324.

Zhou, Z.H. (2012). Ensemble methods: foundations and algorithms. CRC press.