REMAINING USEFUL LIFE PREDICTION OF WATER PIPES USING ARTIFICIAL NEURAL NETWORK AND ADAPTIVE NEURO FUZZY INFERENCE SYSTEM MODELS

by

RAZIEH TAVAKOLI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2018

Acknowledgements

I would like to acknowledge chair of my committee, Dr. Mohammad Najafi, P.E., F. ASCE, and Director of the Center for Underground Infrastructure Research and Education (CUIRE) and Professor at the University of Texas at Arlington. Dr. Najafi has always been a leader, a great motivator and was always supportive throughout all phases of my research. It would have been impossible for me to complete my research without his guidance, expertise and knowledge. His insight has significantly assisted me in choosing my career path, presently and in the future.

I would like to thank my dissertation committee members, Dr. Sharareh Kermanshachi, Dr. Mostafa Ghandehari and Dr. Edmund Prater for their suggestions on improving my dissertation and for taking the time out from their extremely busy schedules for my dissertation. I would also like express my appreciation to my family for their love and support throughout my research.

March 29, 2018

Abstract

REMAINING USEFUL LIFE PREDICTION OF WATER PIPES USING ARTIFICIAL NEURAL NETWORK AND ADAPTIVE NEURO FUZZY INFERENCE SYSTEM MODELS

Razieh Tavakoli, PhD
The University of Texas at Arlington, 2018

Supervising Professor: Mohammad Najafi

The U.S. water distribution system contains thousands of miles of pipes with differing materials, sizes, and ages. These pipes experience physical, environmental, structural and operational parameters that cause corrosion and eventually lead to their failures. The Remaining Useful Life (RUL) is the estimated time before a pipe will experience a failure mode specifically a pipe break. Pipe failure means collapse and deterioration of water pipes overtime. Pipe deterioration results in increased break rates, reduced hydraulic capacity, and detrimental impacts on water quality. Therefore, it is crucial to perform accurate models that can forecast deterioration rates along with estimates of remaining useful life of pipelines to implement essential interference plans that can reduce catastrophic failures. This dissertation discusses a computational model that forecasts the RUL of water pipelines using Artificial Neural Network (ANN) and Adaptive Neural Fuzzy Inference System (ANFIS). Artificial Neural Network and ANFIS are developed using Levenberg-Marquardt backpropagation algorithm and mixture of backpropagation and least squares (hybrid method). Those models are trained and tested with acquired field data. The developed models identify the significant parameters that influence prediction of RUL. It is concluded that, on the average, with approximately 10% of wall thickness loss in

existing cast iron, ductile iron, asbestos-cement and steel water pipes analyzed in this dissertation, the reduction of their remaining useful life will be approximately 50%.

Table of Contents

List of Illustrations

List of Tables

Chapter 1 Introduction and Background

1.1 History of Water Distribution System

The most extensive part of water supply system is water distribution system (Fares and Zayed, 2010). Water distribution systems contain different types of buried pipes (i.e., cast iron, ductile iron, asbestos cement, polyethylene, and polyvinyl chloride). As water distribution system becomes older, its structural condition, hydraulic capacity and performance deteriorates. Several factors impact on the structural deterioration of water mains and their failures, including pipe material, pipe size, pipe age, soil type, climate, and cyclic pressures. However, the physical processes that cause pipe breakage are very complicated. Most water pipes are buried, so there is little data available about how they deteriorate and fail (WRF, 2013). Deterioration of water mains is neither identical nor uniform (Al-Barqawi, 2006). This is because water mains are operated under pressure, and usually unreachable (Al-Barqawi, 2006). Therefore, it is essential to inspect and assess water system condition to efficiently maintain and improve its elements to prevent catastrophic failures and emergency repairs.

Although on-site inspection of a pipeline is the perfect method to analyze and understand its condition; this method can be expensive and usually cannot be cost efficiently applied to smaller diameter distribution lines, which make up most of water systems. Risk of water main failure factors can be divided into likelihood and consequence of failures. Likelihood of failures included, environmental, physical, and operational factors that represent the deterioration factors. Consequence or post failure factors represent the cost of failure in which they should be considered when assessing the risk of water main failure.

1.2 Water Main Classification

Water main can be categorized depending on its features by following several approaches such as material, diameter, and function. There are three main types of pipeline materials that are used in the construction of pressurized pipelines. They are cement-based pipes, plastic pipes, and metallic pipes. Cement–based pipes include concrete pipe as well as asbestos-cement pipe. Both types incorporate Portland cement as the base material. Each of this group of pipeline material contains a variety of materials (Al-Barqawi, 2006). Table 1-1 presents the classification of different types of pipe.

Table 1-1 Types of Pipe Based on Material

Adapted from Al-Barqawi and Zayed, 2008

Metallic Pipe	Concrete Pipes	Plastic Pipes
Cast Iron Pipe (CIP)	Prestressed Concrete Cylinder Pipe (PCCP)	Polyvinyl Chloride Pipe (PVC)
Ductile Iron Pipe (DIP)	Reinforced Concrete Pipe (RCP)	Polyethylene Pipe (PE)
Steel Pipe (SP)	Bar-wrapped Steel-cylinder Concrete Pipe (BWP)	Glass Reinforced Polyester Pipe (GRP)
N/A	Asbestos-Cement Pipe (AC)	N/A

1.2.1 Classification by Diameter

Water main can be categorized according to its diameter into three groups: small diameter (2 in. to 8 in.), medium diameter (10 in. to 30 in.), and large diameter (36 in. to 72 in.). Large diameter pipes have more beam strength than small diameter pipes (Al-Barqawi, 2006). Steel, cast iron (CI), ductile iron (DI), reinforced concrete (RC), pre-stressed concrete cylinder pipe (PCCP), and asbestos cement (AC) are used in the construction of large-diameter water mains, while more recently, polyvinyl chloride (PVC) and polyethylene (PE) pipes have been broadly used, particularly in the lower diameter range (Kleiner and Rajani, 2001).

1.2.2 Classification by Function

Water main can be categorized according to its function into two main classes: transmission and distribution lines. The purpose of transmission pipelines is to transfer water from a main source to a storage system (i.e., water tanks). They are considered the most expensive part of the system because of their greater construction costs (i.e., material, installation, equipment). The purpose of distribution lines is to transport water from the storage system to users (i.e., residential buildings or industrial factories). The minimum diameter for a distribution pipe is 2 in., and the minimum diameter required for serving fire hydrants is 6 in. (Al-Barqawi, 2006).

1.3 Factors Influencing Deterioration of Water Mains

Leaks and breaks due to corrosion of water mains can cause the loss of potable water, water contamination, flooding, real estate damage, and service disruptions to end users. They can be a potential danger if they temporarily disable firefighting abilities (Wang, 2009). In many municipalities, there may be pipes more than 100 years old that are still in operation. However, metallic pipes that are in their early ages may require urgent replacement due to various reasons, for example, very corrosive soil. Kleiner and Rajani (2001) have categorized the deterioration of water mains into two classes: structural corrosion that reduces the capability of pipes to bear various stresses, and deterioration of inner surfaces that decreases their hydraulic capacity and destroys water quality. In cases of severe internal corrosion, the corrosion of inner surfaces can cause structural corrosion as well.

According to Wang (2009), researchers have studied many aspects that lead to the deterioration of water pipes, with the purpose of improving predictive planning models.

As there are many reasons that impact on pipe failures and affect maintenance decisions, it is a complicated process to develop such models to evaluate all these factors. The factors impacting on pipe corrosion can be either time-dependent or static. Those aspects that will not change over time are static parameters, such as pipe diameter or pipe material. On the other hand, pipe age, water pressure and temperature, soil corrosivity, weather and water contents of soil, and previous pipe breaks are samples of random and time dependent factors. These parameters can also be classified as follows (Wang, 2009).

1. The quality and age of the pipe, including pipe material, diameter, length of pipe, and type and number of joints.
2. The environment and temperatures where the pipe is laid, such as the soil corrosivity, frost depth, soil temperature, soil moisture content, traffic and other loadings.
3. The quality of workmanship when a pipe is installed, repaired, or rehabilitated, such as construction method (e.g., embedment type).
4. The operating and maintenance aspects of the pipe network, such as pressure and water hammer, pipe breaks and repair records and details, preventative maintenance (inspection, lining, cleaning, and cathodic protection for metallic pipes).

1.3.1 Failure of Water Mains

Pipes will deteriorate and collapse overtime, but the rate of failure in pipes differs according to pipe's material and exposure to various environmental and operational conditions. Corrosion of pipes will impact on the structural condition and hydraulic capacity of a water main, which reduces the system function (Rajani and Kleiner, 2004). Metallic pipes may deteriorate faster than plastic pipes if they are buried in aggressive soil without proper protection. For example, if metallic pipes are buried in aggressive soil, they should be covered in plastic sheets (polyethylene encasement) to separate the metal from soil, and therefore, prevent pipe corrosion. However, the useful life of the polyethylene sheet is 30 years (Al-Barqawi and Zayed, 2008). Similarly, corrosion is the main cause of failure in Pre-Stressed Concrete Cylinder Pipes (PCCP). When sufficient number of the pre-stressed bars or wires are corroded and destroyed in a section of the PCCP, concrete in that section will not hold pressure. Consequently, the pipe will break due to internal pressure (Al-Barqawi and Zayed, 2008). The failure modes of different types of pipe will explain in following sections.

Water main breaks, because of enforced operational and environmental stresses on a structurally deteriorated pipe due to corrosion, degradation, insufficient installation, and/or manufacturer faults. Pipe break types are categorized into four groups: (1) circumferential breaks, caused by bending stresses; (2) longitudinal breaks, affected by transverse stresses (hoop stress); (3) split bell, affected by transverse stresses on the pipe joint; (4) holes due to corrosion (Nemeth, 2016). Table 1-2 categorizes water main corrosion into three groups. Figure 1-1 graphically represents pipe failures that happen due to straight tension (top), bending or flexure (middle), and hoop stress (bottom).

Table 1-2 Factors that Cause Water System Corrosion

Adapted from Al-Barqawi, 2006

	Factor	Explanation
Physical	Pipe material	Pipes made from various materials fail in various modes.
	Wall thickness	Thinner walled pipe will be corroded more rapidly.
	Pipe age	Results of pipe degradation become more obvious over time.
	Pipe diameter	Small diameter pipes are more vulnerable to beam failure.
	Type of joints	Some sorts of joints have experienced immature failure (e.g., leadite joints).
	Pipe lining and coating	Lined and coated pipes are less vulnerable to deterioration.
	Pipe installation	Pipes can damage due to poor installation practices, making them susceptible to failure.
	Pipe manufacture	Imperfections in pipe walls created by manufacturing faults can make pipes susceptible to failure.
Environmental	Soil type	Pipe loading changes due to corrosive soils that experience important volume changes in reaction to moisture changes,
	Groundwater	Some groundwater is destructive toward specified pipe materials.
	Pipe location	Rate of corrosion can increase due to movement of road salt into the soil.
	Pipe bedding	Inappropriate bedding may result in early pipe failure
Operational	Transient pressure	Changes to internal water pressure will change stresses acting on the pipe.
	Leakage	Leakage corrodes pipe bedding and increases soil moisture in the pipe zone.
	Water quality	Some water is aggressive, encouraging deterioration
	Flow velocity	Unlined dead-ended mains have greater rate of internal corrosion.
	Operations and maintenance practices	Poor practices can cooperate structural integrity and water quality.

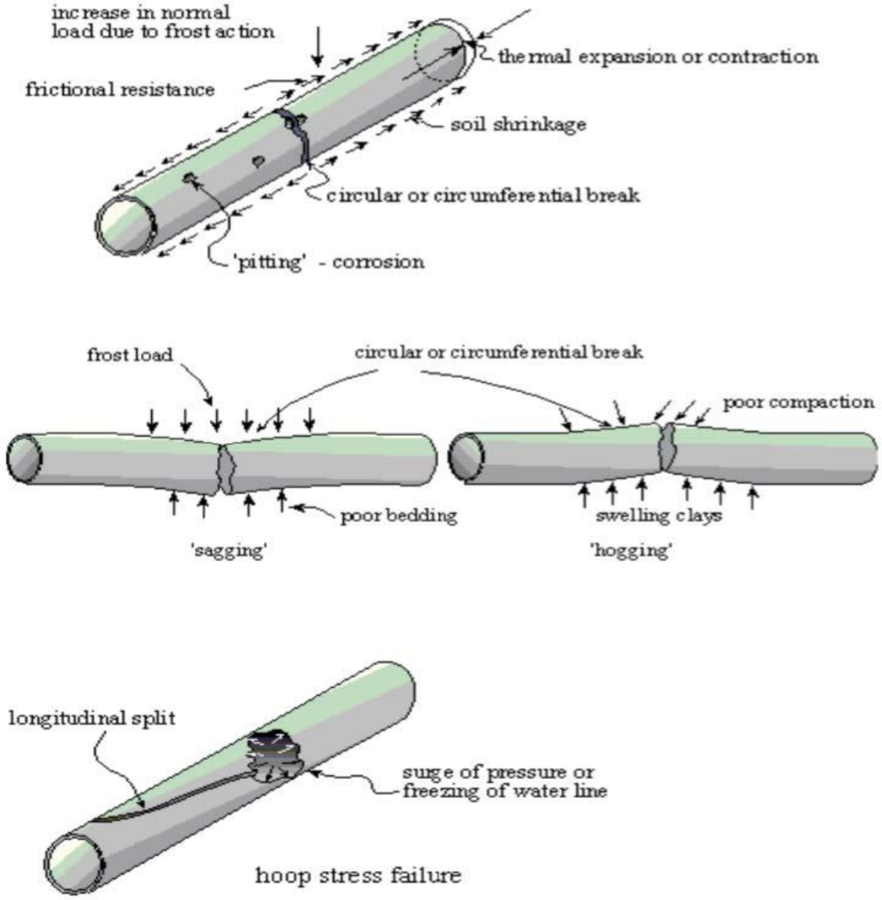


Figure 1-1 Pipe Characteristics and Failure Modes

Rajani and Kleiner, 2004

Different materials have different features and show different behaviors because of certain aspects such as corrosion. In the deterioration analysis of water mains, pipe material is often used as the most significant grouping criteria (Rajani and Kleiner, 2004).

1.3.1.1 Cast Iron Pipes (CIP)

Cast iron pipe is the most common pipe material in the United States, including over 50% of the total US water main network. Some of the earliest cast iron pipe in the U.S. was installed in the 19th century and continued prevalent until the 1970s, when the popularity of ductile iron pipe grew. Cast iron is also mentioned as gray cast iron and is a very strong but fragile material (Najafi and Gokhale, 2005). Pit cast gray iron and centrifugal cast gray iron are two types of cast iron pipe (CIP). One of common failure mode for small diameter cast iron pipes is bell splitting. The failures occur due to the difference between the coefficient of thermal expansion in the joint-sealing compound and the metal in the pipes. In cold temperatures, the leadite expands differently than the cast iron pipe, causing splitting at the bell (Makar, 2001).

1.3.1.2 Ductile Iron Pipes (DIP)

The graphite composition in the CIP was changed from flake form to spherical form by adding inoculants such as magnesium to the molten iron to have an improved pipe material. This process leads to the development of ductile iron pipe (DIP) which has an improved strength, affects resistance and some other properties compared to CIP (Najafi and Gokhale, 2005). The source cause of failure in ductile pipes is extreme forces acting upon the pipe in the forms of internal pressure, bending, soil movement, and thermal expansion due to differences between the temperature of the water pipes and the surrounding soil or the pipes and joint mechanisms (Jenkins, 2014). The failure modes observed among ductile pipes are blowout holes, circumferential cracking, bell splitting, longitudinal cracking, bell shearing, and spiral cracking. Blowout holes are beginning by corrosion pitting which causes wall thinning. Finally, the pressurized water surpasses the strength of the thin pipe wall and a hole is formed. Circumferential cracking, the most common failure mode for pipes less than 14 in., is caused by bending forces or tensile forces due to soil movement (Jenkins, 2014).

1.3.1.3 Polyvinyl Chloride Pipe (PVC)

The 1970s saw a change to the use of Polyvinyl chloride (PVC) pipe because it was cheaper to buy, transport, and install than ductile pipes. Another benefit of PVC is that it does not rust like ferrous pipe. The material composition of PVC makes it more brittle than ductile pipes under certain conditions, but also more disposed to bending and flexure. A study of PVC pipe used for gas distribution showed that manufacturing flaws and installation practices donated the most of PVC pipe failures. With respect to installation practices, pipes left out in the sun too long prior to bedding are subjected to chemical breakdown, degrading the structural integrity of the pipe. An additional cause of pipe failure due to installation of PVC pipe is over insertion at the pipe joint, where the spigot joint is inserted too far into the bell, causing fracture (Jenkins, 2014). There are many different mechanisms for similar failure modes.

Polyvinyl Chloride (PVC/un-plasticized Polyvinyl Chloride (UPVC) pipes have high resistance to deterioration and corrosion, and can be used in very corrosive environments, but they likely to be affected by deterioration if they are exposed to weather, chemical attack, or mechanical degradation from improper installation methods (Al-Barqawi, 2006). The chemical attack resistance for PVC pipes usually reduces with the rise in concentration of a precise chemical. For example, organic chemicals such as solvents and gasoline will weaken PVC/UPVC pipes, resulting in failure of pipe by expansion and rupture.

1.3.1.4 Steel Pipe (SP)

Steel pipes are manufactured in different ways to give them their respective characteristics. Steel pipes can be manufactured using seamless welds, butt-welds or spiral welds. Seamless pipe is formed when a molten steel rod is combined with a clamp. Butt welded steel pipe is formed when hot steel is rolled into a hollow cylinder-like shape giving the pipe a joint. Finally, a spiral welded steel pipe is formed when strips of steel metal are twisted and welded where the edges join each other. Steel pipes are known for their strength and ability to transport water at high pressures (Parisher and Rhea, 2012).

Carbon steel pipes are most commonly used in industry today. The carbon steel material's drawback is the ability to corrode easily via ferrous oxide formation on the inside of the walls, which can sometimes slow down the water flow (Parisher and Rhea, 2012). Use of steel pipe in the U.S. was observed as early as 1863 in San Francisco. Several developments in the technology over the years have made steel more versatile for piping applications (Najafi and Gokhale, 2005).

1.3.1.5 Asbestos Cement (AC) Pipes

Asbestos Cements (AC) pipes can also be weakened and degraded when it is used to transfer aggressive water such as low acidity and low alkalinity waters (Al-Barqawi, 2006). The degradation of pipe will release asbestos fibers, which are harmful to health, and mix it with the carried water through the water distribution system. To prevent this type of damage, the pipe should be lined with epoxy resin or cement mortar (Al-Barqawi, 2006). Due to the harmful effects of asbestos fibers to the human health, its use was discontinued in the early 1980s in North America (Najafi, 2010). However, utilities in U.S. have a small percentage of AC pipe in their distribution network, which are still in service.

1.3.1.6 Reinforced Concrete Pipe (RCP)

These concrete pipes can be reinforced with welded wire fabric, hot-rolled rod made of Grade 40 steel or cold-drawn steel wire made from hot-rolled rods. It can be used for pressure applications up to 55 psi (Najafi and Gokhale, 2005).

1.3.1.7 Prestressed Concrete Cylinder Pipe (PCCP)

This is a composite pipe manufactured using concrete and steel. It is used for high-pressure applications and can handle up to 500 psi. Lined cylinder pipe and embedded cylinder pipe are the two types of PCCP (Najafi and Gokhale, 2005).

1.3.1.8 Polyethylene Pipe (PE)

Polyethylene was discovered in 1933 and its use in pipe applications started in 1950. PE or HDPE pipes have been an alternative for tuberculation and corrosion issues of traditional iron, steel and concrete pipes (Storm and Rasmussen, 2011). HDPE pipes have been used for municipal water applications for almost fifty years, they are still minimally used for potable water transmissions/distributions and wastewater services when compared to other types of pipe (Al-Barqawi, 2006). Similarly, the polyethylene pipes (PE) deteriorate and fail due to joint imperfections, material degradation, and improper pipe installation. In addition, organic chemicals can pass through the walls of the PE pipes (Al - Barqawi, 2006). Table 1-3 shows failure modes and mechanisms for different pipes.

1.3.2 Definition of Remaining Useful Life (RUL)

The RUL is the estimated time before a pipe will experience a catastrophic failure mode (section 1.3.1), specifically a pipe break (Nemeth, 2016). This failure mode may include rupture, breakage, joint separation, longitudinal cracking, circumferential cracking, holes, and other events that make the use of the pipe impossible or impractical. Failure also may include low water pressures, clogging, corrosion, leakage, tuberculation, low flow, water quality issues and similar.

Table 1-3 Failure Modes and Mechanisms

Modified from Jenkins, 2014

Failure Mode		Failure Mechanism	Material
Cracking	Circumferential	Bending moments applied to the pipe and soil movement which, produce tensile forces on pipe.	All
	Longitudinal	Internal water pressure, crushing and compressive forces acting on pipe.	All
	Spiral	Pressure surges and/or combination of bending forces and internal pressure.	All
	Mixed	Combination of stresses.	All
	Ring	Axial tension, bending, traffic load, settlement, uplift, production, fatigue, residual stresses, temperature, and frost.	PVC
	Axial	Internal pressure, bending, traffic load, production, residual stresses, and frost	PVC
	Irregular	Environmental such as chemical, UV, and stress cracking.	PVC
Fracture	Circumferential	Bending moments applied to the pipe and soil movement which produce tensile forces on pipe.	All
	Longitudinal	Internal water pressure, crushing and compressive forces acting on pipe.	All
	Spiral	Pressure surges and/or combination of bending forces and internal pressure.	All
	Mixed	Combination of stresses.	PVC
Buckling	Axial	External pressure, axial compression, temperatures, fire, and interventions.	PVC
	Transverse/ring	External pressure, axial compression, production, residual stresses, temperatures, fire, and interventions.	PVC
	Non-symmetric	Longitudinal bending and brazier effect.	PVC
	Longitudinal	Axial compression and thermal effects.	PVC

1.4 Research Needs

Much research has been conducted on pipe failure assessment and remaining useful life, but no evaluation of significant parameters has been developed that impact on remaining useful life. More condition assessment has been done on wastewater systems than on water systems, therefore there are more uncertainties and unknowns in predicting the possibility of failure (or the remaining useful life) of a water pipeline than a similar wastewater pipeline (WRF, 2013). Furthermore, very few technologies like, Closed Circuit Television (CCTV); ultrasonic inspection, 3D optical scanning, videoscope and etc., have been successful in usefulness of determination of remaining useful life (Agrawal, 2015). Moreover, lack of awareness about the performance and economic parameters of all the commercially available technologies acts as a major drawback in the use of these technologies (Agrawal, 2015).

Additionally, there is also a lack of knowledge about the evaluation parameters for prediction of remaining useful life. There are two approaches used to determine the probability of failure and remaining useful life of underground water pipelines: (1) Statistical approaches based on analysis of available failure records. (2) Physical probabilistic approaches: derived from physical principles of pipeline failure combined with a stochastic representation of input variables. In both of above approaches, various asset parameters are considered in the analysis, such as pipeline diameter, material type, installation year, break history, etc., along with other risk factors such as operating pressure, soil type and soil acidity. Nevertheless, there is no standard procedure for recording data on leaks, breaks, and condition indicators. Therefore, there is a lack of considering wall thickness loss to forecast the remaining useful life. According to previous researches about the water pipes failure and remaining useful life prediction (Manda, 2012; Joshi, 2012; Paradkar, 2012 and Deshmukhi, 2012) there is a lack of a model to evaluate the risk of water pipes failure and determine the remaining useful life for different types of water pipe and there are no researches using combination of Artificial Neural Network (ANN) and Adaptive Neural Fuzzy Inference System (ANFIS) models to predict the remaining useful life of water pipes.

1.5 Research Objectives

The main objective of this dissertation is to develop a model to determine the remaining useful life (RUL) of various types of water pipes based on physical, operational data and inspection datasets. Secondary objectives of this dissertation are:

- To determine the most significant parameters influencing RUL of water pipes using three different methods (statistical analysis, neural network and adaptive neural fuzzy system).
- To evaluate and categorize different condition assessment of water pipes for determination of RUL and to determine the relationship between important parameters and remaining useful life.

For this research soil type, groundwater-table location, operational factors, PVC, and HDPE are not considered. The pipes materials include cast iron, ductile iron, asbestos cement and steel pipes and pipe diameter ranges vary from 4 in. to 24 in.

1.6 Research Methodology

The research methodology employed in this dissertation is based on literature review of the deterioration and condition rating of water mains and the analysis of actual data collected for asbestos-cement pipes, ductile iron pipes, cast iron and steel pipes by following steps:

- Review literature focusing on factors affecting water main deterioration.
- Review collected water main data to analyze condition of the pipes and determining input and output parameters and relationship between them.
- Develop a model to estimate remaining useful life if water pipes using Artificial Neural Network (ANN) and Adaptive Neural Fuzzy Inference System (ANFIS).

Figure 1-2 illustrates the research methodology.

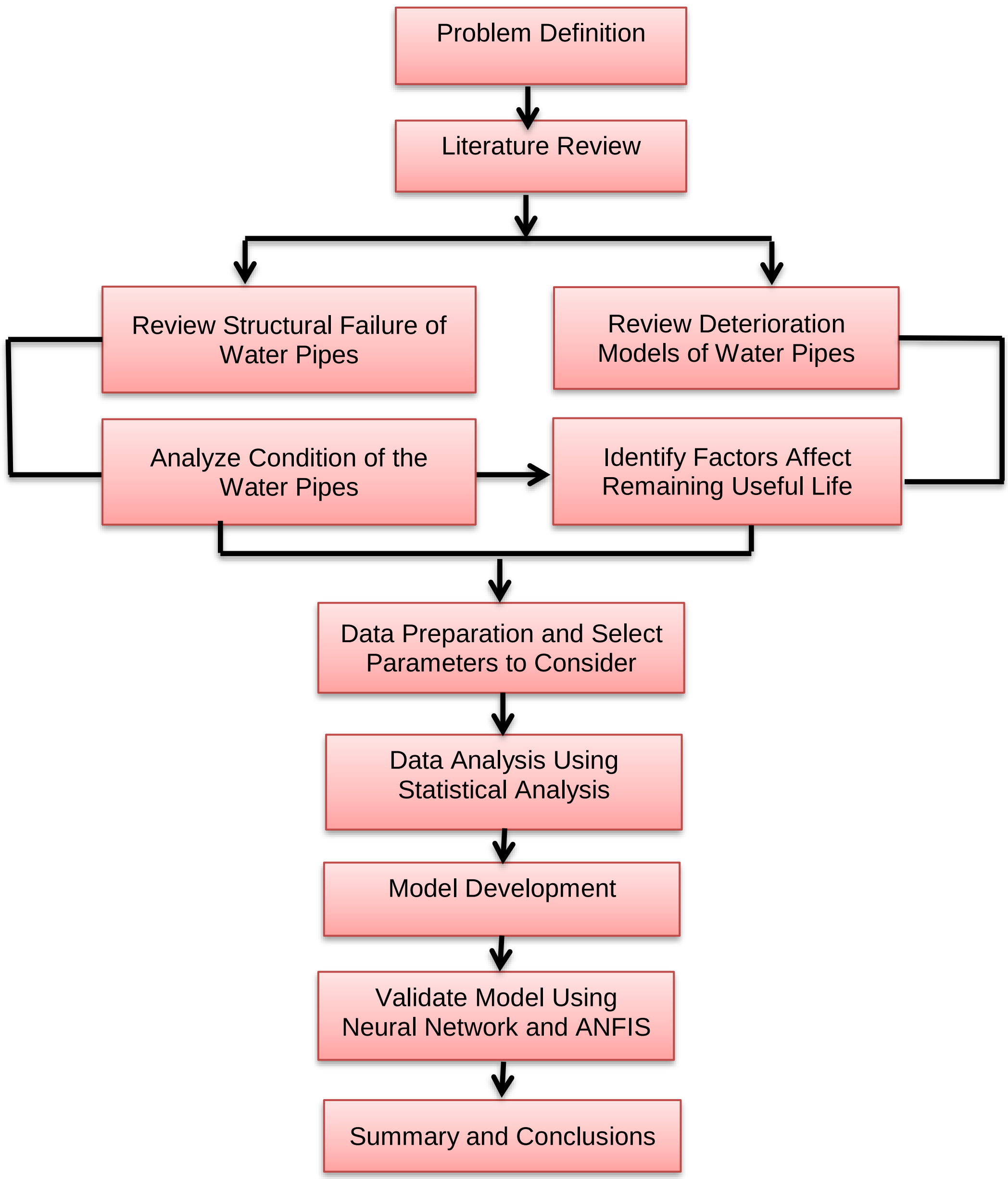


Figure 1-2 Research Methodology

1.7 Hypothesis

With sufficient data, it is expected that ANN and ANFIS models can be utilized to predict remaining useful life (RUL) of water pipes. Condition assessment of water pipes indicates reliability of the ANN model in prediction of remaining useful life. It is expected that pipe age and wall thickness loss are most significant parameters to evaluate RUL. Ductile iron and steel pipes have more remaining useful life compared to cast iron and asbestos cement pipes. It is expected that, on the average, with approximately 10% of wall thickness loss in existing cast iron, ductile iron, asbestos-cement and steel water pipes, the reduction of their remaining useful life will be approximately 50%.

1.8 Contribution to the Body of Knowledge

The major contributions of this dissertation are:

- This dissertation predicts remaining useful life of water pipes using combination of Artificial Neural Network (ANN) and ANFIS.
- Above models have not been used in pervious literature for determination of remaining useful life of water pipes.

1.9 Dissertation Organization

This dissertation consists of six chapters. Chapter 1 presents history of water distribution system, water main classification, factors influencing deterioration of water mains, research needs and objectives, methodology and expected outcome of this dissertation. Chapter 2 provides a literature review on remaining useful life of water pipes and previous studies about deterioration of models of water pipes. Chapter 3 describes the methodology used for this dissertation by outlining a systematic description of the research performed and presents the development and the application of ANN and ANFIS models to water distribution network. Chapter 4 outlines statistical analysis of the research. Chapter 5 analyzes the results achieved in this dissertation. Details of the methodology and results are presented in detail and discussed to demonstrate the practicality and predictive capability of the ANN and ANFIS. Chapter 6 draws conclusions, summarizes the research results, and presents recommendations for further study. Appendices and references are provided at the end of this dissertation.

1.10 Chapter Summary

This chapter discussed the history of water distribution system, types of pipe materials followed by an introduction to cast iron, ductile iron, PVC pipes and asbestos cement pipes, water main classification, factors influencing deterioration of water mains, condition rating of water mains, water system risk assessment, pipe characteristics and failure modes were presented. Additionally, this chapter reviewed the research needs and objectives, methodology and expected outcome of this dissertation.

Chapter 2

Literature Review

2.1 Introduction

The previous chapter illustrated the water pipe failure and characteristic of different types of water pipes. This chapter provides an overview of the researches and various efforts related to remaining useful life prediction of water pipe, structural and risk assessment models of water pipes failures. Researchers have used a broad variety of techniques to predict water pipes deterioration and remaining useful life of water pipes. Although pipelines are designed for a lifespan under standard operating conditions, their deterioration never follows a set pattern. The process of deterioration of pipelines is rather complex because there are many factors which interactively contribute to such deterioration. Environmental, operational and physical interactions ensure that pipeline deterioration is never uniform (Kulandaivel, 2004). Pervious researches about pipeline deterioration techniques are summarized in the following sections.

2.1.1 Multiple Linear Regressions

Multiple linear regression can be used to study the relationship between a single dependent variable and one or more independent variables (Allison, 1999). The general form of a multiple linear regression equation presents in Eq.2.1:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta n X n + \varepsilon \qquad \text{Eq. 2.1}$$

Where Y = dependent variable; α = intercept; $\beta_{1\ldots}$ β_n = regression coefficients; $X_{1\ldots}$ X_n = independent variables; and ε = error term. α, and $\beta_1 \ldots \beta_n$ values are determined by using ordinary least squares method. Multiple linear regressions are one of the most commonly used statistical methods to determine relationships between independent variables and dependent variables.

2.1.2. Logistic Regression

Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). Logistic regression can be considered as a linear regression in which the dependent variable is transformed into the logit of dichotomous output variable Y. The goal of logistic regression is to find the best fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables (Eq.2.2). Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest (Wilson, 2014).

$$logit(Y) = ln\mathrm{p}(Y=1)/1 - P(Y=1) \; = \; \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta n X n + \varepsilon \quad \text{Eq. 2.2}$$

Where Y = dichotomous dependent variable; α = intercept; $\beta_{1\ldots}$ βn = regression coefficients; $X_{1\ldots}X_n$ = independent variables; and ε = error term.

2.1.3 Logistic Function

Cooper et al., (2000) developed a probabilistic, statistical model to predict the probability of failure of large-diameter mains. The model calculates a score for each pipe relative to other pipes in the system based on their calculated probability of failure. The failure model uses a logistics function model to calculate each probability score. The model was developed to be combined with a consequence of failure model to determine a risk score for each main. This consequence model was developed using geographic information system (GIS) data. The failure model discussed here takes the general form:

$$\text{Probe} = \frac{1}{1+e^{-z}} \quad \text{Eq. 2.3}$$

Where $Z = b_0 + X_1b_1 + X_2b_2 + \ldots\ldots + X_ib_i$ are all independent variables, and the '*b*' coefficients are determined using logistic regression. Generalized linear modelling techniques were used on 11 variables to define those that are most significant for use in the model (Wilson, 2014).

2.1.4 Hierarchical Fuzzy Expert System Model

Hierarchical fuzzy expert system model is developed to evaluate the risk of water main failure. The developed model considers four main risk factors, which have sixteen sub-factors that represent both corrosion of the water distribution network and the failure results. Based on literature review, a failure risk scale is planned to help decision makers in water resources management (i.e., companies, municipalities) make informed decisions and start their rehabilitation plans. The developed model is analyzed and verified, and the results show that the model is strong and reliable (Fares and Zayed, 2008).

2.1.5 Delphi Method

Uncertainty is the key concern in strategic decision making for designing sustainable infrastructure and it's believed that the performance models gives reliable prediction mostly in short and medium term rather than long term since the technological and economic factors will change during the time. Therefore, the early step in water researches is to understand the input parameters for prediction model in long term. ''Ranking-type" Delphi is used to develop a group agreement on identifying the important factors that affect deterioration of water mains in distribution networks. In the process of the Delphi, the problem statement should be described at first. Then a panel of experts familiar with the subject is selected. Since Delphi is a group decision technique, it needs several experienced experts familiar with the issue who could deliver correct answers to the questions. Therefore, the process of selecting skilled panelist is of main importance (Zangenemadar and Moslehi, 2016b).

2.1.6 Artificial Neural Network Models

Artificial Neural Networks (ANNs) are forms of mathematical models that simulate the structure and/or function of biological neural networks. It consists of interconnected neurons and processes. In most cases, ANN is adaptive during the learning phase by changing the structure based on the information flow. Modern neural networks are non-linear statistical tools for data modeling. They are usually used to identify patterns in the data, train the network for adjustment, and predict future conditions. Neural Networks can identify complex non-linear relationships between contributing factors, output can adapt and trained to fuzzy or roughly defined problems (Syachrani, 2010). An artificial neural network (ANN) is composed of artificial neurons that are connected to each other and arranged in different layers (input, hidden, and output) (Al-Barqawi and Zayed, 2008). Each connection between neurons has a connected weight, and once the weighted sum of inputs reaches a certain level; each neuron sends a signal that is a resultant value of a start function (Zou et al., 2008). ANNs are trained by using datasets, which include observed input and output patterns.

Advantages of ANNs include the ability to model any complicated nonlinear relationship between input and output variables in cases where the precise form of relationship is not known by using a proper network structure, and to exhibit high error tolerance (Zou et al., 2008). Once trained, a neural network can perform classification, clustering and forecasting tasks. Thus, the pipeline industry can employ this technology to model dynamic deterioration of pipes using the historically available data (Najafi and Kulandaivel, 2005). Once the pipeline deterioration pattern is modeled, it is then possible to predict future condition and deficits of the pipelines (Najafi and Kulandaivel, 2005). However, dependency on a widespread amount of data points is a drawback of this method.

2.1.7 Machine Learning Models

Machine learning based models that are not constrained to a pre-determined model form could be a feasible alternative for developing effective predictive condition assessment models. The machine learning based models developed for water pipes can be classified as neural network based fuzzy logic, polynomial regression, and Bayesian theory (Jenkins, 2014).

2.1.8 Adaptive Neuro-Fuzzy Interference System

Modify network-based fuzzy inference (ANFIS) is a combination of two soft-computing methods of ANN and fuzzy logic (Jang, 1993). Fuzzy logic can change the qualitative aspects of human knowledge and insights into the process of precise quantitative analysis. However, it does not have a defined method that can be used as a guide in the process of transformation and human thought into rule base fuzzy inference system (FIS), and it also takes quite a long time to adjust the membership functions (MFs) (Jang,1993). Unlike ANN, it has a higher capability in the learning process to adapt to its environment. Therefore, the ANN can be used to automatically adjust the MFs and reduce the rate of errors in the determination of rules in fuzzy logic (Suparta and Alhasa, 2016). Table 2-1 presents the advantages and limitations of water pipes deterioration models.

Table 2-1 Deterioration Models of Water Pipes

Adapted from Salman, 2010

Methods	Advantages	Limitations
Multiple Linear Regression	Linear regression is a simple method and can be generated by using common spreadsheet applications.	Validity of the model depends on satisfying several assumptions. The functional relationship between dependent and independent variables is assumed linear.
Logistic Regression	Probability of a pipe segment to be in a deficient state can be obtained by logistic regression.	This method is applicable for identifying the odds ratio associated with dichotomous (such as fail-not fail) variables.
Artificial Neural Networks (ANN)	In ANN models exact functional form does not have to be identified beforehand. Complex nonlinear relationships can be modeled.	An extensive dataset is required for the model to learn all possible combinations.
Delphi Method	Useful method for decision making based on opinions of experts.	It needs several experienced experts familiar with the issue who could deliver correct answers to the questions (Zangenemadar and Moslehi, 2016b).
Hierarchical Fuzzy Expert System Model	Useful for evaluation of water pipes risk failure.	The method require certain factors or levels of belief as means of accounting for uncertainty. All data must be complete and correct to be analyzed (Fares and Zayed, 2008).
Machine Learning Models	Feasible alternative for developing effective predictive condition assessment models.	There is a lack of theory around model creation (Jenkins, 2014).
Adaptive Neuro-Fuzzy Interference System	The system might be initialized with or without prior knowledge in terms of fuzzy rules.	It does not have a defined method that can be used as a guide in the process of transformation into rule base fuzzy inference system (FIS), and it also takes quite a long time to adjust the membership functions (Kruse, 2008).

2.2 Estimating of Remaining Useful Life of Water Mains

There are several groups of factors that affect pipe corrosion, and therefore, the remaining useful life. These factors are wall thickness measurement, water pipe structural failure, pipe condition assessment, risk assessment model, physical, condition, and environmental and operational factors (Figure 2-1). Because of the lack of historical data and availability of physical data in most cases, most studies consider only the physical factors in their research. In this research, the factors of pipe length, diameter, and material, number of break, installation year, condition, and age are used as inputs and the target value to predict the remaining useful life of pipelines.

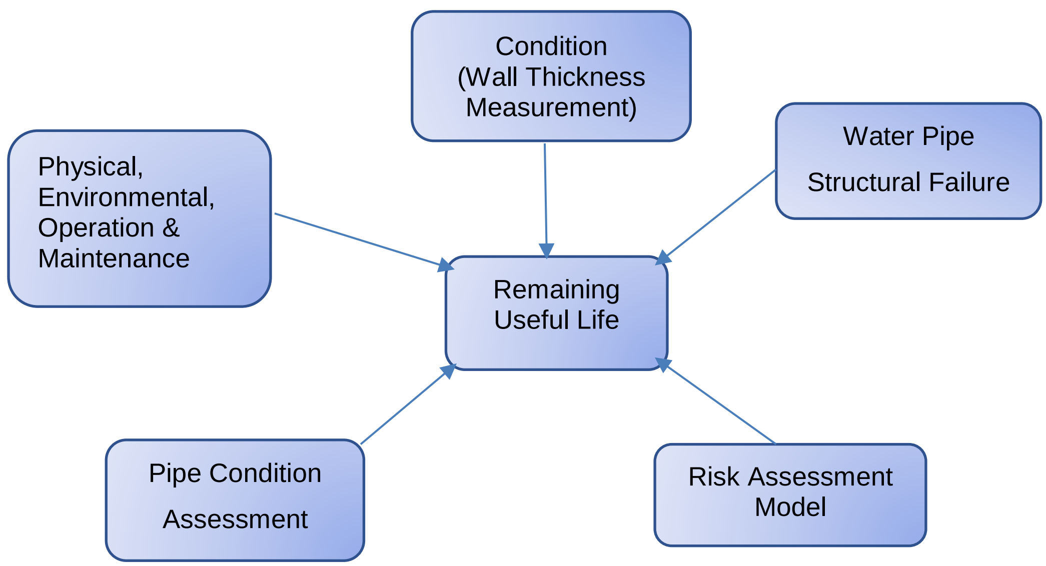


Figure 2-1 Four Groups of Factors Impacting Pipe Deterioration

2.2.1 Pipe Diameter

It has been found that the corrosion of larger diameter pipes is slower compared to that of smaller diameter pipes in general. Kettler and Goutler (1985) have found a strong connection between the pipe break rate and pipe diameter for cast iron pipes but not for asbestos cement pipes. Though the reasons for the reducing tendency for breakage with increasing pipe diameter have not been completely understood, significant evidence implies that the reverse relationship may result from the greater pipe wall thickness of larger diameter pipes. On the other hand, larger diameter pipes can have a larger surface area in contact with surrounding soil, thereby resulting in an increased corrosion of the pipe (Wang, 2009).

2.2.2 Pipe Section Length

In general, the longer a pipe section is the more breaks may happen. This may be because the longer the pipe, the more conditions that it is exposed to. The increased number of conditions may cause breakage (Skipworth et al., 2002). This may because longer pipe runs mean fewer or less sever bends in the pipe to accumulate debris, creating blockages or damage to the pipe from standing sewage (Najafi and Gokhale, 2005).

2.2.3 Pipe Material

As stated in chapter 1, there are varieties of materials, e.g., cast iron, ductile iron, steel, asbestos cement, reinforced concrete, Prestressed Concrete Cylinder Pipe (PCCP), polyethylene, and PVC, for pipeline. Each material has its own benefits and limitations. Pipe material is a significant factor in the corrosion process; pipes made from different materials fail in different ways (Zangenemadar and Moslehi, 2016a).

2.2.4 Pipe Breakage Rate

Breakage rate is the number of breaks per year per unit length of pipeline. When the breakage rate is higher, there are more breaks in one segment, which consequences in a higher corrosion rate. This decreases the durability of the pipeline and increases the possibility of failure (Zangenemadar and Moslehi, 2016a).

2.2.5 Pipe Age

As a pipe gets older, it may require more maintenance and repairs. Many rehabilitation plans have been based only on the age of the pipes. Pipe age is an important factor that indicates the period that the pipes withstand the effect of the surrounded environment, operational condition, and external loads (Fahmy and Moslehi, 2009). Figure 2-2 presents the factors impact on remaining useful life.

Pipe age is determined by calculating the number of years from a pipe's installation year to the present day. It is assumed that the pipe has been frequently in service throughout its whole life span (Nemeth, 2016). Pipe age was calculated by Eq. 2.7:

$$\text{Age} = (\text{present Year} - \text{installation Year}) \qquad \text{Eq. 2.7}$$

Where: Age = age of pipe (years)

Present Year = current year (years)

Installation Year = year of installation (years)

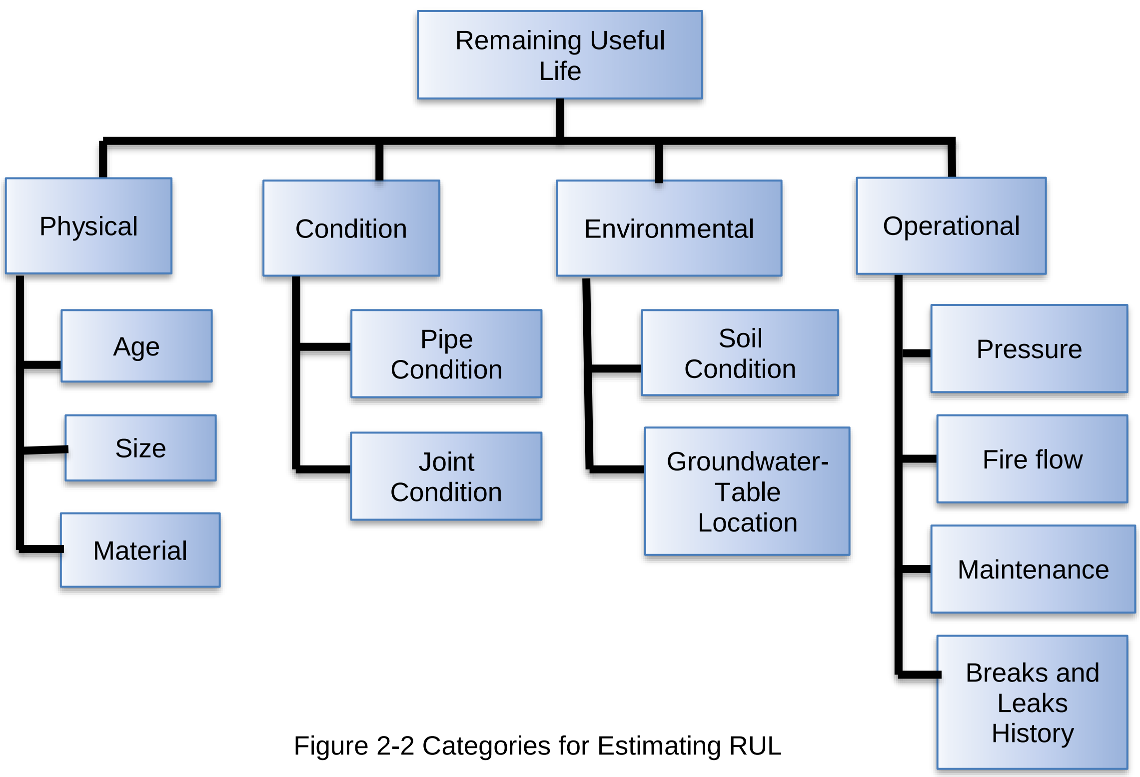


Figure 2-2 Categories for Estimating RUL

Adapted from Nemeth, 2016

2.3 Remaining Useful Life (RUL)

The primary focus of this section is the computation of the remaining useful life of each water main. There are various methods to calculate RUL. According to Devera (2013), RUL is the difference between the pipe's age and Anticipated Service Life. Slightly modified from the equation presented by Devera (2013). Eq. 2.8 presents calculation of RUL.

$$RUL = (ASL - Age) \qquad \text{Eq. 2.8}$$

Where: RUL = remaining useful life of pipe (years)

ASL = anticipated service life (years)

Age = pipe age from year of installation to present (years)

Then, RUL for each water main in the water distribution system is calculated using the Eq. 2.4:

$$RUL = (ASL - Age) \times P_{adj} \qquad \text{Eq. 2.4}$$

Where: RUL = remaining useful life of pipe (years)

ASL = anticipated service life (years)

Age = pipe age from year of installation to present (years)

P_{adj} = break history percent adjustment

Clark et al. (1982) provides a linear model that predicts the number of years from installation until the first break. Cortez (2015) concluded that based on available data the Clark et al. (1982) linear model offered the best results. According to Cortez (2015), the data required to use this model includes time of installation, pipe age, expected service life, pipe diameter, pipe material, pipe length, internal pressure, breakage history and pipe wall thickness loss . Incorporating additional factors efforts to decrease the uncertainty and variation in a pipe's service life based on additional operating conditions.

The Clark et al. (1982) linear model is based on Eq. 2.5:

$$NY = x_1 + x_2 D + x_3 P + x_4 I + x_5 W + x_6 L + x_7 T \qquad \text{Eq. 2.5}$$

Where:

NY = number of years from installation to first repair

X_i = regression Parameters

D = diameter of pipe (in)

P = absolute pressure within pipe (psi)

L = length of pipe

W = wall thickness loss

T = pipe type

Remaining useful life is the difference between the pipe’s age and NY (Cortz, 2015). RUL was calculated based on Eq. 2.6:

$$RUL = (NY – Age) \qquad \text{Eq. 2.6}$$

Where: RUL = remaining useful life of pipe (years)

NY = number of years from installation to first failure (years)

Age = pipe age from year of installation to present (years)

Cortez (2015) adjusted the RUL for breakage history by decreasing the RUL by 10% for each previous break. Using the adjusted RUL, the probability of failure score is determined (Nemeth, 2016). Decreasing the amount of money spent on incorrect or unsuitable amounts of utility data.

Zangenemadar and Moslehi (2016a) proposed a model that considers several factors in prediction of pipelines’ remaining life for different pipe materials in water distribution networks. The main objective of this research was to forecast the remaining useful life of pipelines based on its characteristics, to develop value-driven enhanced intervention plans.

Above study employed Artificial Neural Networks (ANN) to relate the pipe characteristics to the output values, which in this case was the remaining useful life. The model was implemented, validated, and tested using data obtained from the municipality of Canada.

The result showed as pipelines deteriorate, they were more exposed to failure internally and/or externally; the ANN model showed robustness in the prediction of remaining useful life. Modeling the remaining useful life of pipelines through ANN was valuable, as it reduced the time and effort and therefore the costs of analysis.

Fahmy and Moslehi (2009) presented a model designed to forecast the remaining useful life of cast iron water mains. The model considers factors related to pipe properties, its operating conditions and the external environment that surrounds the pipe.

Three different data-driven techniques are considered in the model development; each is used to study the relationship between remaining useful life and a set of deterioration factors, and to forecast remaining useful life of cast iron water mains. These techniques are multiple regression and two types of artificial neural networks: multilayer perceptron and general regression neural network. The data used in model development were acquired from 16 municipalities in Canada and the United States.

The results showed the outcome produced by the developed models correlate well with the actual conditions. The study presented in this paper revealed that data-driven modeling methods are effective in forecasting the remaining useful life of cast iron water mains and it overcomes limitations associated with existing models (Fahmy and Moslehi, 2009).

2.4 Water System Risk Assessment

Risk is determined by considering the Consequence of Failure (COF) and Likelihood of Failure (LOF) of an asset and translating both of those factors into a standardized numerical risk score. The COF focuses on the impact a failure may have on an asset's ability to maintain a target level of service. The LOF is the probability that an asset will fail at any given time and is based on condition and capacity. A COF and LOF score was allocated for each waterline asset (Providence Infrastructure Consultant, 2016).

On the other hand, the numerical score of LOF is related to the asset's condition, capacity, and increases over time as the asset ages and deteriorates or is relied upon to express increasing amounts of flow. Using a risk-ranking concept is an applicable tool for managing infrastructure assets and has been applied as a practical way to identify and rank Capital Improvement Projects (CIP). This approach also reduces the amount of CIP spending overtime. To understand the Likelihood of Failure (LOF) means the utility must know where its pipeline limitations are based on history of breaks, knowledge of age and system, environmental considerations, pipe materials, construction of pipeline, and operational data.

To understand the Consequence of Failure (COF) means the utility must understand the consequences of failure for each linear asset based on their servicing requirements, system redundancies, and impact on stakeholders and surroundings (Wurst, 2016). The more severe the outcome of failure of a pipeline for a utility, the more important it is for the utility to know and understand the condition of the asset. Five categories were developed to characterize the Likelihood of Failure (Wurst, 2016). The categories were:

- Physical Characteristics
- Exposure Factors
- Pipe Condition & Performance
- Design pressure & Load Carrying Capability
- Maintenance Effectiveness

2.5 Water System Asset Management

Asset Management is a systematic process of operating, maintaining and upgrading physical assets cost-effectively. It combines engineering and mathematical analysis with sound business practice and economic theory (Najafi, 2016). Two main aspects of information on pipeline performance that can be used in asset management decision-making processes are (1) information on current condition, which is obtained through field inspection, and (2) information on future performance, obtained using deterioration models and forecasting tools (Najafi, 2016).

2.6 Risk Assessment Model

A risk management framework provides a structured approach to classifying, assessing and managing risks. Because some risk factors were more important or are better predictors of condition than other factors, the individual risk factors were assigned a weighting based on a percentage from 1 to 5, baseline conditions of this study were established in the risk analysis using available information and engineering judgment. The overall values of LOF and COF were assigned scores in five categories to help define the arranging of pipelines for condition assessment activities. The five risk categories are shown in Figure 2-3.

Negligibl	Minor	Moderate	Major	Severe
1	**2**	**3**	**4**	**5**

Figure 2-3 Risk Rating Categories

Wurst, 2016

Using the overall LOF and COF scores from the Risk Model, the pipelines were plotted in a risk matrix as shown in Figure 2. LOF is plotted on the X-axis, while COF is plotted on the Y-axis (Figure 2-4).

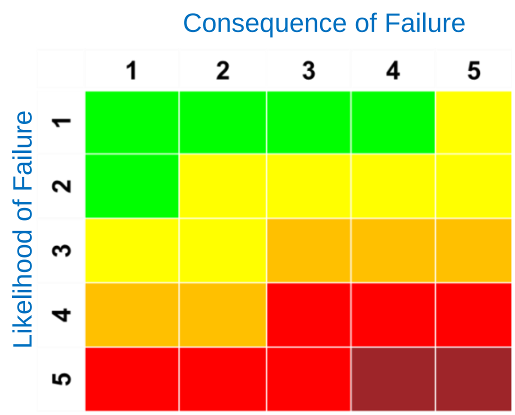


Figure 2-4 Risk Scoring Matrix

Wurst, 2016

2.7 Risk Evaluation of Pipelines

Salman (2010) developed a risk assessment tool at an individual pipe level by combining the probability of failure values determined by statistical corrosion modeling of sewer pipes and consequences of failure values determined by examining the geographical, physical, and functional characteristics of sewer pipes using expert opinions that reflect the relative importance of these features.

To determine probability of failure values, three statistical methods, namely ordinal regression (proportional odds model), multinomial logistic regression, and binary logistic regression, were employed in consecutive steps. Expert opinion was obtained from a local sewer agency to assess geographical, physical, and functional characteristics of pipes in terms of consequences of failure. Three methods were employed to evaluate risk of failure: multiplication, risk matrices, and fuzzy inference systems (Salman, 2010).

The results showed use of risk matrices overcame the limitation of the development method by allowing different levels of risk values to be allocated to different combinations of probability and consequence of failure values (Salman, 2010). Finally, fuzzy inference systems were used to represent the fuzziness in probability, consequence and risk of failure variables; and to allocate risk values based on fuzzy rules. Based on the outcomes, the use of fuzzy inferences led to a better representation of failure risk of sewer pipes.

Fares and Zayed (2008) developed a risk model for water main failure, which evaluated the risk associated with each pipeline in the network. This model considered four main factors: environmental, physical, operational, and post-failure factors (consequences of failure) and sixteen sub-factors, which represented the main factors.

The required data were collected from literature review and through a questionnaire sent to the experts in the field of water distribution network management. To develop the risk of failure model, hierarchical fuzzy expert system (HFES) technique was used to process the input data, which was the effect of risk factors, and generated the risk of failure index of each water main. To validate the developed model, a validated AHP corrosion model and two real water distribution network data sets were used to check the results of the developed model (Fares and Zayed, 2008).

The results presented that pipe age was found to have the most significant factor of water main failure risk, followed by pipe material and breakage rate. Average Validity Percent is 74.8 %, which was reasonable considering the uncertainty involved in the collected data.

Wade (2016) presented a summary of a tiered (or step-wise) approach that has been used by several cities and wastewater utilities to raise their sewer renewal and improvement programs using a risk-based methodology to select the best trenchless solution(s). The methodology was gathering existing information including GIS-based maps, supplemental GPS, work orders, existing inspection data, record drawings, and other asset data to begin the process (Wade, 2016).

The result showed access to a large and well-equipped toolbox that represented multiple inspection and assessment technologies has opened the door for better condition information. This included average and maximum wall loss for concrete pipe, discovery of exterior voids, ovality, condition and location of remaining reinforcement, exact pipeline orientation, and flow characteristics.

2.7.1 Risk Assessment Models

Nemeth (2016) compared two risk assessment models, a statistically complex model and a simplified model. Based on the physical, environmental, and operational conditions of each pipe, these models estimated the probability of failure, quantify the consequences of a failure, and ultimately determine the risk of failure of a pipe.

The risk analysis required the application of Bentley's WaterCAD, Microsoft's Excel and Visual Basic for applications, and ESRI's ArcGIS. WaterCAD provides the hydraulic properties of each pipe within the system. This study verified that a risk assessment model was applicable in identifying critical components and developing a pipe maintenance schedule. Utilization of a risk evaluation model would allow municipalities to efficiently assign funds and enhance their water distribution system. Application of computer modeling and data analysis programs was suggested for future research (Nemeth, 2016).

Tuhovcak and Mika (2013) summarized a basic theory of indirect condition assessment including several definitions of condition assessment, factors contributing to deterioration and possible outputs of the condition assessment. Furthermore, the legal requirements concerning this issue specifically in the Czech Republic and the Slovak Republic were listed.

The methodology was based on condition indicators (CI) formulated using a multi-objective optimization. The results showed that proposed methodology achieved good results and it showed certain significant benefits e.g., satisfactory level of detail, optional user modification (Tuhovcak and Mika, 2013).

Kulandaivel (2004), developed a pipeline condition prediction model based on neural network algorithm, which can recognize pipelines at risk of degradation. Furthermore, he identified useful pipeline performance data sources for deterioration model development and evaluated the performance of the neural network model with test data. The result showed this model assisted the municipal agencies in prevention of the risks involved with pipeline failures by prioritizing the parts of the network that needs immediate action and enhance their limited inspection and maintenance budget by applying resources where they are most effective (Kulandaivel, 2004).

2.8 Condition Rating of Water Mains

As most of the municipalities are in short of capital funds in planning the repair, rehabilitation, and replacement of their deteriorated water mains, condition rating practices are very important in helping them to classify their water main assets and assign their limited budgets. Unfortunately, there is no standard condition rating practice so far. In practice, many municipalities have developed or are using modified weighting systems for rating their water mains. However, some municipalities may not even have such a system in place to help them do the decision-making, and they rely on the spot valuation of a pipe break to decide a follow-up solution. Pipe needed to be replaced when there were three or more pipe breaks per 300 meters of the pipe. However, the manual did not give details how this was decided, and it did not remark the time interval between these three or more breaks. Walski and Pellicia (1982) have questioned this rule that, wondering, for example, whether, in a pipe of less than 327.75 feet, if one break has occurred, does the pipe need to be replaced according to this rule (Wang, 2009).

2.9 Condition Assessment of Water pipes

Condition assessment is a continuing process to evaluate the condition of a water system, data and information are gathered through observation, direct inspection, investigation, and indirect monitoring and reporting as a result, better determine the likelihood of failure of those assets. An analysis of the data and information helps determine structural and operational issues, and performance of the system. Condition assessment also includes failure analysis to determine the causes of infrastructure failures and to develop ways to avoid future breakdowns. Condition assessment improves the ability of utilities to make technically sound judgments regarding asset management (EPA, 2013).

2.9.1 Condition Assessment Technologies

Condition assessment technologies can be roughly divided into two categories: indirect methods, and direct methods. Indirect methods contain analysis of break history data, performing potential corrosion surveys, and soil sampling and testing. These methods are quite inexpensive and provide understanding into potential condition of the pipeline, but do not provide direct indication of the condition of the pipeline (Agarwal, 2010). On the other hand, direct methods, while usually more expensive, provides data that is obtained by direct observation and examination of pipe condition. Direct methods cover three broad areas of visual inspection, leak detection, and pipe wall thickness evaluation. Selection of suitable technology for a specific application is related to two factors (Agarwal, 2010). Figure 2-5 presents classifications of condition assessment technologies.

The first factor is to match the technology with the level of risk. For low risk assets, the utility may limit the scope of assessment to only indirect methods. Moreover, for very high-risk assets, the utility may consider not only the indirect methods, but also the more expensive direct methods. The second factor to consider is to match the cost with efficiency and level of coverage (Habibian, 2016). Most technologies for condition assessment of large numbers of a current buried pipeline include placing some type of device within the pipe. Given the two considerations of pressurized flow and typically smaller size, access to the interior of water pipes for condition assessment is usually much more difficult than access to the interior of wastewater pipes. Further, for wastewater pipelines the application of closed circuit television (CCTV) evaluation of the pipe has been found to be very valuable in pipe condition assessment of wastewater systems (Habibian, 2016).

On the other hand, CCTV assessment of water pipelines has had only limited application. The results of these differences in condition assessment of water and wastewater systems is that the capacity to exactly and reliably predict the probability of failure is greater in wastewater systems than in water systems. Thus, more condition assessment has been done on wastewater systems, than on water systems, so there are more uncertainties and unknowns in predicting the possibility of failure (or the remaining expected lifespan) of a water pipeline than a similar wastewater pipeline (WRF, 2013).

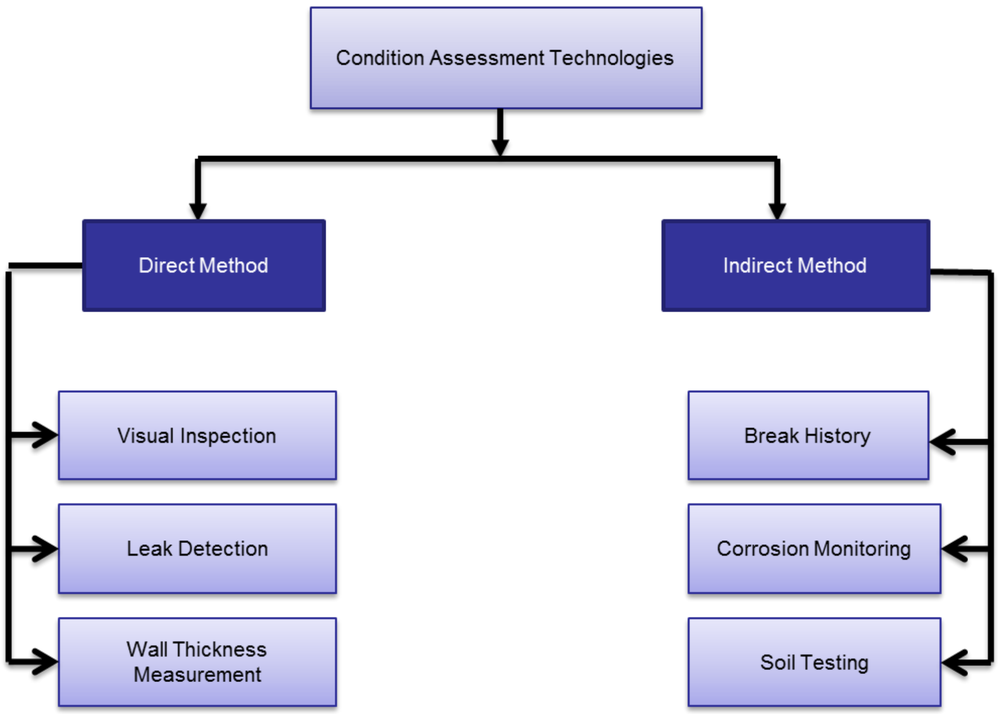


Figure 2-5 Condition Assessment Technology for Water Mains

Habibian, 2016

2.9.2 Direct Methods for Inspection of Water Mains

These are portable microprocessor-based devices that pinpoint leaks automatically based on a cross-correlation method. In this method, acoustic leak signals are measured with vibration sensors or hydrophones at two pipe contact points (usually fire hydrants or valves) that bracket the location of a suspected leak. The leak signals are transmitted from the sensors to the correlator wirelessly. In most cases, the leak is located asymmetrically between measurement points (Fahmy, 2009). Consequently, there is a time lag between the measured leak signals. The time lag is found from the cross-correlation function of the leak signals. In the presence of a leak, the cross-correlation function has a distinct peak at the time shift between leak signals. The location of the leak is calculated based on an algebraic relationship between the time lag, the sensor-to-sensor distance and the propagation velocity of sound waves in the pipe (Makar, 2001).

2.9.2.1 ePulse® Inspection Technology and Procedure

The ePulse® technology uses acoustic field data along with a pipeline's characteristics to estimate remaining structural wall thickness. For field-testing, sensors were magnetically attached to the operating nut of two valves and the section of pipe between these valves was "in bracket". After installation of the sensors, a baseline recording was taken for the bracketed pipe segment. If during the recording a leak was suspected, the EchoWave® leak detection technology was employed to determine the location of the leak (Providence Infrastructure Consultants, 2016).

Following the baseline recording, a sound wave, or noise, was created by tapping on a nearby hydrant that was outside of the bracketed pipe segment. The sensors measured the time it took for the sound wave to travel between the sensors and the speed of the sound wave was used to estimate the pipeline's structural wall thickness. This data was transmitted to a receiver and base unit where the data was stored for processing later offsite. A graphical depiction of the field-testing procedure is presented in Figure 2-6.

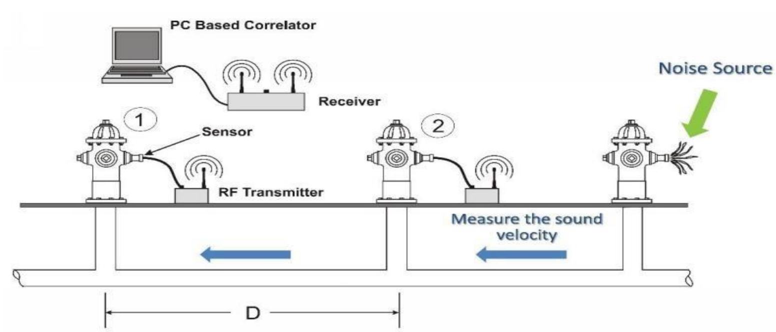


Figure 2-6 Acoustic Testing Procedure

Providence Infrastructure Consultants, 2016

2.9.2.2 EchoWave® Leak Detection Technology and Procedure

EchoWave® is a proprietary technology employed by Echologics for leak detection. Leak detection was performed simultaneously with the ePulse® acoustic testing on selected pipes. A correlator listened for leak noise, and if a leak was identified, the time delay between the sensors was used to locate the leak. The two (2) sensors record the sound and the correlator uses the distance between the sensors to measure the time it takes for the leak noise to arrive at each sensor. Ground sounding microphones were also used to confirm the location of leaks (Providence Infrastructure Consultants, 2016). A graphical depiction of the leak detection procedure is provided in Figure 2-7, however, in this figure, the sensors are attached to hydrants instead of valves.

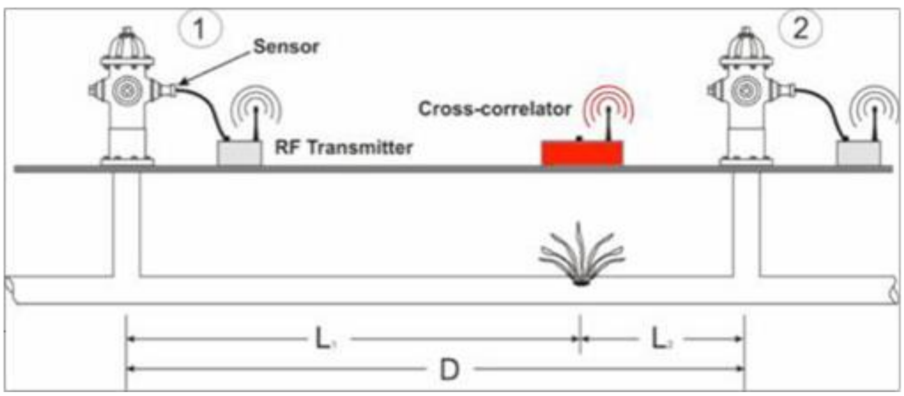


Figure 2-7 Acoustic Testing Procedure

Providence Infrastructure Consultants, 2016

2.9.3 Indirect Methods for Condition Assessment of Water Mains

A general diagnosis of the actual structural state of municipal water infrastructure systems is needed. To make a diagnosis, one must collect and analyze data on the characteristics of water pipes and on their breakage histories (Fahmy, 2009). Unfortunately, many municipalities have been carefully recording breakage histories only for a decade, while their pipes have been in the ground for much longer. Rajani and Kleiner

(2004) noted that the most important information that should be collected for the condition assessment of water mains includes the following:

- Pipe information (i.e., pipe diameter, wall thickness, date of installation, depth of burial, and manufacturing spun/ cast).
- Soil condition, (i.e., soil type, soil· pH, soil density, soil resistivity, and aeration quality).
- Installation information (laying condition, load factor).
- Operational conditions (water pressure, surge pressures, summer and winter air and water temperatures, wheel loads, vehicle impact factor, frost load factor).

The location and timing of the samples collected each year should be based either on a set number of pipe breaks or on a set of specific locations within the water system (Fahmy, 2009). Furthermore, tests on soil samples collected near the excavated pipe samples provide additional information to forecast deterioration pit growth. It should be noted that indirect methods have been used to get corrosion pit features by using empirical methods for pipes of the same age that were buried in the same soil type.

2.10 Structural Inspection and Monitoring Techniques for Water Pipes

2.10.1 CCTV with Laser Profile Adapters

CCTV is one of the most common visual inspection techniques used in water pipelines. However, this technique only provides unprocessed data, and assessment of change in pipe shape and size is difficult and inaccurate. Profiling adapter is a special attachment that can be added to the front of the CCTV camera. These cameras are mounted on remote-controlled platforms and video recording is controlled remotely (Covilakam, 2011).

2.10.2 Magnetic Flux Leakage (MFL)

This method works on the principle that when a magnet is placed next to a pipe wall, most of the magnetic flux lines would pass through the pipe wall. That is, pipe wall is a preferred path for the flux. This method can be used to measure leakages and the remaining metal loss of water pipes (Covilakam, 2011).

2.10.3 Remote Field Eddy Current (RFEC)

RFEC method is based on the principle that when an energized coil is brought near the surface of a metal component, eddy currents are induced in the system. A typical set up consists of an exciter coil that generates a direct electromagnetic field that travels inside the pipe. A small magnetic field sensor is positioned some distance away (Covilakam, 2011). Concurrently, the exciter generates another indirect field that travels through the pipe wall with minor attenuation. Changes in field strength and attenuation are dependent on pipe wall thickness and thus the signature of these changes enables the determination of pipe wall thickness (Covilakam, 2011).

2.10.4 Ultrasonic Technologies (US)

Ultrasonic measurements are among some of the best-established methods for simple external testing of points along a steel pipeline wall. Ultrasonic monitoring measures the propagation time for high frequency, short wavelength mechanical waves through metallic pipe wall, and this data is correlated with the pipe wall thickness (Covilakam, 2011).

2.10.5 Fiber Optic Sensors

Optical fiber is a thin, flexible, transparent fiber that can be used widely for remote sensing. Fiber-optic sensors (also called optical fiber sensors) are fiber-based devices for sensing some quantity, typically temperature or mechanical strain, but sometimes also displacements, vibrations, pressure, acceleration, rotations, or concentrations of chemical species. The general principle of fiber optic sensor is that light from a laser or any other super luminescent source is sent through an optical fiber and when these optical fiber experiences subtle changes in parameters there will be a change in the measurement of light reaching the detector (Covilakam, 2011).

2.11 Structural Failure of Water Mains

The structural deterioration of water mains and their following failure are compound processes. Many factors can affect the rate of the deterioration of water distribution systems and lead to their failure. These factors can be physical, external environmental, or operational. Kleiner and Rajani (2001), Kroon (2001) have studied the degradation of metallic and PCCP water mains, which represent over 70% of the total network of water mains in North America. Their studies revealed that corrosion is largely responsible for both metallic pipes and PCCP failure.

Kroon (2001) described pitting corrosion, which is initiated by a localized anodic point on the metal surface. Their studies disclosed that corrosion is largely responsible for both metallic pipes and PCCP failure. Kroon (2001) described pitting corrosion, which is initiated by a localized anodic point on the metal surface. The penetration of the metal continues at this same point because a large area around the pit is cathodic to the pit itself. Pitting corrosion is commonly encountered at pinholes or flaws in dielectric coating systems. For steel and stainless steel, chloride ions are well known as a cause of pitting attacks.

In such cases, either a corrosion pit in the pipe wall grows from the inside or the outside surface until the pipe has been completely penetrated, allowing water leakage from the pipe. Water mains typically break when the amount of corrosion (or degradation) is such that the main is no longer able to withstand the forces acting on it (Fahmy, 2009).

The potential consequences of failure in a given pipeline segment is the most significant factor in determining the level and type of effort that should be invested in collecting the various types of data about the water mains. The structural failure modes for each of the common water main materials show in Table 2-2.

2.12 Bathtub Curve

The life cycle of a water main pipe may be represented by a bathtub curve (Rajani and Kleiner, 2004). A bathtub curve defines the rate of failure in respect to the service life of the pipe (Figure 2-8). The early part of the curve shows “infantile failure” which for pipes is demonstrative of failure due to human factors in the actual laying of the pipe (manufacturing faults, tend to appear during that part). A period follows in which failure rate is generally low (Najafi and Gokhale, 2005). When failure does happen, it may depend on many factors, such as excessive loads not designed for, or settlement. As the pipes tend towards the end of their useful life the failure rate increases exponentially. The “Bath Tub” curve can be applied to an individual pipe, a group of pipes with similar characteristics or the whole population of a pipe network (Najafi and Gokhale, 2005).

Table 2-2 Structural Failure Modes for Common Water Main Materials

Adapted from Fahmy, 2009

Water Main Material	Structural Failure Modes
Cast Iron (CI) Small diam (<15 in.) Large diam (>20 in.) Medium diam (15-20 in.)	• Circumferential Breaks, split bell, corrosion through holes • Longitudinal breaks, bell shear, corrosion through holes • Same as small, plus longitudinal breaks and spiral cracking, blown section
Ductile Iron (DI)	• Corrosion through holes
Steel	• Corrosion through holes, large diameter is susceptible to collapse
Polyvinyl Chloride (PVC)	• Longitudinal breaks due to excessive mechanical stress • Susceptible to impact failure in extreme cold condition
High Density Polyethylene (HDPE)	• Joint imperfections, mechanical degradation from • improper installation methods, susceptible to vacuum collapse for lower pressure ratings
Asbestos Cement (AC)	• Circumferential breaks, pipe degradation in aggressive water • Longitudinal splits
Concrete Pressure Pipe (CPP) and Prestressed Concrete Cylinder Pipe (PCCP)	• Ruptures due to loss of pre-stressing upon multiple wire failure. Pipe degradation in particularly aggressive soils, corrosion of pipe canister, concrete damage due to improper installation methods

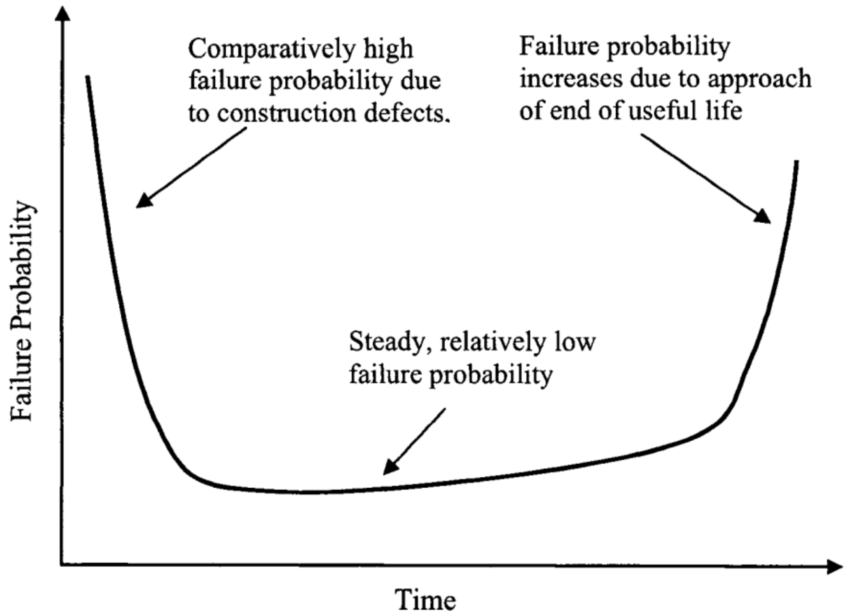


Figure 2-8 Bathtub Curve of Pipe Performance with Age

Najafi and Gokhale, 2005

2.13 Chapter Summary

This chapter reviewed the existing literature on the several deterioration models that used for risk assessment and calculation remaining useful life of water pipes. The deterioration models discussed in this chapter included fuzzy logic, Analytical Hierarchy Process (AHP), Artificial Neural Network (ANN), hierarchical fuzzy expert system and logistic regression. Water condition assessment, structural failure and asset management of water pipes were discussed. The literature review implies that there is a good possibility to develop a successful neural network-based model if the critical parameters that provide the deterioration of pipelines are achieved. A neural network model for predicting pipeline performance trends based on historical condition assessment data will be developed in this dissertation.

Chapter 3 Neural Network and Neuro Fuzzy Application

3.1 Introduction

This chapter presents the methodology used in this dissertation for the development of a water pipeline remaining useful life model. The remaining useful life model is based on neural network modeling technique. A detailed description of neural networks is presented in this chapter, along with the ANFIS method appropriate for remaining useful life prediction.

3.2 Artificial Intelligence and Neural Networks

The term artificial intelligence usually used to refer to the field of computer science devoted to producing programs that try to be as smart as humans are. Expert systems and neural networks are two forms of artificial intelligence, each with separate strengths and weaknesses. Most applications of artificial intelligence are programs that simulate either the deductive or the inductive intelligence of human being (Kulandaivel, 2004). Deductive systems, which can be simulated by expert systems, require rules or instructions performed one at a time to arrive at the answer. By contrast, induction takes in a large amount of information all at once and then draws a conclusion. Neural networks can be used to simulate the inductive behavior of humans (Kulandaivel, 2004). Once trained, the neural network can look at input data and produce an appreciate answer. In a comparison to expert system, Garrett (1992) presents the following advantages to the use of neural networks:

- Neural networks have the capability to present a model for a situation where only examples are presented.
- Expert systems require “certain factors” or “levels of certainty’, whereas neural networks are trained to deal with uncertainty since training data is gained from situations very close to the situations in which the network will operate.

- Expert systems are very brittle in that all data must be complete and correct for a system to be analyzed. On the other hand, neural networks can allow for minor errors in the input data and for minor deviations from existing training cases.

Neural Networks can be used to:

- Identify patterns and images
- Construct a decision tree to solve a problem
- Classify data
- Predict outcomes
- Study thematic evolution of a process and construct cost effective models (Kulandaivel, 2004).

3.2.1 The Neural Network Algorithm

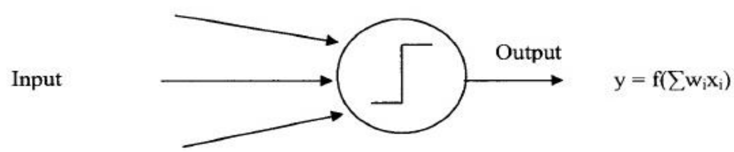
There are many types of neural networks, but all have three things in common. A neural network can be expressed in terms of its individual neurons, the connections between them (topology), and its learning rule. Both biological and artificial neural networks contain neurons, real or simulated. These neurons have many connections to each other, which transfer information. The knowledge of a network is allocated across the interconnections between the neurons. A typical neuron receives input, from many other neurons (Kulandaivel, 2004).

A neuron calculates its own output by finding the weighted sum of its inputs, generating an activation level and passing that through an output on transfer function. The point where two neurons communicate is called a connection. The strength of the connection between two neurons is called a weight.

The collection of weights arranged in rows and columns is called the weight matrix. A neural network learns by changing its response as the inputs change. Because the weights in the network can change, the relationship of the network's output to its inputs can be changed as well. In this sense, the learning rule determines the behavior of the network and how that behavior can change over time (Kulandaivel, 2004).

Single neuron artificial neurons as information processing devices were first proposed more than fifty years ago. As shown in Figure 3-1, a neuron calculates a weighted summation of its *n* inputs, the result of which is then threshold to give a dual output y, which is either +1 or -1. The bias weight, 0, is presented whose input is fixed at +1. This bias weight is adaptive like the others and its use allows better flexibility of the learning process.

Figure 3-1



Schematic Diagram of an Artificial Neuron

Kulandaivel, 2004

For a classification problem, the neuron allocates input patterns, signified by the vector of numbers X = (X_1, X_2...., X_n), either to class A (for which y would be +1) or class B (for which y would be -1). The function is presented in Eq. 3.1:

$$Y= f(\sum_{i=1}^{n} wixi)= \begin{cases} 1 \text{ when } \sum_{i=1}^{n} wixi > 0 \\ -1 \text{ when } \sum_{i=1}^{n} wixi \geq 0 \end{cases} \quad \text{Eq. 3.1}$$

In the above equation (3.1), y is the neuron output and f is a hard-limiting of threshold function, sometimes known as the neuron's transfer function, which gives an output of +1 whenever $\sum wixi$ is greater than zero (the threshold value) or -1 whenever $\sum wixi$ is less than (or equal to) zero.

The learning process is to correct all the weights and let the output y method the desired output so that the neuron performs the classification task properly. Multi-class problems can also be solved by having several neurons operating similar (Kulandaivel, 2004).

3.2.2 Backpropagation Neural Network (BPNN)

Back propagation Neural Network was created by generalizing the Widrow-Hoff learning rule to multiple layer networks and non-linear differentiable transfer function. Input vectors and corresponding target vectors are used to train a network until it can estimate a function, associate input vectors with specific output vectors, or classify input vectors in a correct way. Networks with biases, a sigmoid layer and a linear output layer are able of like any function with a finite number of breaks. The back-propagation algorithm contains of two paths, forward path and backward path (Fahmy, 2009). Forward path contains creating a feed forward network, adjusting weight, simulation and training the network. The network weights and biases are updated in the backward path. Feed forward networks often have one or more hidden layers of sigmoid neurons followed by output layer of linear neurons. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output paths (Fahmy, 2009).

The linear output layer allows the network produce values outside the range -1 to +1 (Figure 3-2). On the other hand, if we want to confine the outputs of the network between 0 and 1, then the output layer should use a log-sigmoid transfer function. Before training a feed forward network, the weight and biases must be adjusted. Once the network weights and biases have been set, the network is ready for training. Random numbers are used around zero to prepare weights and biases in the network. The training process requires a set of appropriate inputs and targets as outputs. During training, the weights and biases of the network are adjusted to reduce the network performance function. The default performance function for feed forward networks is mean square errors, i.e., the average squared errors between the network outputs and the target output (Fahmy, 2009).

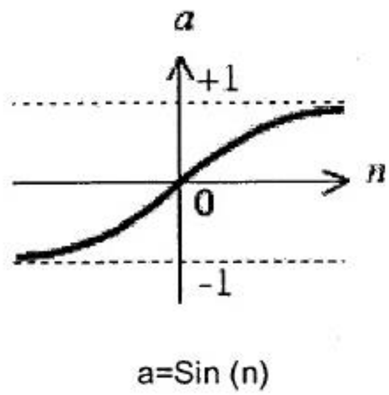


Figure 3-2 Sigmoid Transfer Function

Fahmy, 2009

By far, the BPNN is the most common one used for mathematical modeling backpropagation learning scheme by which a layered neural network with constantly valued neurons is trained to become a pattern-matching machine. It offers a way of using examples of a target function to find the weights that make a certain mapping function hidden in the neural network estimated the target function as carefully possible. As shown in Figure 3-3, the neurons of the networks are structured in multiple layers: input, hidden, and output. Each hidden-layer neuron obtains input from all neurons in the input layer through weighted connections (w). In addition, each neuron is related with a bias term, called the threshold, 0. This bias term works as a horizontal shift for the source of the transfer function to adjust the level of incoming signals to the neuron. Values of both w and 0 for a given neural network are determined during the training phase.

Back Propagation (BP) algorithm, which is used in this study, introduced in 1986, BP has developed as the most prevalent learning algorithm for the training of MLP model (Haykin, 1999).

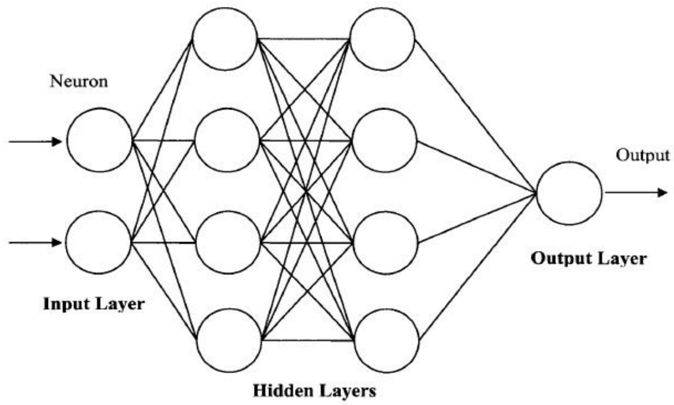


Figure 3-3 Regression Plot

Haykin, 1999

3.2.3 BPNN Modeling

The BPNN network operates in two modes: mapping and training mode. In mapping mode, information flows forward through the network from inputs to outputs. In the training mode, the information flow substitutes between forward and backward. In the mapping mode, the network processes one example at a time, producing an estimate of values of the dependent variables based on the values of the independent variables for the given example. First, a set of values for the independent variables is loaded onto the input layer of the network. The input-layer neurons do no calculation - each neuron merely sends a copy of its value to all the hidden-layer neurons. Each hidden neuron computes the weighted sum of the inputs using its single connection strengths as weights. Next, each hidden neuron calculates a transfer function of its input sum and sends the result to all the output-layer neurons. Then, each output-layer neuron completes a similar calculation and outputs the resulting value as an estimate of the dependent variable it represents.

The training mode refers to the procedure in which the network is showing examples with correct output values known. The training algorithm contains of three steps. In the first step, the training patterns attained from the database are fed into the input layer of the network. These inputs are propagated through the network until reaching the output layer. The output of each neuron is calculated by transfer functions Eq. 3.2 and Eq. 3.3 (Lou et al., 1999):

$$a = \sum_{i=1}^{n} wixi \qquad \text{Eq. 3.2}$$

$$O = f\,(a) = \frac{1}{1+e^{-ga}} \qquad \text{Eq. 3.3}$$

Where;

O = neuron output,

a = input to the transfer function

X_i = i^{th} input,

W_i = weight of connection i,

g = gain of sigmoid function, and

n = number of inputs to one neuron

In the second step, the neural network outputs are subtracted from the desired values to achieve an error signal. This error signal is the source for the coming backpropagation step. The Eq. 3.4 expresses the error signal (Lou et al, 1999):

$$E_{RMS} = \sqrt{\frac{\sum_{j=1}^{N}\sum_{j=1}^{N}(T_{JK}-O_{JK})^2}{N_0 N_e}} \qquad \text{Eq. 3.4}$$

Where:

E_{RMS} = root mean square error

N_o = number of neurons in the output layer

N_e = total number of patterns in an epoch,

T_{jk} = target (desired) value of the j_{th} neuron, and the k_{th} pattern, and

O_{jk} = output of the j_{th} neuron, and the k_{th} pattern

In the third step, error is reduced by the backpropagation of the error signal through the neural network. In this process, the involvement of each hidden neuron is calculated, and matching weight corrections needed to reduce the error are derived. For each output neuron k, compute the 5 values, Eq. 3.5 defines as follows:

$$\boldsymbol{\delta}_j = f\ (x_j)\ \sum_k \delta_K W_{JK} \qquad \text{Eq. 3.5}$$

Where:

$\boldsymbol{\delta}_j$ = adjusted error of hidden neuron j;

X_j = input to the hidden neuron j;

$\boldsymbol{\delta}_k$ = adjusted error of output neuron k connected with hidden neuron j; and

$W_{\mathbf{jk}}$ = connection weight between neuron j and k.

The weight connecting any two neurons is updated by Eq. 3.6:

$P \longrightarrow q$

$$Vw_{qb} = \alpha\ \boldsymbol{\delta}_q\ O_p \qquad \text{Eq.3.6}$$

Vw_{qb} = adjustment of weight between preceding layer neuron p and proceeding layer neuron q;

$\boldsymbol{\delta}_q$ = adjusted error of proceeding layer neuron q;

O_p = output of preceding layer neuron p; and

α = learning coefficient (a positive constant).

The training process repeats steps 1 through 3 for all patterns in the training set until the overall error is adequately low based on a given standard. If the network has not met then go back to the step 1, otherwise stop training. Once trained, the neural network has the ability of adjusting to changing input. If the trained network results in good accuracy on the testing and validation data set, the development process is done (Lou et al., 1999).

To determine whether or not the database covers enough samples, two factors will be measured important: (1) the form of the target function: to maintain a given accuracy, sample size needs to increase as the target function becomes more compound, and (2) the noise in the data: to continue a given correctness, sample size needs to increase as noise increases. Given a target function of a certain difficulty, and a certain amount of noise in the data, there will be a complete limit to the precision the model can succeed. An infinite sample size would be needed to attain the limit of exactness (Kulandaivel, 2004).

For neural network modeling, if the sample is large enough, the difficulty of the network's mapping function could be improved to match the difficulty of the target function. Therefore, as sample size rises, neural network model's exactness will be restricted only by the noise in the data. Generally, neural network can take advantage more from large samples than regression can. Because larger samples let us to use more hidden neurons or to remain training longer, the precision can be enhanced by increasing sample size.

On the other hand, neural network model does not entail a larger sample than a regression model. As the sample size gets smaller, we can use fewer hidden neurons or stop training earlier to avoid overfitting. The basic rule, therefore, is to use the largest sample available.

3.2.4 Neural Network Training and Testing

Training a neural network includes continually giving a set of examples (facts) to the network. The network takes each input, makes an estimate as to the output, checks this estimate contrary to the output (correct answer), and adjusts the original connections (weights) if its estimate is incorrect. This process is repeated for each fact until the network learns the facts sufficient to be useful (Najafi and Kulandaivel, 2005). The training file generated using Net Maker is accessed through Brain Maker for modeling analysis. Histograms and the Network Progress Display, two useful tools provided by Brain Maker, can help define whether the network is making advancement in training and still has ability to learn (Najafi and Kulandaivel, 2005).

The processes for importing, training, testing, and saving results are explained here. First, the data is imported from a file and must be formatted correctly. Second, the ANN tool in MATLAB is used to select the data used for training, testing and validation, along with the number of hidden layers before training. Third, the training is started to build the neural network in an iterative manner. Several algorithms are available within MATLAB to train the neural network and all of them are gradient-descent methods (also known as back-propagation method).

Fourth, once the neural network is constructed, statistical measures are assessed to evaluate the performance of the ANN model. Multiple training simulations were run to confirm the created neural network was precise (Nishiyama, 2013).

Testing the network is the same as training it, but the network is shown with the examples it has never seen before, and no weight changes are made during testing. The testing process simply employs the constructed model from training and validation to assess performance. Validation happens after the neural network has been developed. Validating a network consists of presenting it with new input data and collecting the network outputs. Unlike testing, there is no identified output, only identified inputs in the validation.

Validation and testing are both a process of certifying overfitting does not happen. Overfitting is the case of an over-trained model that identically imitators the input data pattern and does not model the relations between variables. If validation and testing are significantly under-performing compared to the training process it shows possible overfitting. To guarantee this is not the case the validation and testing process is detected during the construction process, and if their performances show signs of diminishing then training should be stopped and re-evaluated (Nishiyama, 2013). While the model is being trained, validation and testing happen concurrently. Figure 3-4 illustrates the overall flowchart of the model development.

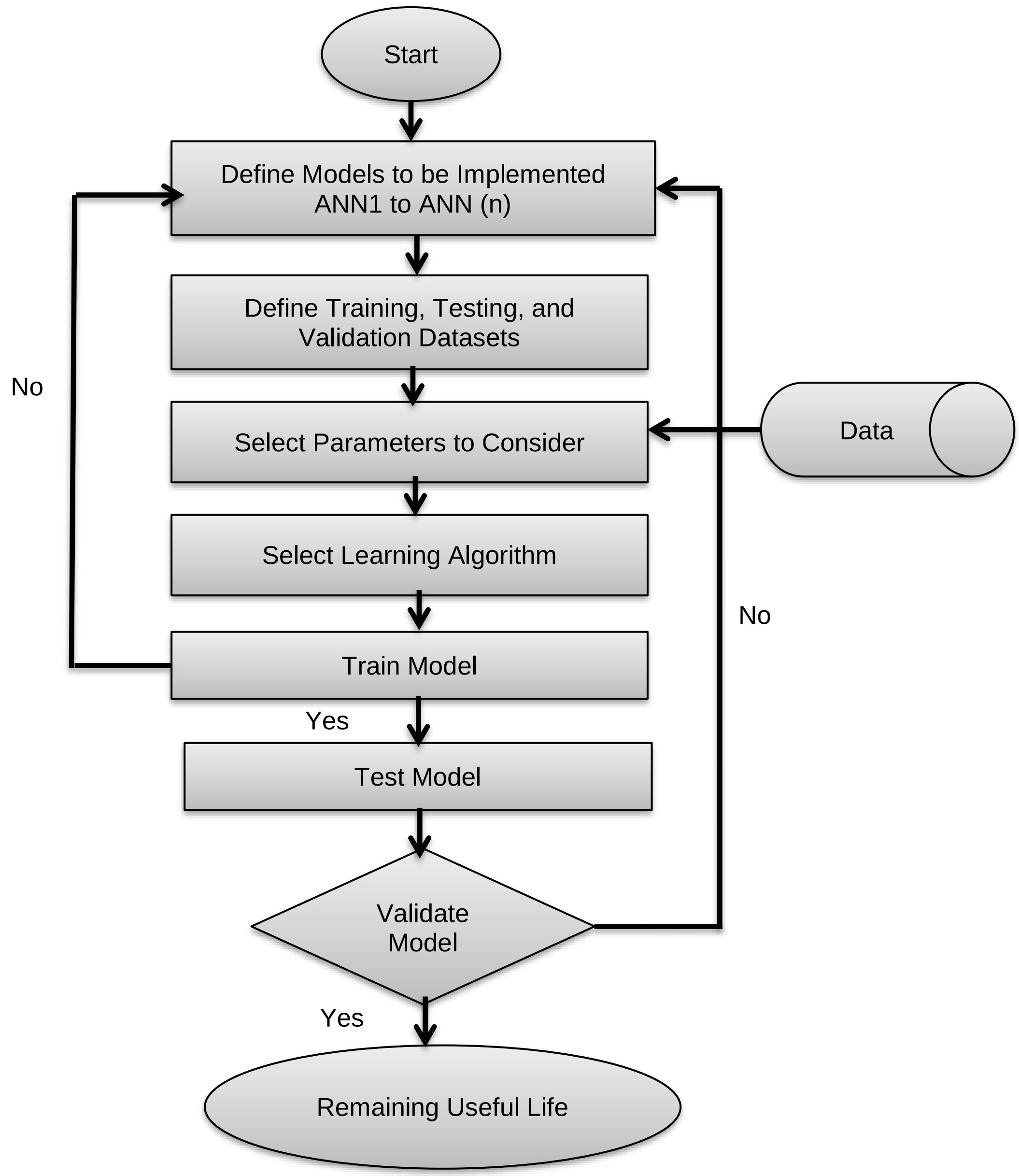


Figure 3-4 Model Development Framework

3.2.5 Neural Networks and Statistical Modeling: A Comparison

The statistical modeling aimed to find an equation that detention the general shape of a relationship, which is usually derived from observed examples. Therefore, the fields of statistical modeling and neural networks are closely connected in the context of input-output mapping.

The main difference between these two fields is that traditional statistical models usually need an equation to be quantified, which could be difficult in complex nonlinear cases, while neural networks have been mostly used to deal with nonlinear problems without needing a pre-specified function form. However, with the advent of the backpropagation neural network (BPNN), trace carefully in solving mathematical modeling problems. This technique resolves one of the central problems in neural networks, and it is a useful modeling tool as well (Kulandaivel, 2004).

The training of the network is repetitive for many examples in the set until the network reaches a stable state, where there are no more important changes in the weights. Thus, the network learns from the samples by building an input-output mapping for the problem at hand. Statistical modeling techniques are used to derive correlations between variables from examples as well.

3.2.6 Selection of Optimal Number of Hidden Neurons

Selection of ideal number of hidden neurons is a significant subject in the neural network training process. The aim of training is to gain a neural network with best simplification ability. Simplification is defined as the ability of a neural network to store in its weights overall features, which are common to a group of samples. Usually, a neural network with too few hidden neurons will not be capable to learn adequately from the training data set, whereas a neural network with too many hidden neurons will let the network to remember the training set instead of simplifying the learned knowledge for future hidden instances. Unfortunately, there is no accurate formula for defining the ideal number of hidden neurons for a given application. There are numerous ways to define a good number of hidden neurons. One solution is to train the neural networks with the number of hidden neurons calculated using Eq. 3.7 (Kulandaivel, 2004):

$$\text{Number of Hidden Neurons} = \frac{\text{of Data sets} - \text{Outputs}}{\text{c}\ (\#\ \text{input}+\#\text{output}+1)} \qquad \text{Eq. 3.7}$$

Where, C = 2 - 5.

Therefore the # of Neurons = (269 - 1) / 3 × (18 + 1 + 1) ~ 4 Neurons

The second equation suggested in Brain Maker manual is:

Number of Hidden Neurons = (# Inputs + # Outputs)/2

= (18 + l)**/**2 ~ 9 Neurons

The third solution is to start with a small number of hidden neurons and add more while training if the network is not learning. In this research, the first method (4 Hidden Neurons) was used to train firstly and regularly more neurons are added to train several neural networks with varying number of hidden neurons. The neural network that resulted in the least testing error was selected, resulting in the best simplification ability. The "testing while training" method was used to trace the testing errors (simplification ability) of the neural network during training process. After training, it was suitable to find the best network with the least testing errors.

3.2.7 Application of Artificial Neural Networks in Pipeline Prediction Modeling

Software ANN was chosen for this modeling because of its ability to cover nonlinear and compound behavior of water networks. Furthermore, it covers many variables that increase the system's performance reliability (Lawrence, 1994). In this research the ANN models were developed, trained, validated, and tested in MATLAB. Eight ANN models were developed with one hidden layer that is different in two aspects: the number of neurons in the hidden layer and the random groups of data sets. The ANN1, ANN2 … and ANN10 models have 3, 4,5,6,7 … and 10 neurons in their hidden layers, respectively. There is a close relationship between age, length, material, wall thickness loss and remaining useful life. Thus, pipe material, wall thickness loss, length, diameter, and age are selected from the database as the input values for ANN models. Figure 3-5 depicts the typical structure of the neural network with the given set of input parameters. The number of hidden layers was determined through trial runs of the model. The dataset is split randomly into training (75%), validation (10%) and testing (15%). The performance charts are presented in Figure C-1 through Figure C-8 (Appendix C).

For each ANN model, trials were performed to reach the lowest error. The performance of the models was assessed based on R^2, mean absolute error (MAE), relative absolute error (RAE), root relative square error (RRSE), and mean absolute percentage error (MAPE) according to the Eq. 3.8 – Eq. 3.12:

$$R^2 = 1 - \frac{\sum i\,(ti-oi)^2}{\sum i(ti-\frac{1}{n}\sum ti)^2} \qquad \text{Eq. 3.8}$$

$$MAE = \frac{1}{n}\sum i\,|ti - oi| \qquad \text{Eq. 3.9}$$

$$RAE = \frac{\sum i\,|ti-oi|}{\sum i\,|ti-1/n \sum \sum_i ti|} \qquad \text{Eq. 3.10}$$

$$RRSE = \sqrt{\frac{\sum i\,(ti-oi)^2}{\sum i\,(ti-\frac{1}{n}\sum_i ti)^2}} \qquad \text{Eq. 3.11}$$

Where:

oi = output parameter;

$$MAPE = \frac{100}{n}\sum_i \frac{|ti-0i|}{ti} \qquad \text{Eq. 3.12}$$

R^2 = coefficient of determination;

ti = target parameter;

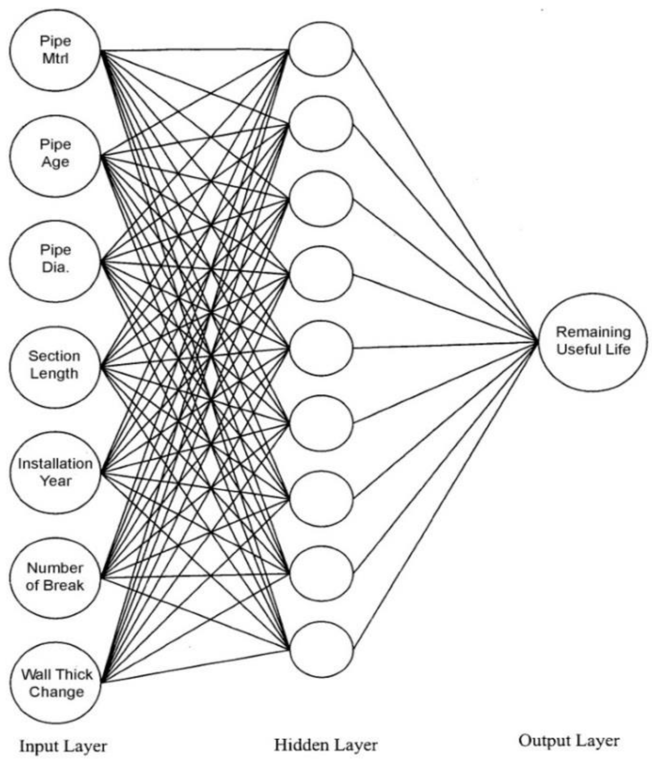


Figure 3-5 Structure of the ANN Models

The coefficient of determination (R^2) is used in statistical analysis repeatedly, as it is easy to calculate and understand. It fluctuates between [0, 1] and assesses the percentage of total differences between estimated and target values with respect to the average. MAE is an absolute measure and ranges from 0 to $+\infty$. One advantage of MAE is that it is not affected by outliers and can be calculated as an alternative for mean square error (MSE). RAE is less influenced by outliers, same as MAE; however, it is contaminated by extremely large or small values. The relative absolute error (RAE) and root relative square error (RRSE) assess the performance of a forecasting model in the same way. In fact, lower RAE and RRSE result in better performance of the forecasting model. In this research, MAPE is used mostly to evaluate the accuracy of a model, because of its simplicity.

It identifies error as the proportion of actual data, and higher precision comes with a lower MAPE (Zangenemadar and Moslehi, 2016a). The training is stopped when the neural network settles at the lowest possible training and testing errors and there is no visible convergence in the training statistics. Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Regression R values measure the correlation between outputs and targets. An R-value of 1 means a close relationship.

3.2.8 Calculation of Neuron Output

The basic unit of computation in a neural network is the neuron, often called a node or unit. It receives input from some other nodes or from an external source and computes an output. Each input has an associated weight (w), which is assigned based on its relative importance to other inputs (Karn, 2016). The node applies a function f (defined below) to the weighted sum of its inputs as shown in Figure 3-6.

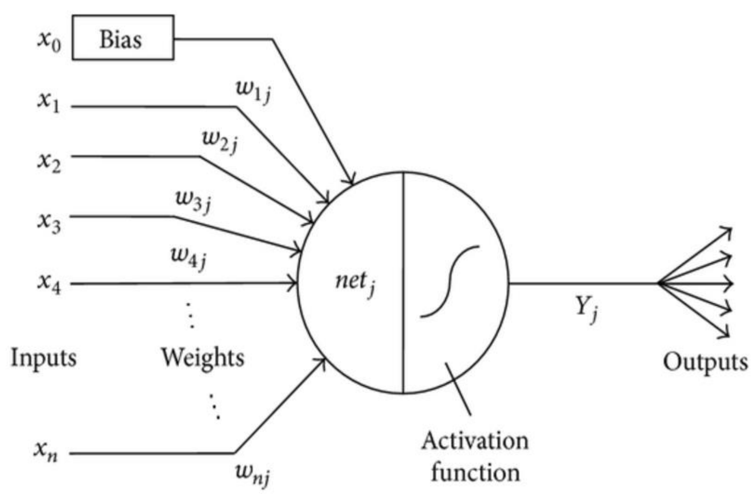


Figure 3-6 Output of Neuron

Karn, 2016

The above network takes numerical inputs x_1 and x_2 and x_n has weights w_1 and w_2 and w_n associated with those inputs. Additionally, there is another input 1 with weight b (called the Bias) associated with it. The output of neuron for each input parameters calculated using Eq. 3.13. The purpose of the activation function is to introduce non-linearity into the output of a neuron. Every activation function (or non-linearity) takes a single number and performs a certain fixed mathematical operation on it (Karn, 2016). Activation function in Backpropagation algorithm is sigmoid logistic function and it varies between -1 and +1 (Eq. 3.14):

Output for each input parameter: $\sum_{i=1}^{n} WiXi + bi$ Eq. 3.13

$$Ƒ\left(\sum_{i=1}^{n} WiXi + b\right) = \frac{1}{1+e^{-(\sum wx+b)}} \quad \text{Eq. 3.14}$$

Where:

X_i = i^{th} input,

W_i = weight of connection i,

B_i = bias of connection i,

n = number of inputs to one neuron

3.3 Adaptive Neural Fuzzy Inference System

Adaptive Neural Fuzzy Inference System (ANFIS) Using a given input/output data set, the toolbox function ANFIS creates a fuzzy inference system (FIS) whose membership function parameters are (adjusted) using either a backpropagation algorithm alone, or in combination with a least squares type of method. This allows your fuzzy systems to learn from the data they are modeling. ANFIS was built on the three main components, namely basic rules, where it covers the selection of fuzzy logic rules “If-Then;” as a function of the fuzzy set membership; and reasoning fuzzy inference techniques from basic rules to get the output. Figure 3-7 shows the detailed structure of the ANFIS.

ANFIS will work when the input that comprises the actual value is transformed into fuzzy values using the fuzzification process through its membership function, where the fuzzy value has a range between 0 and 1. The basic rules and databases are referred to as the knowledge base, where both are key elements in decision-making. Usually, the database contains descriptions such as information on fuzzy sets parameter with a function that has been defined for every existing linguistic variable (Suparta and Alhasa, 2016).

There are several types of ANFIS, namely Takagi–Sugeno, Mamdani, and Tsukamoto (Cheng et al., 2005). An ANFIS of Takagi–Sugeno model was found to be widely used in the application of ANFIS method.

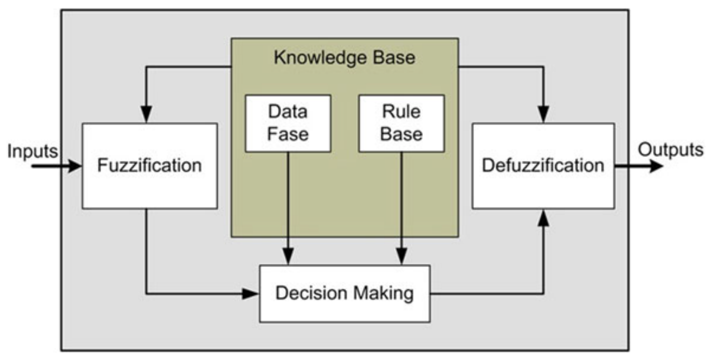


Figure 3-7 Fuzzy Inference System

Suparta and Alhasa, 2016

3.3.1 ANFIS characteristics

Compared to a common neural network, connection weights and propagation and activation functions of fuzzy neural networks vary a lot. Although there are many different methods to model a fuzzy neural network, most of them agree on certain characteristics such as the following:

- A neuro-fuzzy system based on an underlying fuzzy system is trained by means of a data-driven learning method derived from neural network theory. This only considers local information to cause local changes in the fundamental fuzzy system (Kruse, 2008).

- It can be represented as a set of fuzzy rules at any time of the learning process, i.e., before, during and after. Thus, the system might be initialized with or without prior knowledge in terms of fuzzy rules.
- The learning process is constrained to assurance the semantic properties of the underlying fuzzy system.
- A neuro-fuzzy system estimates an n-dimensional unknown function, which is partly represented by training examples. Fuzzy rules can therefore be interpreted as unclear examples of the training data (Kruse, 2008).

3.3.2 Adaptive Neural Fuzzy Inference System (ANFIS) Application

ANFIS techniques provide a method for the fuzzy modeling procedure to learn information about a data set, to calculate the membership function parameters that best allow the related fuzzy inference system to track the given input/output data (Gershteyn and Perman, 2003).

This learning method works similarly to that of neural networks. Algorithm defined by Jang (1993) and creates a fuzzy decision tree to categorize the data into linear regression models to diminish the sum of squared errors (SSE) base on Eq. 3.14:

$$SSE = \sum e_j^2 \qquad \text{Eq. 3.14}$$

Where:

ej: the error between the desired and the actual output

These techniques provide a method for the fuzzy modeling procedure to learn information about a data set, to calculate the membership function parameters that best allow the related fuzzy inference system to track the given input/output data (Gershteyn and Perman, 2003).

3.3.3 ANFIS Approach

ANFIS is employed to model the relationship between the input parameters (material, age, diameter, length, installation year, number of break and wall thickness change). To apply the ANFIS, seven inputs (material, age, length, diameter, number of break, year, and wall thickness loss) and a single output (RUL) are considered in the fuzzy inference system. The Sugeno fuzzy model is employed in fuzzification and defuzzification of the system (Lin and Huang, 2013).

The complete ANFIS consists of five layers, the fuzzy layer, production layer, normalization layer, de-fuzzy layer, and total output layer. Each layer includes several nodes, which are defined by the node function (Figure 3-8). The node function in each layer, which performs the same operation, is detailed below:

Layer 1 is the fuzzy layer, in which every node is an adaptive node with node function. The output is the product of all the incoming signals. Each node represents the fire strength of the rule. The Gaussian function was adopted in this research. Where we have seven inputs of nodes and μ_{Ai}, μ_{Bj}, μ_{Ck}, μ_{Dk}, μ_{Ek}, μ_{Fk} and μ_{Gk} denote the membership functions of the fuzzy set.

$O^1_{Ai} = \mu_{Ai}$ (Material), $\quad i = 1, 2, 3, 4, 5, 6, 7$

$O^1_{Bj} = \mu_{Bj}$ (Age), $\quad j = 1, 2, 3, 4, 5, 6, 7$

$O^1_{Ck} = \mu_{Ck}$ (Length), $\quad k = 1, 2, 3, 4, 5, 6, 7$

$O^1_{Dk} = \mu_{Dk}$ (Diameter), $\quad L = 1, 2, 3, 4, 5, 6, 7$

$O^1_{Ek} = \mu_{Ek}$ (Break), $\quad M = 1, 2, 3, 4, 5, 6, 7$

$O^1_{Fk} = \mu_{Fk}$ (Year), $\quad N = 1, 2, 3, 4, 5, 6, 7$

$O^1_{Gk} = \mu_{Gk}$ (Wall Thickness Loss), $\quad O = 1, 2, 3, 4, 5, 6, 7$

Layer 2 is the production layer, in which every node is a fixed node with node function to multiply input signals to serve as output signal (Eq. 3.15).

$O^2_{ijklmno} = \mu_{Ai} \times \mu_{Bj} \times \mu_{Ck} \times \mu_{Dk} \times \mu_{Ek} \times \mu_{Fk} \times \mu_{Gk} = w_{ijklmno}$ i, j, k, l, m, n, o = 1,2,3,4,5,6,7

Eq. 3.15

Layer 3 is the normalization layer, in which every node is a fixed node with node function to normalize firing strength by calculating the ratio of this node firing strength to the sum of the firing strength (Eq. 3.16):

$$O^3_{ijklmno} = \overline{w}_{ijklmno} = \frac{\mathrm{Wijklmno}}{\sum i \sum j \sum k \sum l \sum m \sum n \sum o\ (\mathrm{Wijklmno})} \quad i, j, k, l, m, n, o = 1,2,3,4,5,6,7$$

Eq. 3.16

Where $\overline{w}_{ijklmno}$ denotes the normalized firing strength (output).

Layer 4 is the de-fuzzy layer, in which every node is an adaptive node with node function to calculate the consequence of each fuzzy rule using Eq. 3.17:

$O^4_{jklmno} = \overline{w}_{ijklmno} f_{ijklmno} = w_{ijklmno} (p_{ijklmno}$ (Material) + $q_{ijklmno}$ (Age) + $r_{ijklmno}$ (Length) + $s_{ijklmno}$ (diameter) + $t_{ijklmno}$ (Break) + $u_{ijklmno}$ (Year) + $v_{ijklmno}$ (Wall thickness) + $w_{ijklmno})$

i, j, k, l, m, n, o = 1,2,3,4,5,6,7   Eq. 3.17

Layer 5 is the total output layer, in which the single node is a fixed node with node function to calculate the overall output (Lin and Huang, 2013) based on Eq. 3.18:

$RUL = O^5_1 = \sum_i^7 \sum_j^7 \sum_k^7 \sum_l^7 \sum_m^7 \sum_n^7 \sum_o^8 \ \overline{w}\mathrm{ijklmno}\ f_{iijklmno}$ i, j, k, l, m, n, o = 1,2,3,4,5,6,7

Eq. 3.18

Where RUL represents the inferred output (i.e., the predicted remaining useful life) of the ANFIS.

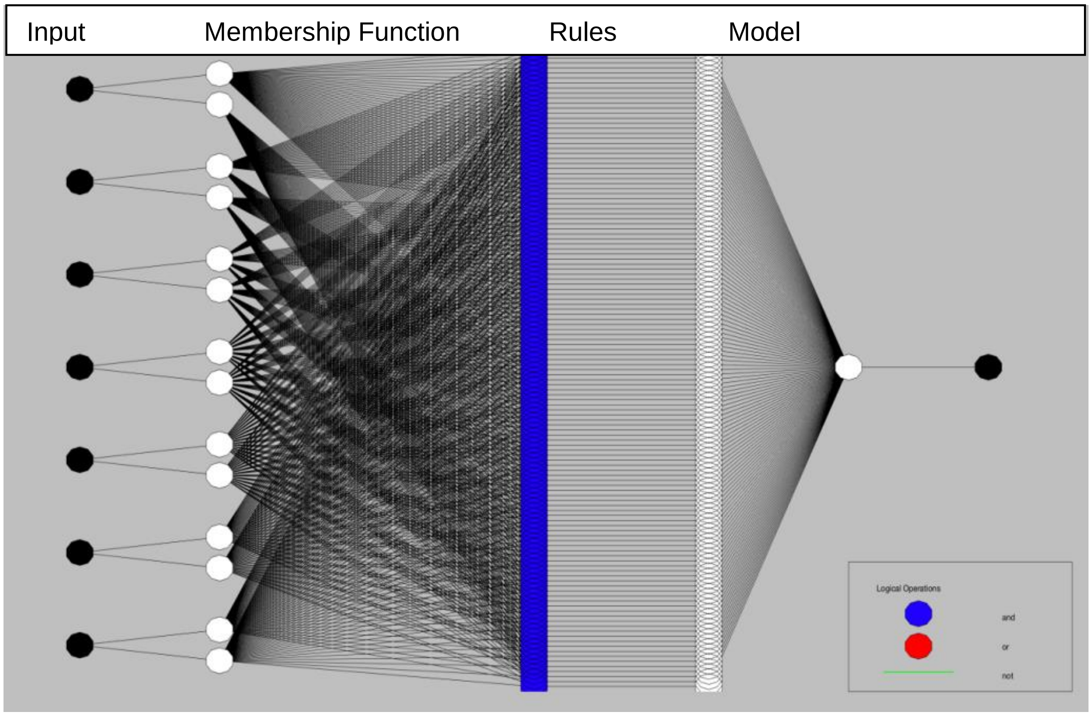


Figure 3-8 Frameworks of ANFIS

MATLB R2017

3.3.4 Construction of the ANFIS System

The ANFIS theory is used to create the model to predict the remaining useful life. MATLAB R2017 Fuzzy Logic Toolbox was used in our research. The input and output parameters are used as the training data for the prediction model. The training data has seven input parameters and one output (Figure 3-9).

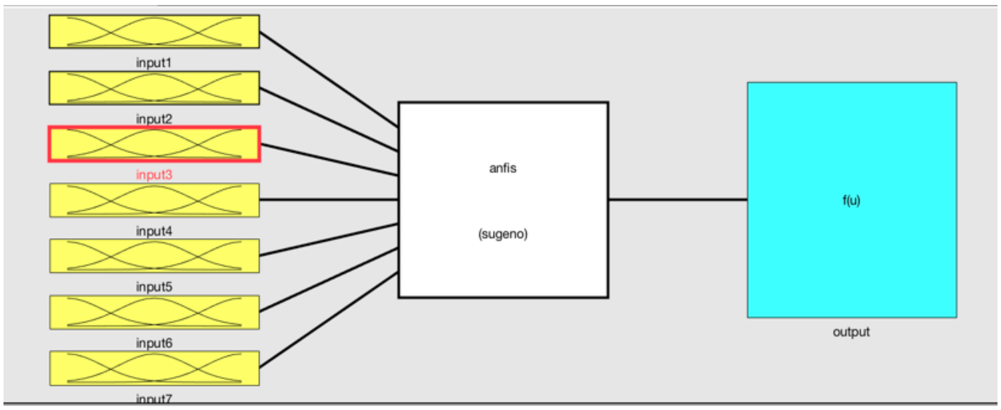


Figure 3-9 Fuzzy Rule Architecture of ANFIS Model

MATLB R2017

First, the training data are loaded into MATLAB Fuzzy Logic Toolbox, and a membership function is chosen. Different membership functions for each of the input parameters can also be chosen. The chosen membership functions are Gaussian membership function. After the model is trained using the hybrid- learning rule, the results output by different membership functions were tested against the verification data. The precision of each membership function is determined using the root-mean-square-error (RMSE). The prediction model with the smallest RMSE is the best (Lin and Huang, 2013).

3.4 Chapter Summary

The discussions in this chapter reinforced the suitability of using neural network methodology for predicting the condition of pipelines. A comprehensive list of parameters that affect the condition of the water pipes along with the modeling methodology was presented. The study develops an Artificial Neural Network model (ANN) that can make individual and group prediction depending on data availability. The model includes the potential effect of related attributes (e.g., age, length, diameter, installation year…) in addition to condition and operational attributes. The proposed methodology uses a fitting model to group water pipes based on physical attributes and their operational conditions.

Chapter 4 Data Collection and Preprocessing

Pervious chapter discussed various models of water pipe deterioration. This chapter provides the effectiveness of an ANN model that depends on the availability of reliable input data. Finding data that represents or corresponds to the possible factors reviewed was important for representing the physical cause-effect relationships. The reliability of data is measured by amount of “noise” inherent in the data (Sacluti, 1999). Noise is data patterns that contain inaccuracies and discrepancies, which does not allow the model to make proper associations between input and output patterns. Use of data with little apparent noise would result in a more accurate and precise model.

Data collection involves evaluating all available data based on accessibility, relative ease of obtaining long-term relevant data, and the prospect of future availability of the same type of data for future models. This data must have characteristics that are significant for model convergence. If all the proposed model input parameters are used for the model, the run times for model training will be exceedingly long, and hence would result in insufficient use of time. Also, if insignificant (or inappropriate) data is not eliminated initially, the redundant input parameters will be treated as “noise” by the ANN model, and as such may decrease the likelihood of the model convergence (Kulandaivel, 2004).

4.1 Data Collection

The dataset used for the development of the prediction model using artificial neural network for water pipes is described in this section. Four case studies were considered for this research. Following sections provide information on the background of case studies.

4.1.1 Southgate Water District (Colorado, Denver)

The first project is the Southgate Water District (SWD) located in Colorado, Denver. The project was selected based on information about water pipes failure and inspection dataset and included Ductile Iron (DI) and asbestos-cement (AC) pipes. Those pipes were aged and wall thickness losses have increased due to deterioration. The inspection was performed in 2016 to evaluate the condition of pipes. The Southgate Water District (SWD) is approximately 14.9 square miles and serves about 44,000 customers. The SWD owns and operates approximately 235 miles of water distribution mains, ranging in size from 4-in. to 36-in. The SWD is supplied treated water by Denver Water (Providence Infrastructure Consultants, 2016). The SWD is located south of Denver, Colorado, in Arapahoe and Douglas Counties, and is generally positioned on the west side of Interstate 25 from Belleview Avenue to Ridgegate Parkway. The SWD has about 29 miles of transmission pipelines throughout the District that function as the backbone of the water system. The SWD water system contains approximately 40% AC pipe, 34% DI pipe, 25% PVC pipe and about 1% of CI and steel pipe combined that obtained from the SWD's Geographic Information System (GIS) pipeline database. The distribution of pipe materials within the SWD (Providence Infrastructure Consultants, 2016) presents in Figure B-5 (Appendix B).

4.1.2 Laval, Moncton and Quebec cities (Canada)

Second project is in Canada. Three cities of Canada (Laval, Moncton and Quebec) are considered for data acquisition (Figure B-1 and B-3). This project was selected because they have large, typical water distribution networks and deteriorated over the years. The inspection was performed on 2006 to evaluate the condition of pipes. The total lengths of the water distribution network are 996 miles, 321 miles, 268.43 miles respectively. Iron pipes that were installed in 1954 have 95 breaks from 1987 to 2001 and 11 breaks in 1999 all the breaks occurred when the pipes are of ages between 20 and 50 years (Wang, 2006).

4.1.3 Montreal, Canada

The third project uses database related to city of Montréal and was selected for this research. The project was considered due to a large database and different types of water pipes. It was extracted from the comprehensive shape files of city of Montreal and imported to Excel worksheets. Based on the description of the data file, the database has been updated whenever a breakage happened, or a modification has been done to the network (Figure B-2). The water network of city, which consists of 125,829 data points, and covers all over the Island of Montreal with the length of 3318 miles are provided in Appendix B. The network includes cast iron pipes. Database was initially filtered, and insufficient data was removed (Zangenemadar and Moslehi, 2016).

4.1.4 Denver Water's Distribution System

The Fourth project uses Denver water's distribution system, which selected due to the mixture of old and new materials ranging from "pit" cast iron installed in the late 1800s to recently installed polyvinyl chloride (PVC). Pipe inventory and break record data were generated using the utility's geographic information system (GIS). This allowed queried data to be imported directly into Microsoft Excel, simplifying the model initialization and data entry processes. The process of importing the pipe inventory information (diameter, material, installation date, etc. was streamlined by including only the 296 miles of pipe with break histories (Rogers, 2016). This simplification also speeded up the computational time associated with linking the pipe installation information to the break records. The inspection was conducted on 2013 to evaluate the condition of pipes. There were total of 4,112 different pipes constituting the 296 miles of water mains with a total of 5,610 break records (Figure B-4). The database for this research are provided in Appendix D.

4.2 Input Parameters and Analyses

The database after preliminary preprocessing comprises the variables that are considered to have an impact on pipe condition and for training the model. The variables are Pipe material (cast iron, ductile iron, asbestos cement and steel), pipe age, pipe size (diameter), section length, number of break, installation year and wall thickness change. It was found in this study that data preprocessing is necessary for BPNN model development. This preprocessed database was then used to train, test, and validate the BPNN model.

4.2.1 Software Selection for Data Preprocessing

The original data was stored in Microsoft Access database. Microsoft Excel was selected as a data processing and analysis tool because of its adaptability for spreadsheet analysis and was used to accomplish statistical tests on the collected data. The amount of data in this study required a combined statistics software package, which could provide complete control over data access, management, analysis and presentation.

4.2.2 Data Analysis

Seven variables that affect the water pipes deterioration during their lifetime were identified. The parameters are determined to have substantial impact on water pipe deterioration. Seven variables were used in modeling process as shown in Table 4-1.

Table 4-1 Variables Used for Neural Network Modeling

Name of Variable	Description of Variable
Section Length	Length of pipe segments in feet
Size	Diameter of pipe segments in in.
Type of material	Asbestos cement, Cast Iron, Ductile iron and Steel
Age of pipe	Age of pipe grouped on a five-year period
Installation year	The year of installation is a corrosion indicator
Number of break	The number of breaks for length of pipe
Wall thickness change	Loss of wall thickness due to corrosion of pipe

Data analysis involved the analysis of all collected data as means of defining initial factor effects on the water pipe remaining useful life. Furthermore, the analysis was used as means of revealing data inconsistencies and errors. Minimum, maximum, mean, mode, standard deviation and correlation values were developed for all factors. The correlation of an input provides an indication of whether an input will correctly, or acceptably, train with a neural network.

A histogram was also developed for each input factor. The purpose of the histogram is to provide an illustration of the range and constancy of the collected data. The following Figures 4-1 to 4-8, represent the various statistical analyses performed with the data.

Figure 4-1 represents the histogram representing the different classes of Materials in the water pipe database in consideration. Most of the pipes were cast iron pipes constituting about 46% of the sample with asbestos cement pipes being the second highest number followed by ductile iron and steel pipes.

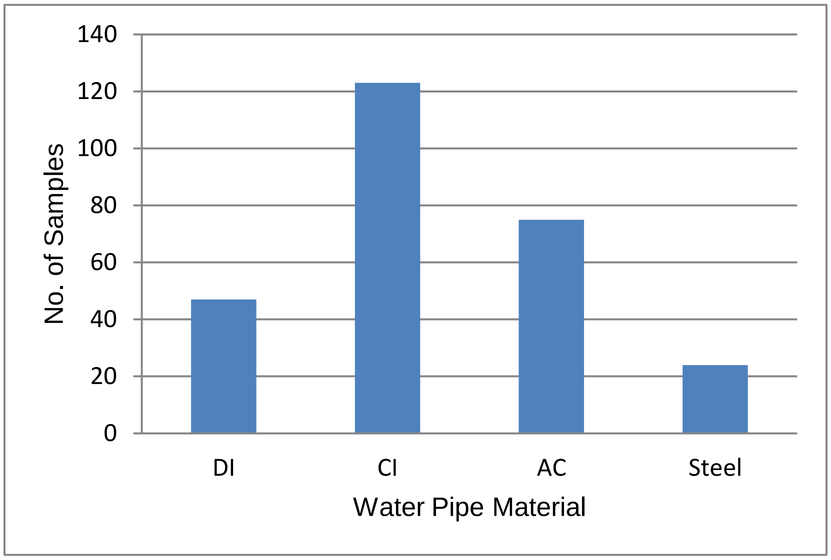


Figure 4-1 Water Material Distribution in Database

Figure 4-2 represents the distribution of the water pipes age of the samples in database. The distribution is lognormal. It can be observed that the age of pipes in this study ranges between one and 130 years. Pipe age is the difference between installation year and date of performing inspection and evaluating the results.

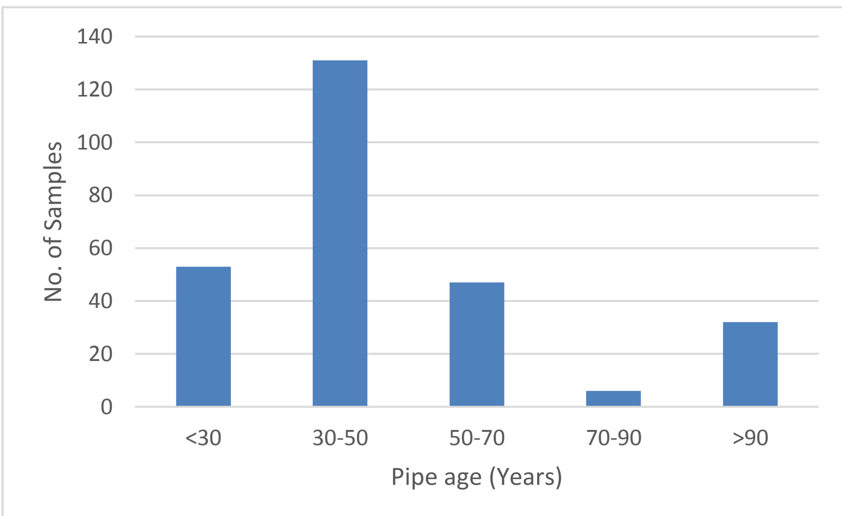


Figure 4-2 Water Pipe Age Distribution in Database

Figure 4-3 presents the water pipes diameter distribution in the database. Most of water pipes are in the 4 to 6-in. category and pipe diameter ranges are from 4 in. to 24 in.

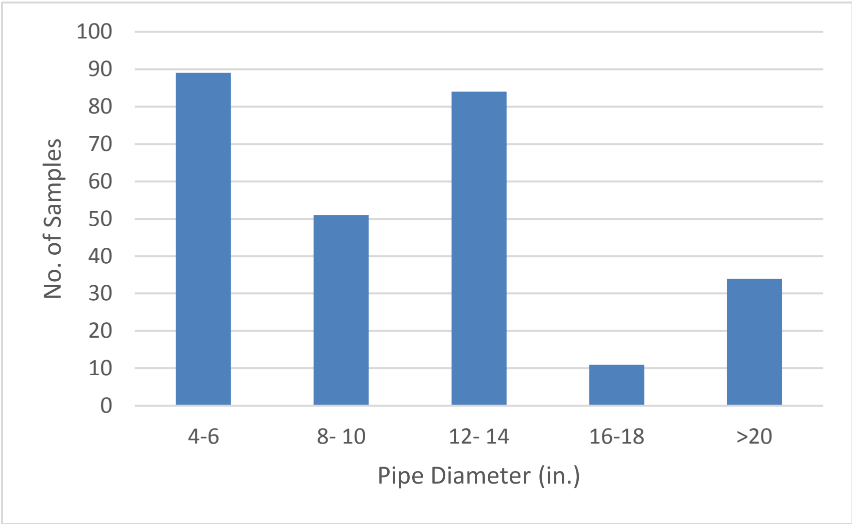


Figure 4-3 Water Pipe Size Distribution in Database

Figure 4-4 shows the distribution of pipe installation year within the group in database. The pipe installed from 1887 to 2011. Most of the pipes installed from 1960 to 1969.

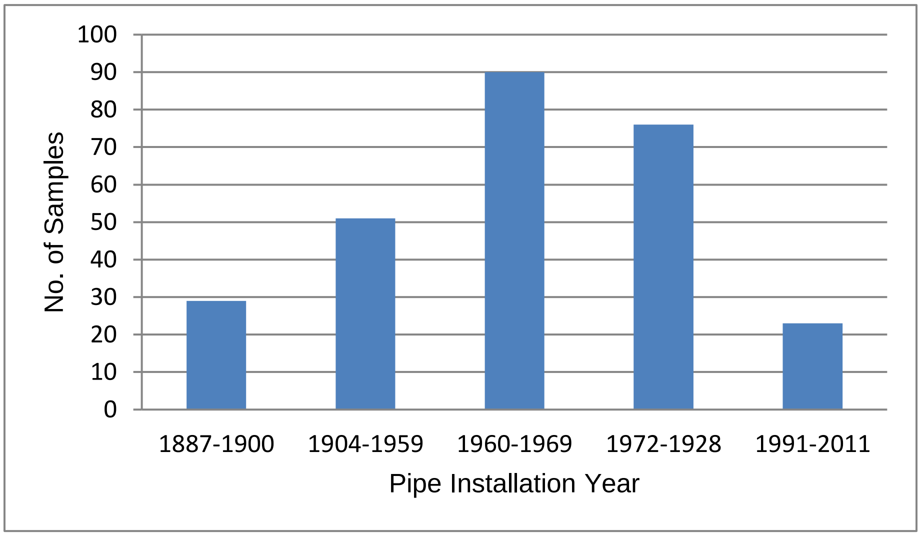


Figure 4-4 Water Pipe Installation Year Distribution in Database

Figure 4-5 shows the water pipes length distribution obtained in database indicating that majority of pipes are between 303 (feet) and 800 (feet).

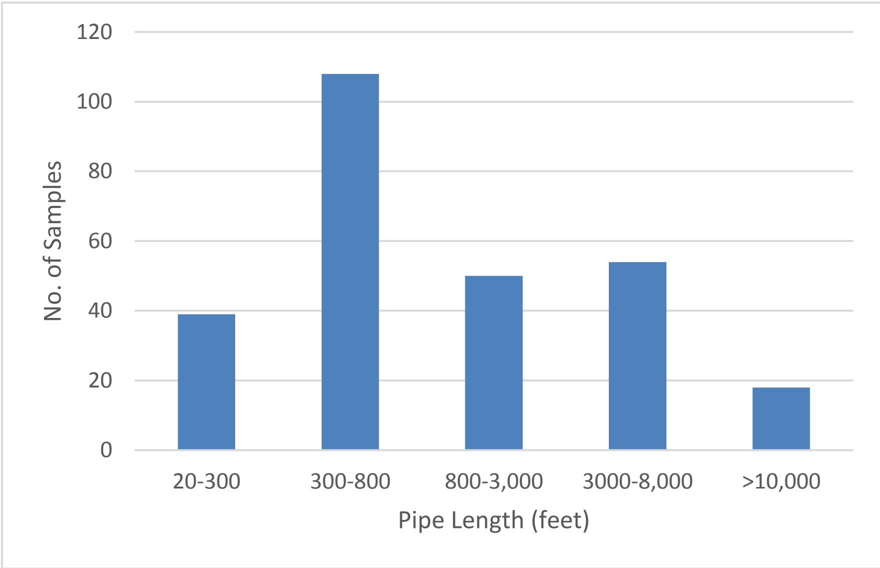


Figure 4-5 Water Pipe Length Distribution in Database

Figure 4-6 represents the water pipes number of breaks in database illustrating that most of the pipes breaks from 0 to 2 and 8 to 10 during to their lifetime.

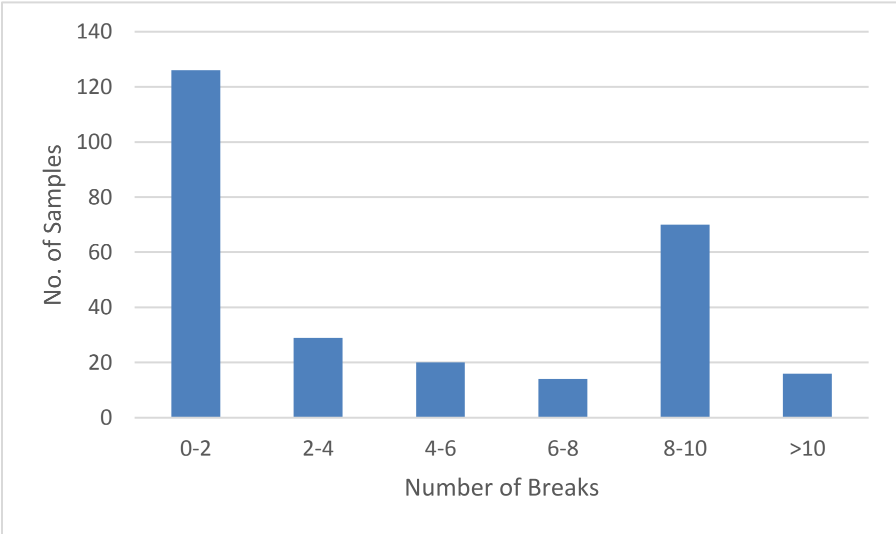


Figure 4-6 Water Pipe Number of Breaks Distribution in Database

Figure 4-7 shows pipes wall thickness losses in database illustrating that most of the water pipes have wall thickness loss around 30-40 percent.

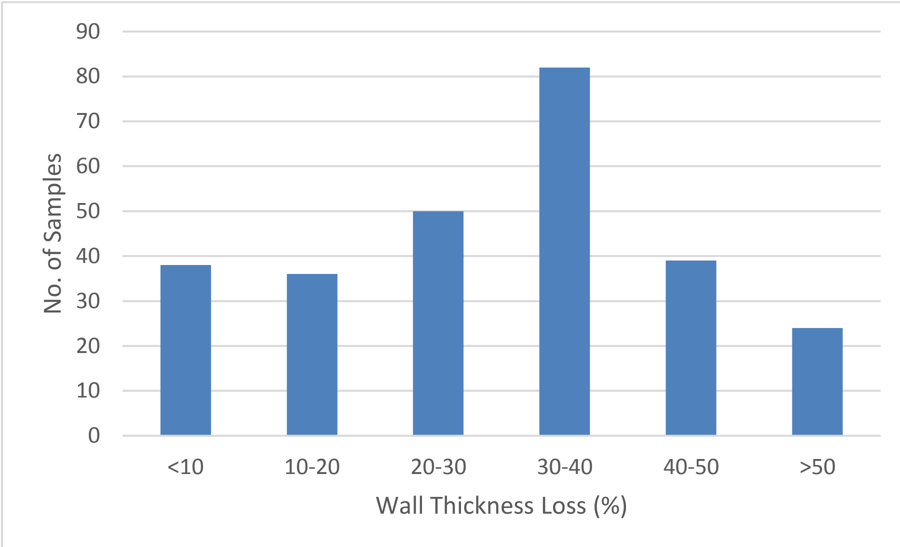


Figure 4-7 Water Pipe Wall Thickness Loss in Database

Figure 4-8 presents the relationship between pipe material and pipe dimeter, in database illustrating that most of CI, DI, AC and steel pipes are 6 in.,12 in., 8 in., and 12 in. respectively.

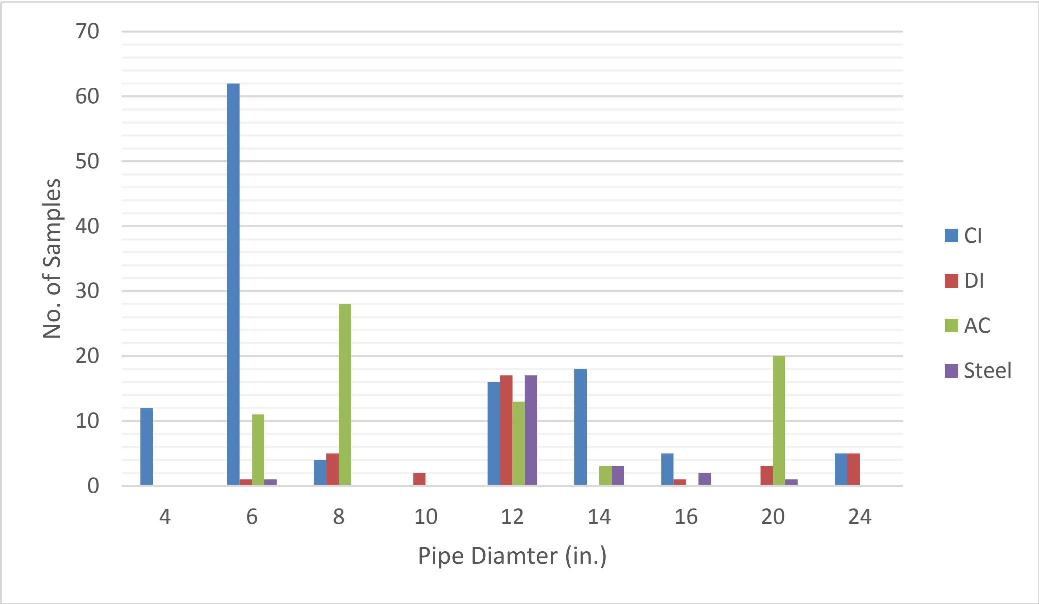


Figure 4-8 Pipe Material and Pipe Diameter in Database

4.3 Prediction of Remaining Useful Life

The remaining useful life is predicted based on the effects of each parameter (age, diameter, installation year, material, number of breaks, length and wall thickness loss). According to the literature review (section 2.3), above parameters have most impact on remaining useful life, all above parameters considered for the prediction of remaining useful life. The remaining useful life of water pipes for this dissertation is calculated based on number of years from installation to first repair and age of the pipe (Eq. 2.5 and Eq. 2.6) and categorized to different classes from below 20 years to more than fifty years. Figure 4-9 illustrates remaining useful life of the water pipes based on actual data. The results show most of remaining useful life varies from 30 to 40 years and the distribution is lognormal.

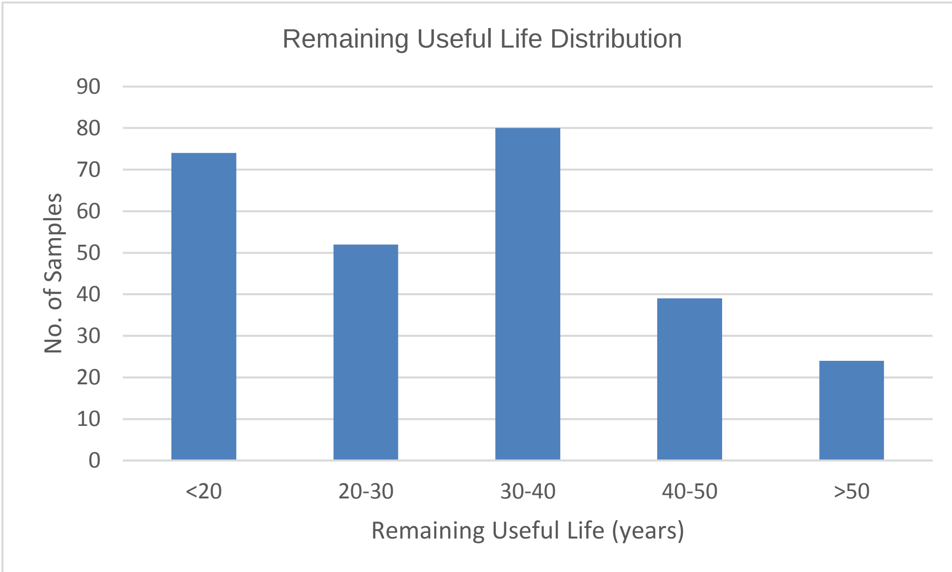


Figure 4-9 Water Pipe Remaining Useful Life Distribution in Database

4.4 Multiple Regressions and Testing of Significance

Multiple regressions and one-way analysis of variance are generated to check the correlation between each parameter (input) with remaining useful life (target). The input with a good correlation (value close to either 1 or - 1) will usually be more significant than an input with a poor correlation (value close to 0).

Figure 4-10 illustrates the relationship between wall thickness loss and remaining useful life in database. The X-axis presents the wall thickness loss (%) and the Y-axis presents the remaining useful life. The relationship between wall thickness loss and remaining useful life as observed in Figure 4-10 indicates a correlation between two. Remaining useful life has a lower value with the increase of wall thickness loss. The relationship is polynomial regression because it provides best fit for the trend with x degree of two due to higher value of R square the polynomial regression is considered.

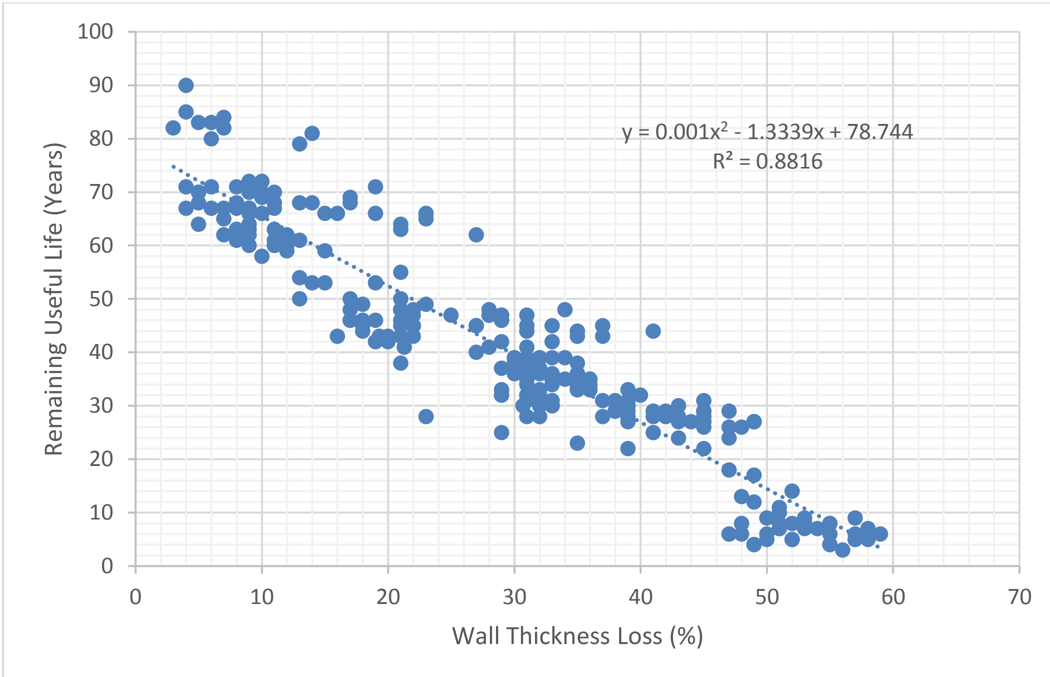


Figure 4-10 Relationship between Remaining Useful Life and Wall Thickness Loss

Figure 4-11 illustrates the relationship between the water ages and remaining useful life in database. The X-axis presents the pipe age (years) and the Y-axis presents the remaining useful life. The age of the pipes shows a pattern of increase with decreasing of remaining useful life and after age 100 has a higher value of remaining useful life. Moreover, the relationship between age and remaining useful life is polynomial regression because it provides best fit for the trend with x degree of two and high R^2 of 82%.

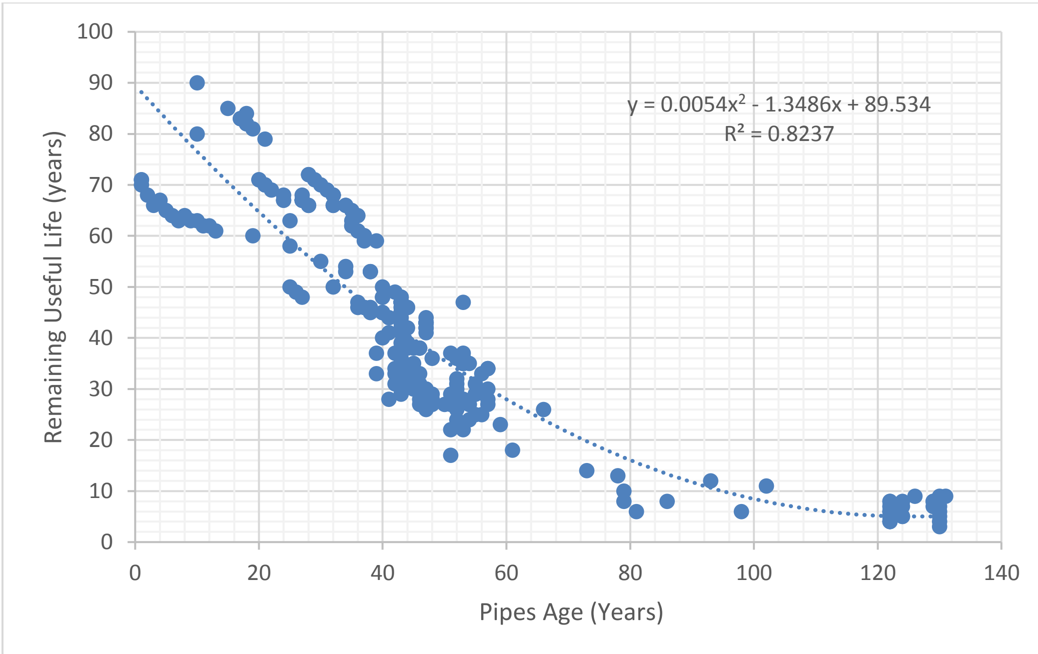


Figure 4-11 Relationship between Remaining Useful Life and Age of Pipes

The trend between age and remaining useful life presents the possibility of exponential relationship. The semi-log distribution is considered for log (RUL) to find the relationship between pipe age and log (RUL) and apply exponential equation (Eq. 4.2). Figure 4-12 illustrates the linear relationship between pipe ages and log (RUL). The X-axis presents the pipe age (years) and Y-axis presents the log (RUL).

$Y= b_0 + b_1A$ $\quad$ $Y= -0.0101A + 2.0251$

$RUL = Ce^{b_1 A}$ $\quad$ $C= e^{2.0251} = 7.58$

$RUL = 7.58e^{-0.0101A}$ $\quad$ Eq. 4.2

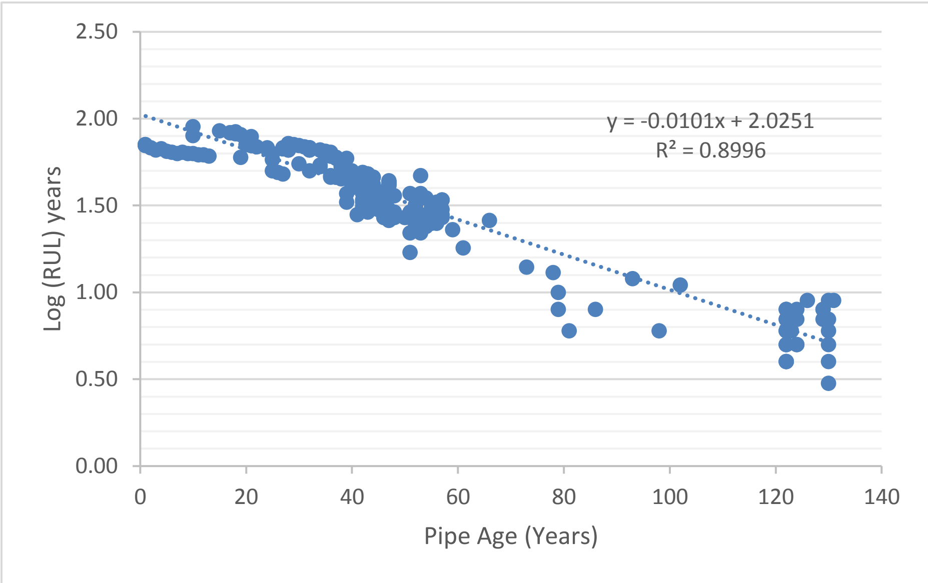


Figure 4-12 Relationship between Log (RUL) and Age of Pipes

Figure 4-13 illustrates the relationship between age and wall thickness loss in database. The X-axis presents the pipe age (years) and the Y-axis presents the wall thickness loss. Wall thickness loss has a higher value with the increase of age of pipes. The relationship is considered the polynomial regression due to higher value of R^2 the polynomial regression is considered because it provides best fit for the trend with x degree of two and high R^2 of 78%.

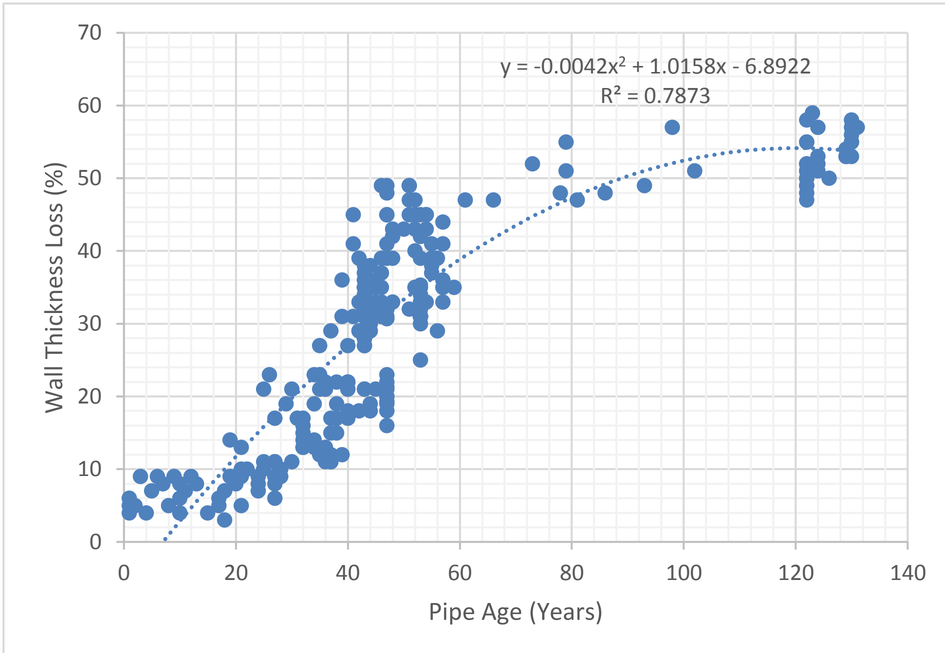


Figure 4-13 Relationship between Wall Thickness Loss and Age of Pipes

4.4.1 One-way ANOVA Analysis

There is no substantial correlation seen with the remaining useful life and the other parameters. Analysis of variance and t test are generated to determine the statistical significance of the other input variables (Montgomery, 1994). The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of Remaining Useful Life (RUL) for three or more independent (unrelated) groups. When the model passes both ANOVA and t test, it is statistically significant (P value <0.05), which means that the dependent variable (response) and the independent variables (predictors) have a significant relationship. Moreover, the null hypothesis is rejected, and it is concluded that population means of RUL are not the same. Furthermore, the f-value shows how far the results are from the hypothesis. The ANOVA test results are listed below:

- Pipe Material: The f-ratio value is 34.07805. The p-value is < .001. The result is significant at $p < .05$.
- Pipe Length: The f-ratio value is 5.60639. The p-value is < .00024. The result is significant at $p < .05$.
- Pipe Installation year: The f-ratio value is 3.9938. The p-value is 0.03656. The result is significant at $p < .05$.
- Pipe Diameter: The f-ratio value is 3.9938. The p-value is 0.03656. The result is significant at $p < .05$.
- Pipe Number of Breaks: The f-ratio value is 6.63956. The p-value is < 0.00042. The result is significant at $p < .05$.

4.5 Statistical Analysis

The primary basis of what is presented in this research lies in the fundamentals of statistics. The population mean, μ, provides the average of the group characteristic (i.e. average Remaining Useful Life) (Daly et al., 2016). The variance, S^2, indicates how far data points are spread out from one another. A variance of zero specifies the values are all identical. Smaller variances indicate the data points are closer to the mean while larger variances indicate the data is more spread out. The sample standard of deviation, $\boldsymbol{\sigma}$, is simply the square root of the variance and provides a statistic that is expressed in the same units as the mean. This statistic provides a normalized value for statistical tests.

Confidence interval shows the range of values that are considered within the accuracy of the measurement. The confidence level, e.g. 95%, must be known to define this interval. A confidence level of 95% will involve that any sampling results will fall between the confidence interval bounds with 95% confidence. The statistical analysis of this data set is provided in this section, in which the pipe material (M), diameter (D), length

(L), number of break (B), installation year (Y), wall thickness change (W), and Age (A) represent the selected input parameters, and the remaining useful life (RUL) represents the output parameter that is estimated from pipe age (Table 4-2). These parameters were considered for the ANN models. The data set was randomly selected which were used for training, validation, and testing the results, respectively. Pipe material classified as four main groups of asbestos cement, cast iron, ductile iron, and steel pipes.

4.5.1. Mean and Standard Deviation for Population

For this study, the interval estimation is used to find amount of Mean (M) and Standard deviation (σ) for population using 95% level of confidence (Eq. 4.3).

α = 0.05 $\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ Eq. 4.3

Example for age:

$\bar{x}$ = 49.78

N= 269, $Z_{\alpha/2,}$ **= 1.96** (Montgomery, 1994)**, σ = 30.31. So, 46.15**$\leq \mu \leq 53.40$**,**

Or 46 $\leq \mu \leq 53$

Table 4-2 Statistics of Water Pipe Database Used in this Dissertation

Variables	Minimum	Maximum	Mean ($\overline{x}$)	Standard deviation (α)	Skewness	Mode
Age (Years)	1	131	49.78	30.31	1.45	43
Diameter (in.)	4	24	10.66	5.13	0.94	6
Length (ft)	20.5	36,161.4	2870.51	5008.58	3.55	5280
Material	6.14	8.35	6.146	2.75	-0.89	8.35
Number of Break	0	95	5.09	7.74	6.22	6
Installation year	1887	2011	1961.15	28.78	-1.25	1969
Wall Thick loss (%)	1	59	29.64	14.81	-0.11	33
RUL (Years)	3	90	40.65	20.46	0.07	36

Population variance (σ^2) can be computes using Eq. 4.4:

$$\frac{(n-1)S^2}{\chi^2\alpha/2} \ll \boldsymbol{\sigma}^2 \ll \frac{(n-1)S^2}{\chi^2 1-\alpha/2} \quad \text{(Two-sided Interval)} \qquad \text{Eq. 4.4}$$

Where: S= Sample Standard Deviation, $\chi^2\alpha/2$ = value of a random variable having a chi-square distribution.

$\chi^2_{\ 0.025,\ 268}$ =295.689, $\chi^2_{\ 0.975,\ 268}$ = 208.098 (Montgomery, 1994)

822.80$\leq \boldsymbol{\sigma}^2 \leq 1169.13$ 28.68 $\leq \boldsymbol{\sigma} \leq 34.19$ $\mathbf{29 \leq \sigma \leq 34}$

It concluded that based on Table 4-7, if assumed that the average age of pipe is 50 years, and standard deviation is approximately 30. Approximately 68% of the pipes have the age between 20 and 80 years because the amount of Z is between -1 and 1 and area of $-1 \leq$ Z≤ 1 is 68% (Montgomery, 1994). It is good to inspect before age 20 in conservative decisions. In addition, the latest time for inspection would be in age of 80. The same processes have been done for remaining useful life**.** The result for this database shows if average remaining useful life is 40 years and standard deviation is 20, approximately 68% of pipes have a remaining useful life between 60 and 20 years, because the area of $-1 \leq Z \leq 1$ is around 68%.

$\mathbf{38.20 \leq \boldsymbol{\mu} \leq 43.09}$ $\mathbf{38 \leq \boldsymbol{\mu} \leq 43}$

$379.41 \leq \boldsymbol{\sigma}^2 \leq 539.11$ $19.47 \leq \boldsymbol{\sigma} \leq 23.21$ $\mathbf{19 \leq \boldsymbol{\sigma} \leq 23}$

4.5.2 Goodness-of-fit Test

The goodness-of-fit test applies to situations in which we want to determine whether a set of data may be looks upon as a random sample from a population having a given distribution (Montgomery, 1994). The degree of freedom, number of class interval and theoretical frequency are provided in Eq. 4.5, 4.6 and 4.7 respectively. Chi- square good ness of fit test is considered in this section with 95% level of confidence and lognormal distribution (Table 4-3).

Degree of freedom = # class interval - # parameters -1: 9-2-1 =6 Eq. 4.5

Number of class Intervals: $1+3.3 \log_{10} N$ = $1+ 3.3 \log_{10} 269$ = 9.01 ~ 9 Eq. 4.6

e_i (Theoretical Frequency) = N × (Prob. of the interval) = $269 \times \frac{1}{9}$ = 29.88 Eq. 4.7

O_i = Observed Frequency

Table 4-3 Data of Age for Chi-square Goodness-of-fit Test

Class Interval	O_i	e_i	$\frac{(Oi-ei)^2}{ei}$
≤ 20	25	29.88	0.80
20-30	26	29.88	0.50
30-40	35	29.88	0.88
40-50	98	29.88	155.30
50-60	45	29.88	7.65
60-70	2	29.88	26.01
70-80	4	29.88	22.42
80-90	2	29.88	26.01
≥ 90	32	29.88	0.15

The data presents lognormal distribution, so the probability of failure is calculated based on Eq. 4.6 and Eq. 4.7 (Table 4-4):

Probability of failure if age is ≤ 20 and ≤ 90, σ_x =30.31, μ_x = 49.78

$\sigma^2_y = \ln(1 + (\frac{\sigma x}{\mu x})^2)$, σ^2_y =0.315 σ_y = 0.561 Eq. 4.6

$\mu_y = \ln \mu x - \frac{1}{2}\sigma^2 y$, μy = 3.75 Eq. 4.7

P ($x \leq 20$) = P ($y \leq 2.995$) , P ($Z \leq \frac{2.995-3.75}{0.561}$), P ~ 0.091, P ~ 9 %

P ($x \leq 90$) = P ($y \leq 4.499$) = P ($Z \leq \frac{4.499-3.75}{0.561}$), P ~ 0.9082, P ~ 91%

P ($x \geq 90$) = P ($y \geq 4.499$) = P ($Z \geq \frac{4.499-3.75}{0.561}$), P ~ 0.092, P ~ 9%

4.5.3 Conditional Probability

The conditional probability of an event B is the probability that the event will occur given the knowledge that an event A has already occurred. This probability is written P

(B|A), notation for the probability of B given A. In the case where events A and B are independent (where event A has no effect on the probability of event B), the conditional probability of event B given event A is simply the probability of event B, that is P(B).

Using Conditional probability considering age and remaining useful life, we have:

P (RUL$|$Age≤ 20) = $\frac{P(RUL\ and\ age \leq 20\ years)}{P(\text{x} \leq 20)}$ = 0.15 ~ 15%

P (RUL$|$Age≥ 90) = $\frac{P(RUL\ and\ age \geq 90\ years)}{P(\text{x} \geq 90)}$ = 0.003 ~ 0.3%

Table 4-4 Data of Remaining Useful Life for Chi-square Goodness-of-fit Test

Class Interval	O_i	e_i	$\frac{(\text{Oi} - \text{ei})^2}{ei}$
≤ 10	34	29.88	0.57
10-20	6	29.88	19.08
20-30	36	29.88	1.25
30-40	64	29.88	38.96
40-50	52	29.88	16.38
50-60	15	29.88	7.41
60-70	40	29.88	3.43
70-80	13	29.88	9.54
≥ 80	9	29.88	14.59

The same process has been done for Probability of failure considering lognormal distribution and if remaining useful life is ≤ 10 and ≥ 80:

σx = 20.46, μ_x = 40.65

$\sigma^2_y = \ln(1 + (\frac{\sigma x}{\mu x})^2)$, σ^2_y = 0.225 σy = 0.474

$\mu_y = \ln \mu x - \frac{1}{2}\,\sigma^2 \text{y}$, μ_y = 3.592

$P(x \leq 10) = P(y \leq 2.302) = P(Z \leq \frac{2.302-3.592}{0.474})$, P~ 0.958, P ~ 95%

$P(x \geq 80) = P(y \geq 4.382) = P(Z \geq \frac{4.382-3.592}{0.474})$, P ~,0.05, P ~ 5%

The results show probabilities of water pipe failure for pipes having age ≤ 20 and ≤ 90 and are approximately 9% and 91%. Moreover, the probabilities of pipes failure having remaining useful life ≤ 10 and ≥ 80 are 95% and 5% respectively.

4.6 Chapter Summary

This chapter presented a detailed discussion about the data preprocessing and analysis. The raw database was transformed into a standardized format is ready for neural network model development. The available parameters for the model development were identified and their relevance examined through statistical analysis. Although certain variables have a low amount of correlation to the condition of water pipes and deterioration, it is important to include such parameters as neural network can capture even subtle relationships.

Chapter 5 Model Development

The previous chapter described the preprocessing and statistical analysis of the acquired data. This chapter deals with the detailed account of the model development process. The model development in this dissertation includes training and testing. The following sections provide a detailed discussion about the model development process. As described in Chapter 4, Training a neural network involves repeatedly presenting a set of examples (facts) to the network. The network takes each input, makes a guess as to the output, checks this guess against the output (correct answer), and makes corrections to the initial connections (weights) if its guess is incorrect. This process is repeated for each fact in turn until the network learns the facts well enough to be useful.

5.1 Artificial Neural Network (ANN) Results and Analysis

5.1.1. Histograms of Errors

The ANN models were developed using MATLAB R2017. The histogram of the errors and performance graphs are presented in Appendix C. Histogram of errors gives a quick snapshot of the distribution of errors, making it easy to see how close the network to attaining the pre-specified tolerance level. The X-axis presents the errors that calculated based on difference between targets and outputs. The Y-axis presents the number of instances. Moreover, the blue bars represent training data, the green bars represent validation data, and the red bars represent testing data. The histogram provides an indication of outliers, which are data points where the fit is significantly worse than most of data. The outliers determine if the data is bad, or if those data points are different from the rest of the data set. If the outliers are valid data points, but are unlike the rest of the data, then the network is extrapolating for these points. The number of bins show the range of numerical value.

5.1.2. Performance Charts

Performance chart presents the best validation performance compared to testing performance and the progress of the mean square error, during training. The training was stopped when the neural network settled at the lowest possible training and testing error and there was no observable exposure in the training statistics. The X-axis presents the number of epochs. An epoch is a measure of the number of times all the training vectors are used once to update the weights. The Y-axis presents the mean square error. The results are reasonable if the mean-square error is small, the test error and the validation error have similar characteristics and no significant overfitting has occurred. The performance charts of ANN models are provided in Appendix C.

The summary of training, validation and testing statistics is given in Table 5-1 through 5-3. Eight ANN models were developed for the training, validation and testing. The mean absolute error, relative absolute error, root relative square error and mean absolute percentage error are calculated based on Eq. 3.8 through Eq. 3.12 (section 3.2.7). Figures 5-1 through 5-5 present the relationship between ANN models and mean absolute error (MAE), relative absolute error (RAE), root relative square error (RRSE), mean absolute percentage error (MAPE) and R^2.

Table 5-1 Training Indices of Various ANN models Used to Predict the RUL

Model	Training				
	R^2 (%)	RAE	RRSE	MAPE	MAE
1	98.1	0.007	0.012	1.076	0.170
2	97.8	0.031	0.053	3.724	0.700
3	93.7	0.212	0.249	20.678	4.710
4	96.2	0.119	0.147	10.665	2.650
5	97.3	0.039	0.053	4.531	0.870
6	89.5	0.294	0.325	29.998	6.500
7	98.2	0.016	0.037	2.243	0.370
8	97.2	0.020	0.041	3.171	0.450

Table 5-2 Validation Indices of Various ANN models Used to Predict the RUL

Model	Validation				
	R^2	RAE	RRSE	MAPE	MAE
1	98.3	0.000	0.001	8.047	1.304
2	96.1	0.001	0.003	27.866	5.240
3	94.7	0.011	0.015	154.706	35.230
4	95.2	0.006	0.009	79.791	19.830
5	96.3	0.002	0.003	33.902	6.520
6	90.1	0.016	0.020	224.435	48.640
7	97	0.002	0.002	16.780	2.790
8	97.2	0.001	0.002	23.725	3.360

Table 5-3 Testing Indices of Various ANN models Used to Predict the RUL

Model	Testing				
	R^2	RAE	RRSE	MAPE	MAE
1	98	0.007	0.001	5.431	0.880
2	97.3	0.002	0.005	18.810	3.540
3	92.1	0.018	0.024	104.426	23.780
4	94.8	0.010	0.014	53.859	13.390
5	95.9	0.003	0.005	22.884	4.400
6	91.8	0.025	0.031	151.493	32.830
7	97.4	0.002	0.004	11.330	1.88
8	96.8	0.001	0.004	16.014	2.270

Figure 5-1 presents the Relative Absolute Error (RAE) of eight ANN models. The X-axis presents the eight ANN models and the Y-axis presents the relative absolute error. The minimum value of RAE in training, validation and testing are 0.007, 0 and 0.007 respectively. The results show models one and seven have the lowest value of relative absolute error.

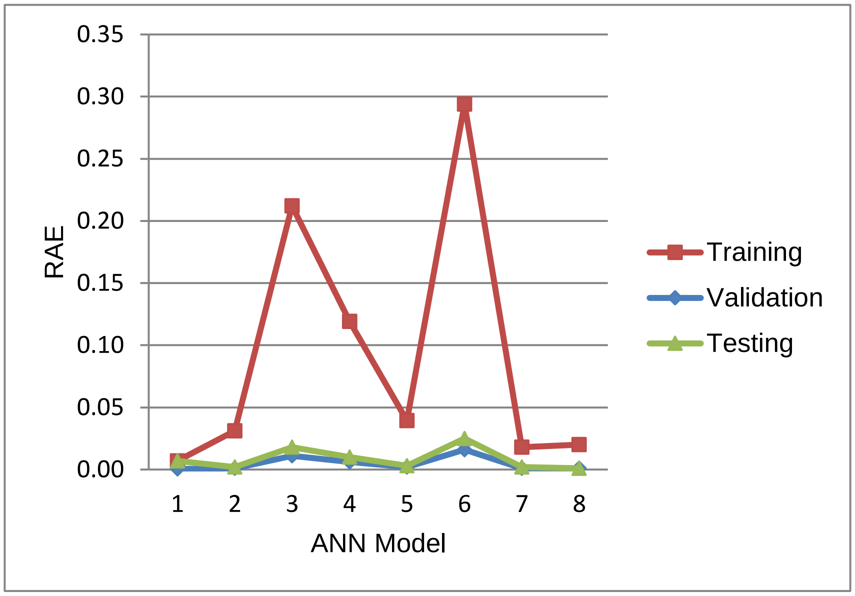


Figure 5-1 Relative Absolute Error

Figure 5-2 presents Root Relative Square Error (RRSE) of eight ANN models. The X-axis presents the eight ANN models and the Y-axis presents the root relative square error. The minimum value of RRSE in training, validation and testing are 0.012, 0.001 and 0.001 respectively. The results show models one and seven have the lowest value of root relative square error.

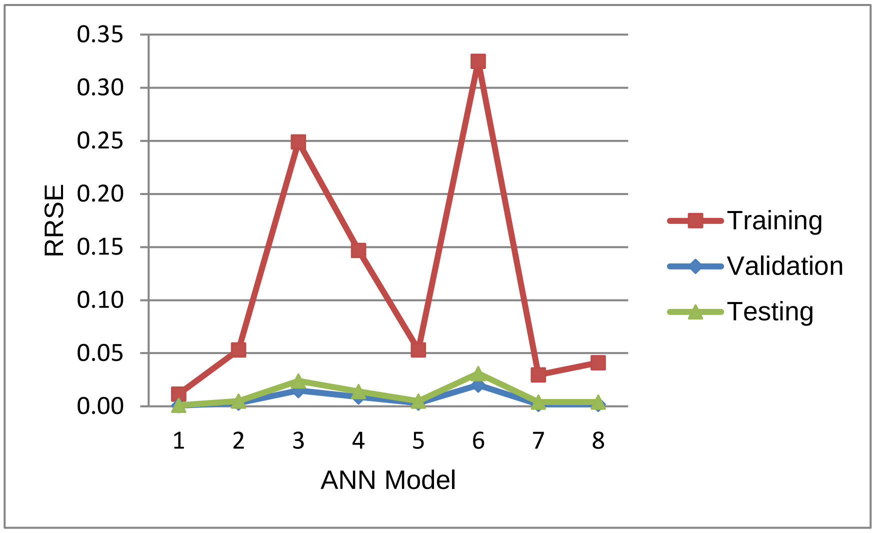


Figure 5-2 Root Relative Square Error

Figure 5-3 presents the Mean Absolute Percentage Error (MAPE) of eight ANN models. The X-axis presents the eight ANN models and the Y-axis presents the mean absolute percentage error. The minimum value of MAPE in training, validation and testing are 1.076, 8.047 and 5.431 respectively. The results show models one and seven have the lowest value of mean absolute percentage error.

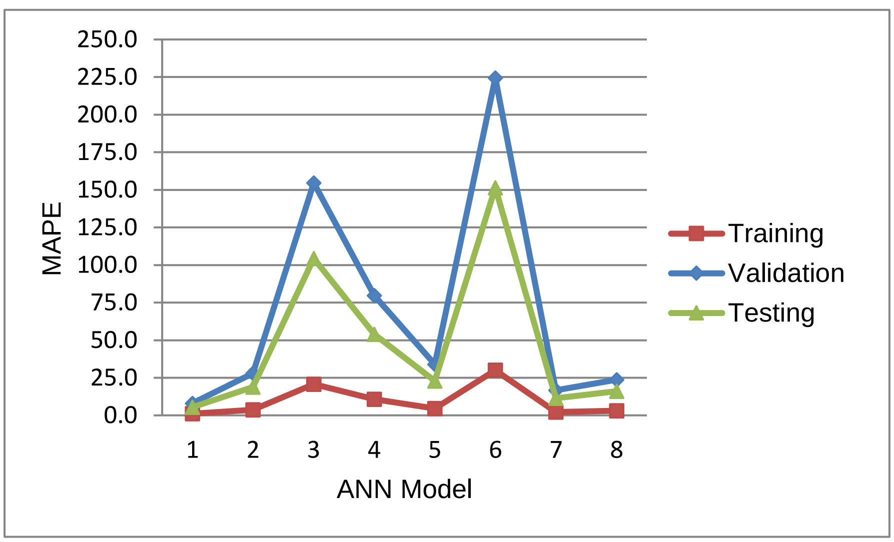


Figure 5-3 Mean Absolute Percentage Error

Figure 5-4 illustrates the Mean Absolute Error (MAE) of eight ANN models. The X-axis presents the eight ANN models and the Y-axis presents the mean absolute error. The minimum value of MAE in training, validation and testing are 0.17, 1.304 and 0.88 respectively. The results show models one and seven have the lowest value of mean absolute error.

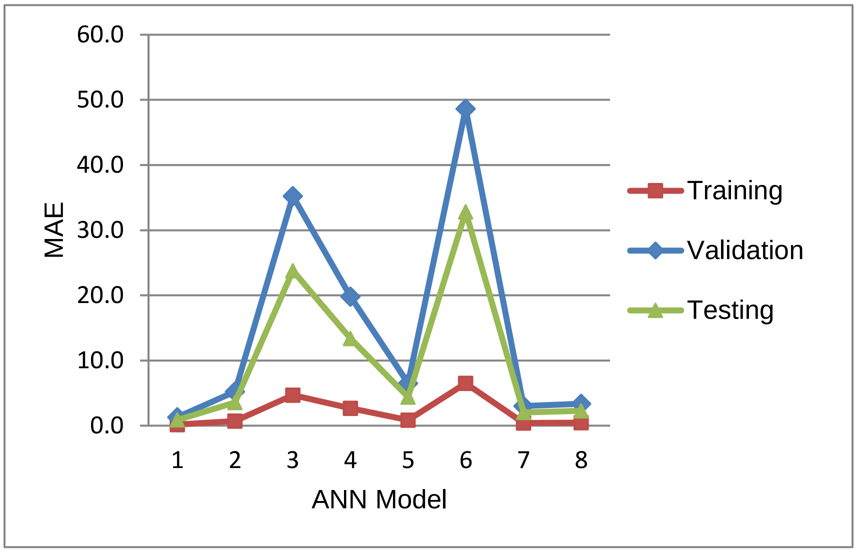


Figure 5-4 Mean Absolute Error

Figure 5-5 shows the calculated value of coefficients of determination (R^2) for training, testing, and validating phases for all ANN models. The X-axis presents the eight ANN models and the Y-axis presents the R^2. The coefficients of determination, in all cases, appear to be higher than 90%. Therefore, it is suggested that all the models can be used suitably for predicting the remaining useful life.

It also demonstrates that ANN1 and ANN7 have the highest R^2 in comparison to other models. They are 98.1, 98.3, and 98% for the training, validation, and testing phases for ANN1, respectively and for ANN7, they are 98.2%, 97%, and 97.4% for the training, validation, and testing phases, respectively. However, the differences between ANN1 and

ANN2, ANN5, ANN7, ANN8 are less than 2% and considered negligible. They could all be considered the highest performance model.

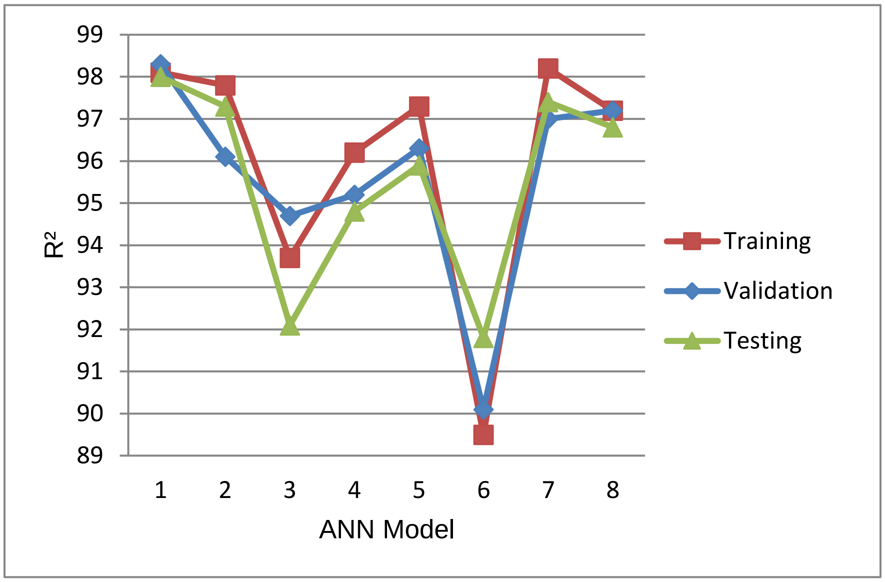


Figure 5-5 R^2 Values of ANN Models

5.1.1 Analysis of Results

The minimum values of MAE, RRSE, MAPE, and RAE in the training phase are 0.17, 0.012, 1.076, and 0.007, respectively, and are all associated with ANN1. The minimum values of MAE, RRSE, MAPE, and RAE in the validation phase are 1.304, 0.001, 8.047, and 0, respectively, and are all associated with ANN1. Finally, the minimum values of MAE, RRSE, MAPE, and RAE in the testing phase are 0.88, 0.001, 5.431, and 0.007, respectively, and are all associated with ANN1. Comparing the results, all the proposed models can predict the remaining useful life values well. However, because the remaining useful life should be determined with the highest accuracy, it is recommended to assess the results of the ANN models in more detail and to identify the best model even though the differences in performance are negligible. In general, the highest R^2 and lowest error are the indicators of best performance for an ANN model. However, the accuracy of testing and validating phases decreases by increasing the R^2 in the training phase of complex models. This phenomenon is called overfitting and must be avoided in ANN modeling. In this study, the R^2 in training, testing, and validating phases are very close except in ANN6 and ANN3, which have differences less than 4%. Consequently, overfitting was not used in any of the nine proposed models.

The ANN6 model had the lowest R^2 in the training, validation and testing phase. The trend performs to be the same for all the phases, and the differences are insignificant. The RRSE values are approximately equal in all ANN models in the three phases except in ANN6. The RRSE values are higher in the ANN6 model in training, validation and testing, which proves the previous assumption that this model is the least accurate one. The MAPE values are higher in validation rather than testing and training; however, the differences are more than 10%. Because R^2 is the highest and MAE, RAE, RRSE, and MAPE are the least for the ANN1 and ANN7 model, it appears to be the most precise models. Moreover, the MAPE value is less than 10, which categorizes this forecasting as a high-accuracy prediction. Figure 5-6 present the R^2 for the total samples of ANN models. The R^2 are higher in ANN1 and ANN7.

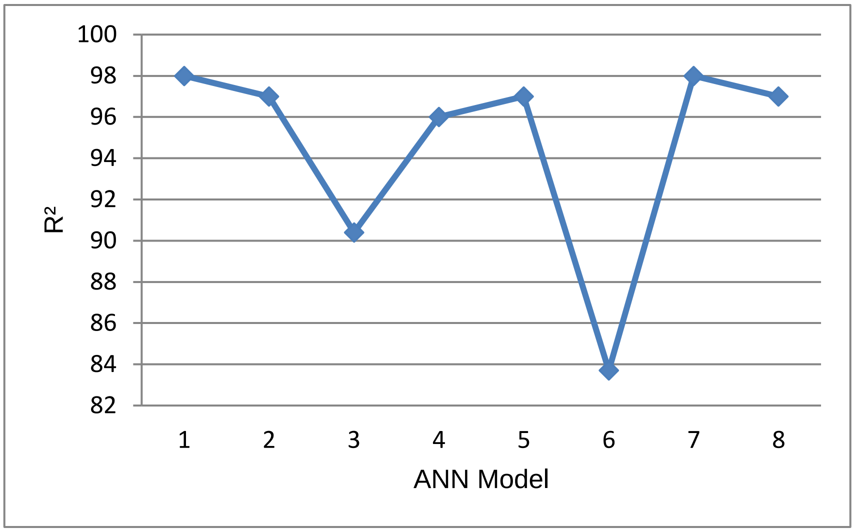


Figure 5-6 R^2 Values of ANN Models for Total Samples

Figure 5-7 presents the error results for total samples of ANN models. The X-axis presents the eight ANN models and the Y-axis presents the all error results (RRSE, MAPE, RAE and MAE. The results show ANN1 and ANN7 have the least error compared to the other ANN models.

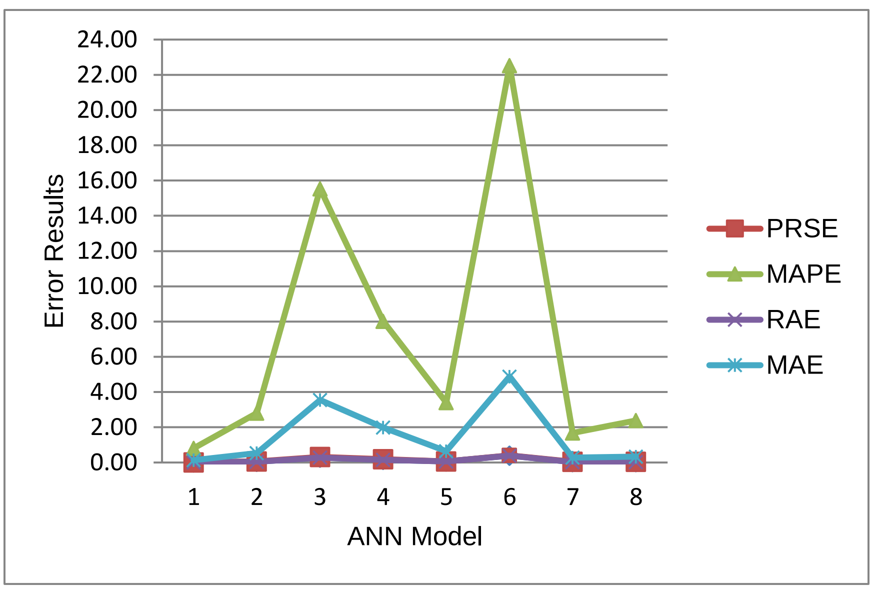


Figure 5-7 Error Results for Total Samples

The comparison of predicted results and estimated results are shown in Figure 5-8 for the best ANN model. The X-axis presents the estimated RUL and the Y-axis presents the predicted RUL. Estimated RUL is calculated based on actual data and predicted RUL is based on ANN results for best model. Most of the results fall in the near area of y=0.9112x+3.7663. The coefficient of determination is 89%, which indicates that the proposed models have predicted the remaining useful life of the pipeline exactly, and is reliable for further analysis of the network. The model's precision can be increased, as more input parameters identified to affect the water condition are available for modeling.

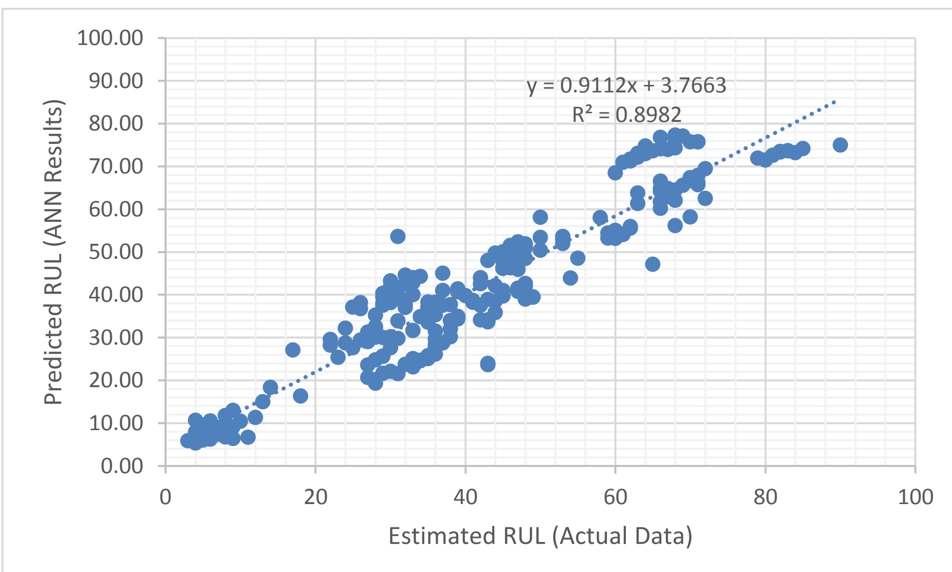
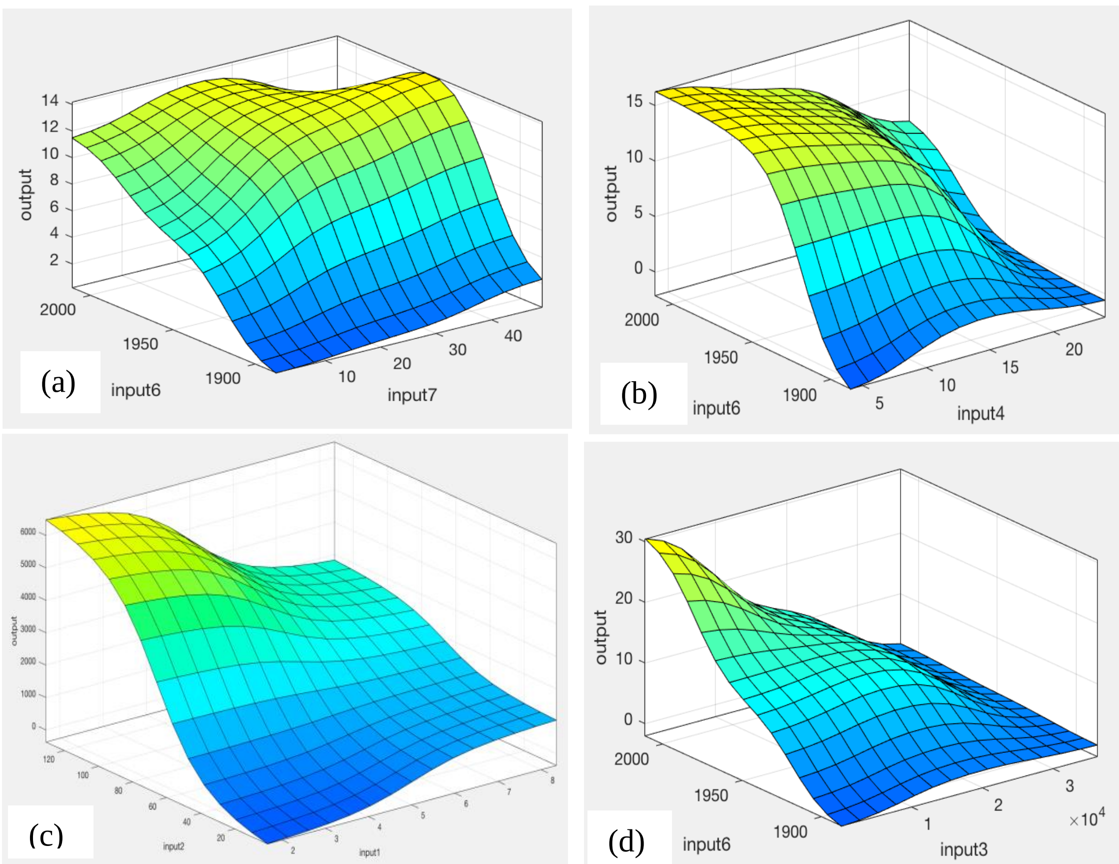


Figure 5-8 Predicted Results versus Estimated Results

5.2 ANFIS Results

Figure 5-9 shows the impacts of the installation year (input 6) and wall thickness loss (input 7) (a), Diameter (input 4) and installation year (b), material (input 1) and age (input 2) (c), length (input 3) and installation year (d), age and installation year, wall thickness loss and number of break (input 5) on the remaining useful life in database. As shown in the figure, the smaller Input parameters can reduce output angle. The slopes of material, diameter and length are smaller than the other parameters Figures b, d and e). The slopes of wall thickness, age and installation year are higher than the other parameters (Figures a, c and f). Therefore, those parameters have the most impact on the remaining useful life.



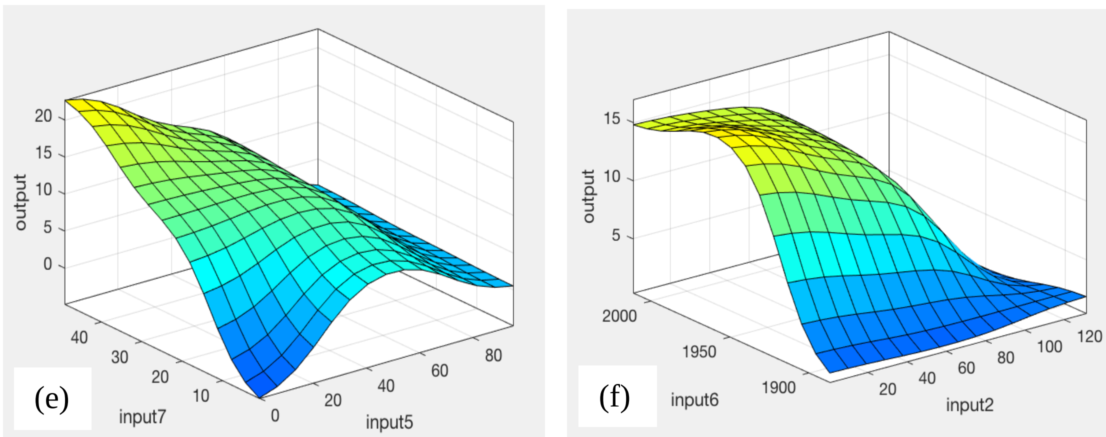


Figure 5-9 Contour Surface for Relationship between Input Parameters and Output

MATLAB R2017

5.2.1. ANFIS Training Error

The training error chart presents the number of errors versus the number of epochs in ANFIS. After loading the training data and generating the initial ANFIS structure, the ANFIS is trained using hybrid as an optimization method. The optimization methods train the membership function parameters to emulate the training data.

Figure 5-10 illustrates the error decreasing with increasing the number of epochs. The result of MATLAB is presented on Appendix C (Figure C-9).

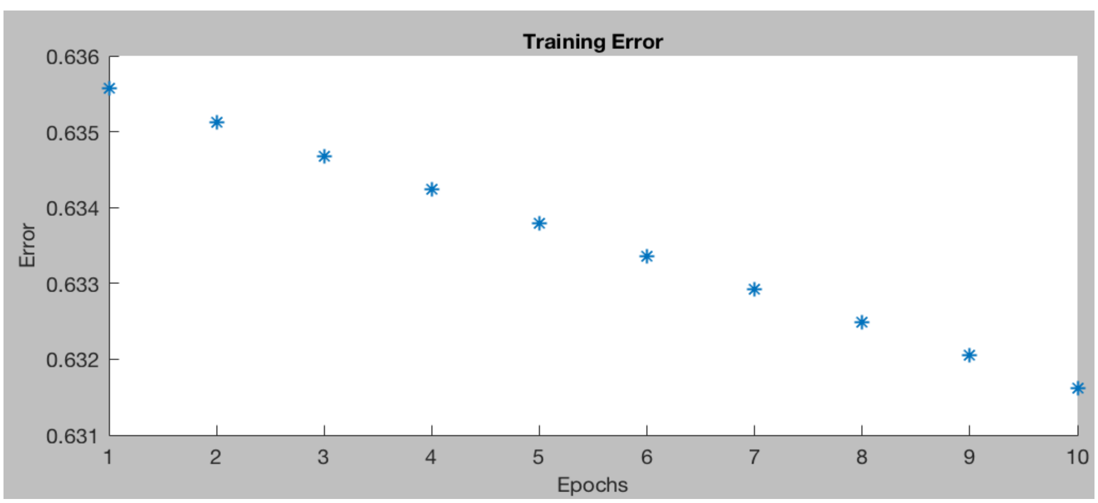


Figure 5-10 Training Error

5.2.2. ANFIS Training Data

The next step is to check if the training data matches with ANFIS output. The best match between training data and ANFIS output presents the high accuracy of the ANFIS model. Figure 5-11 presents training data matches (input data) with ANFIS output. The result of MATLAB is presented on Appendix C (Figure C-10).

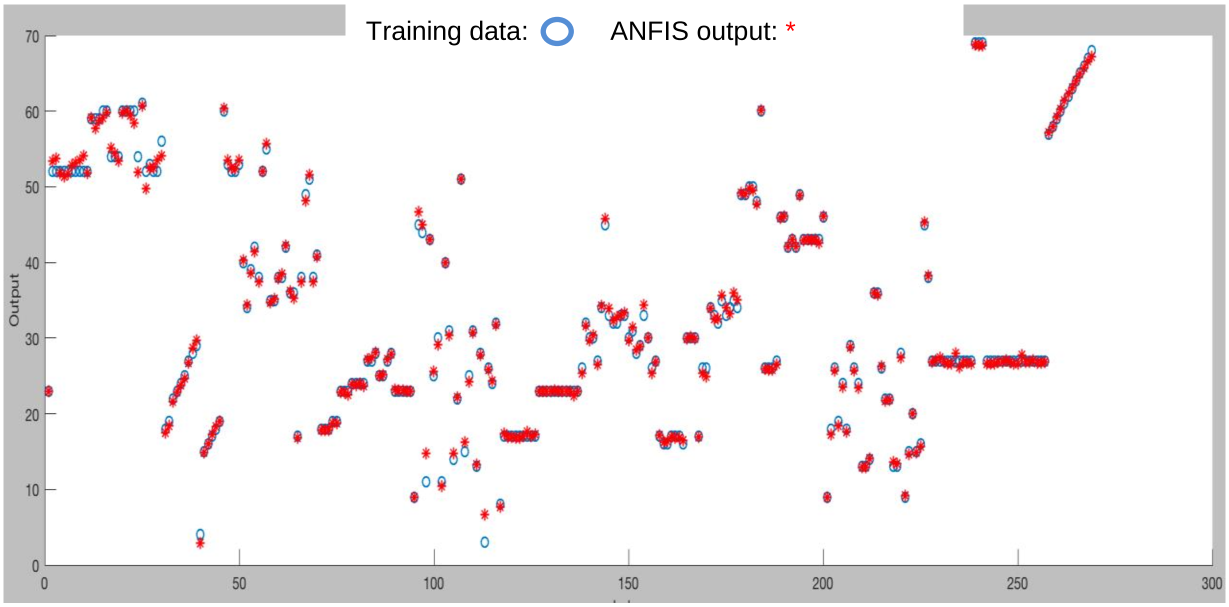


Figure 5-11 Training Data and ANFIS Output

MATLAB R2017

5.3 Discussion of Results

The remaining useful life of wall thickness loss for ANN models are calculated based on Eq. 3.13 (section 3.2.8). The remaining useful life is predicted based on best ANN model. Figure 5-12 illustrates the relationship between remaining useful life of wall thickness loss and predicted remaining useful life of best ANN model using neural network. The X-axis presents the estimated RUL of wall thickness loss and the Y-axis presents the predicted RUL of best ANN model. The correlation is a polynomial regression with R^2 of

93%. The high correlation presents the significance of wall thickness parameter compared to other input parameters.

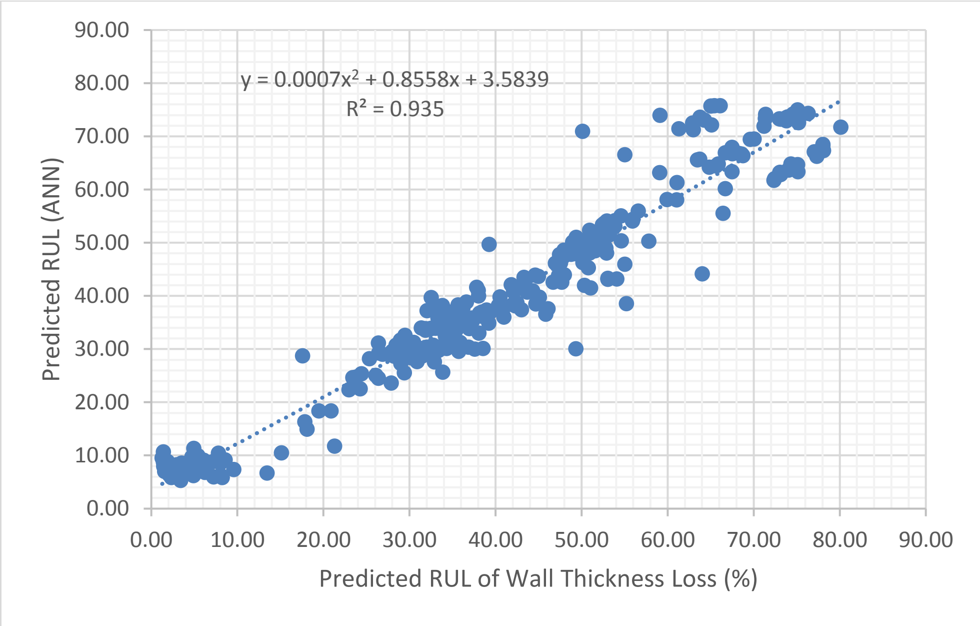


Figure 5-12 Relationship between Wall Thickness Loss and RUL

The remaining useful life of pipe age for ANN model is calculated based on Eq. 3.13 (section 3.2.8). The remaining useful life is predicted based on best ANN model. Figure 5-13 illustrates the relationship between pipe age reaming useful life and predicted remaining useful life of best ANN model using neural network. The X-axis presents the estimated RUL of pipe age and the Y-axis presents the predicted RUL of best ANN model. The correlation is a polynomial regression with R^2 of 94%. The high correlation presents the significance of pipe age parameter compared to other input parameters.

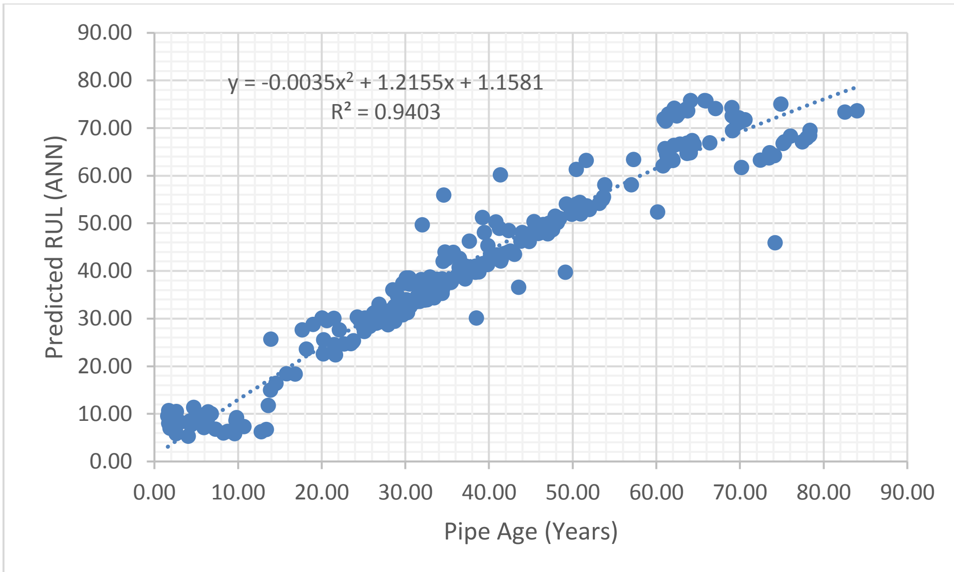


Figure 5-13 Relationship between Pipe Age and RUL

According to the results obtained from statistical analysis, neural network and adaptive fuzzy inference system, it was demonstrated that neural network and ANFIS were adept in capturing the relationships that give an indication of the prediction of remaining useful life. According to neural network results, age, and wall thickness loss were most significant parameters. Based on ANFIS, age, wall thickness loss, installation year; and from statistical analysis age and wall thickness loss were the most important parameters. It is concluded that age and wall thickness loss have most significant relationships with remaining useful life (output). Table 5-4 presents the factors that have most impact on remaining useful life.

Table 5-4 Most Significant Input Parameters in Database

Models	Material	Age	Length	Year	Wall Loss
ANN	X	X	X	X	X
ANFIS	N/A	X	N/A	X	X
Statistical	N/A	X	N/A	N/A	X

5.3.1. Condition Assessment of Entire Water System

The remaining useful life of entire water system in database is categorized to five classes based on statistical analysis results (section 4.5). The classification is from critical to very good condition described in chapter 2 (section 2.6). The total condition assessment of entire water system based on remaining useful life is presented in Table 5-5 and shown graphically in Figure 5-14. The results show 27% of total water system is in very good condition and 15% is in critical condition.

Table 5-5 Condition Assessment of Entire Water System in Database

Color	Description	RUL	% of the Entire Water System
	Very Good	> 50 years	27%
	Good	41 – 50 years	20%
	Fair	31 – 40 years	22%
	Poor	21 – 30 years	16%
	Critical	< 20 years	15%
Total			**100%**

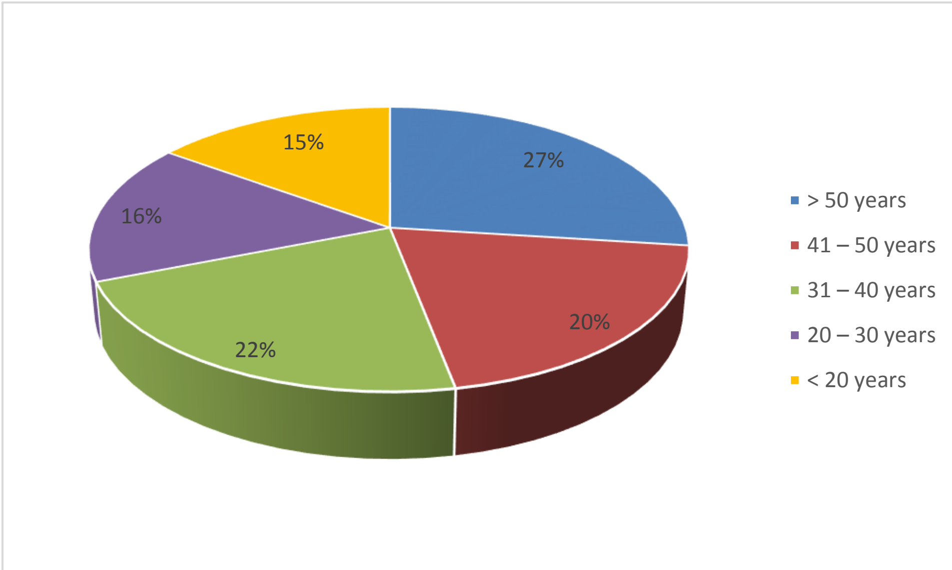


Figure 5-14 Condition Assessment of Entire Water System in Database

The remaining useful life of various types of pipe in database are provided in Table 5-6 and show graphically in Figure 5-15. The results show the remaining useful life of most cast iron and asbestos cement pipes are between 21 to 30 years. Moreover, the remaining useful life of ductile iron and steel pipes are higher than 40 years. In addition, steel pipes and ductile iron pipes last longer than cast iron and asbestos cement pipes.

Table 5-6 Percentage of Pipe Material and Remaining Useful Life in Database

Remaining Useful Life	Cast Iron	Ductile Iron	Asbestos Cement	Steel
> 40 years	25%	66%	1%	62%
31 – 40 years	2%	30%	24%	4%
21 – 30 years	45%	2%	52%	30%
11 – 20 years	24%	2%	23%	0%
0 – 10 years	4 %	0%	0%	4.0%

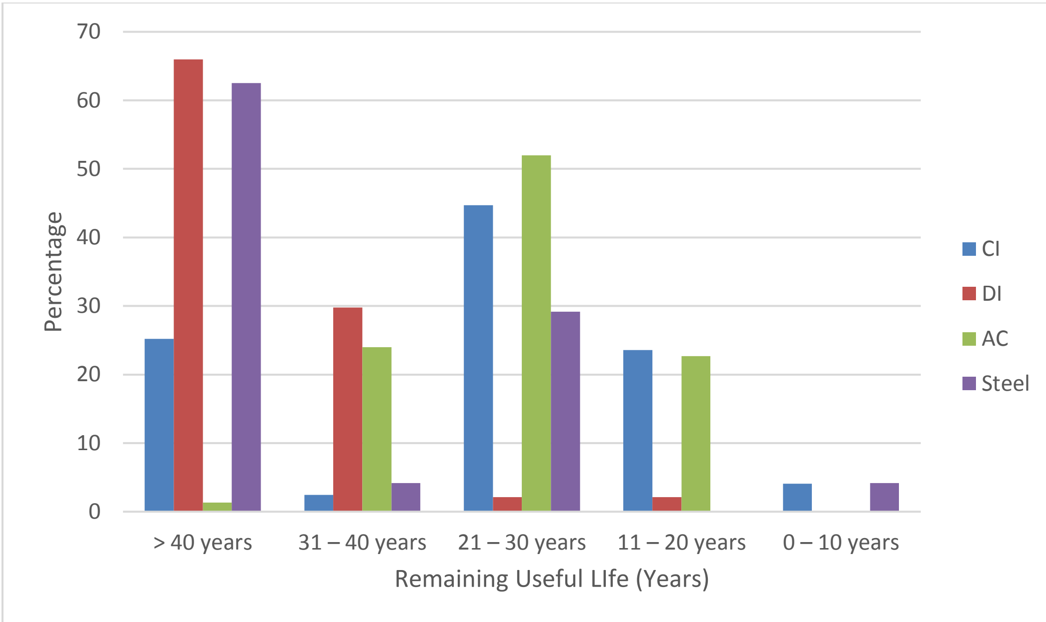


Figure 5-15 Pipe Material and Remaining Useful Life in Database

Figure 5-16 illustrates the relationship between wall thickness loss and remaining useful life in database in different ages for cast iron pipes. The results show with increasing wall thickness loss remaining useful life decrease and pipes in old ages have a high renege of wall thickness loss compared to the pipes in young ages. The results show increasing 8% of wall thickness loss for pipes greater than 60 years old, the remaining useful life decrease 70% approximately. Similarly, increasing 20% of wall thickness loss for pipes between 50 to 60 years old, the remaining useful life decrease 25% approximately.

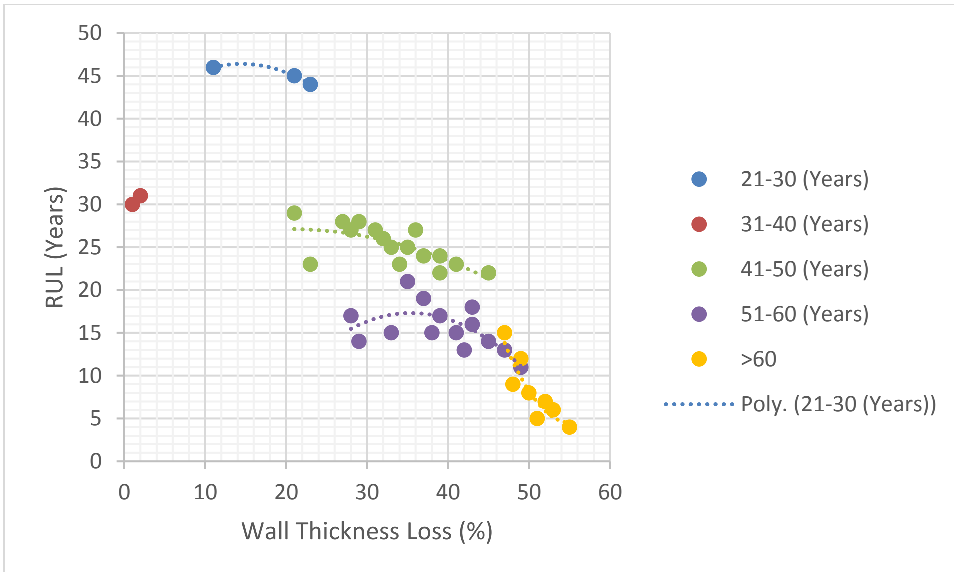


Figure 5-16 Remaining Useful Life Prediction for Cast Iron Pipes in Database

Figure 5-17 illustrates the relationship between wall thickness loss and remaining useful life in database in different ages for ductile iron pipes. The results show with increasing wall thickness loss remaining useful life decrease and pipes in old ages have a high renege of wall thickness loss compared to the pipes in young ages. The results show increasing 12% of wall thickness loss for pipes between 31 to 40 years old, the remaining useful life decrease 10% approximately. Similarly, increasing 14% of wall thickness loss for pipes between 21-30 years, the remaining useful life decrease 20% approximately.

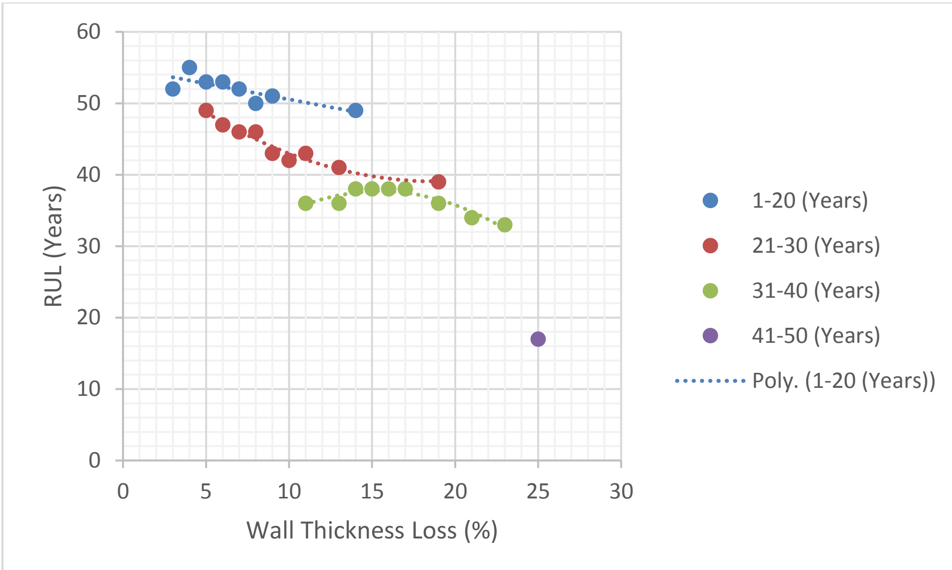


Figure 5-17 Remaining Useful Life Prediction for Ductile Iron Pipes in Database

Figure 5-18 illustrates the relationship between wall thickness loss and remaining useful life in database in different ages for asbestos cement pipes. The results show higher value of wall thickness loss with decreasing of remaining useful life and pipes in old ages have a high renege of wall thickness loss compared to the pipes in young ages. The results show increasing 20% of wall thickness loss for pipes between 51 to 60 years old, the remaining useful life decrease 60% approximately. Similarly, increasing 30% of wall thickness loss for pipes between 41-50 years old, the remaining useful life decrease 40% approximately.

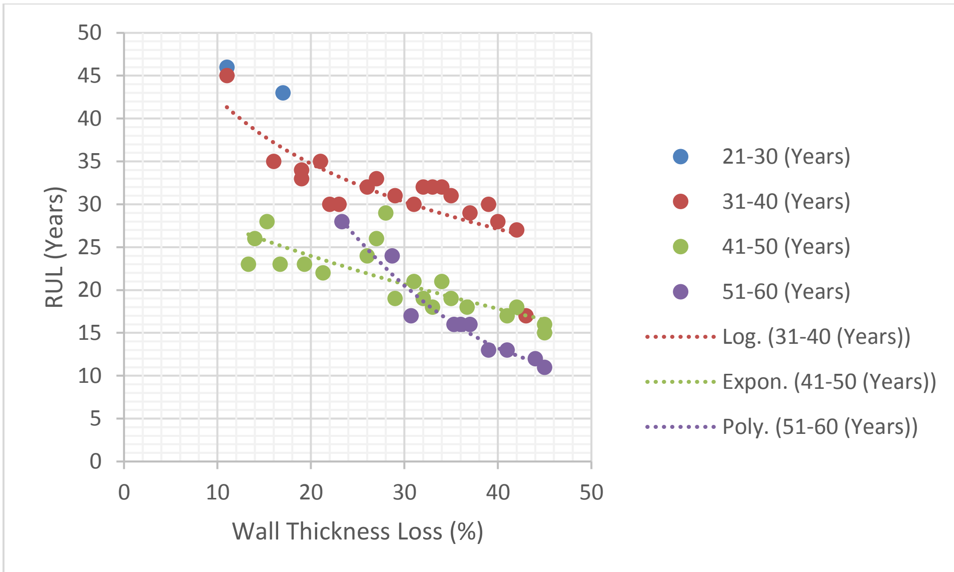


Figure 5-18 Remaining Useful Life Prediction for Asbestos Cement Pipes in Database

Figure 5-19 illustrates the relationship between wall thickness loss and remaining useful life in database in different ages for steel pipes. The results show higher value of wall thickness loss with decreasing of remaining useful life and pipes in old ages have a high renege of wall thickness loss compared to the pipes in young ages. The results show increasing 8% of wall thickness loss for pipes between 1 to 20 years old, the remaining useful life decrease 23% approximately. Similarly, increasing 8% of wall thickness loss for pipes between 41-50 years old, the remaining useful life decrease 20% approximately.

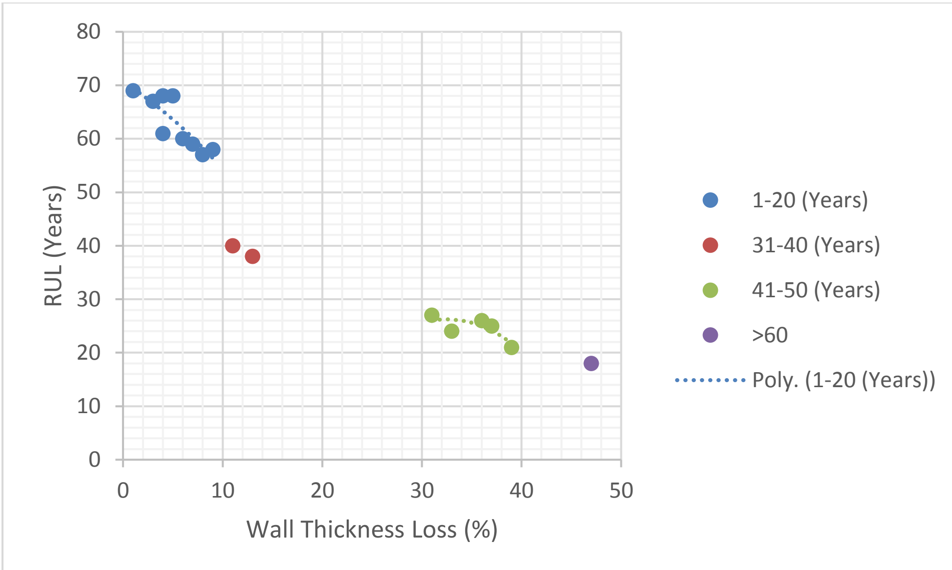


Figure 5-19 Remaining Useful Life Prediction for Steel Pipes in Database

5.3.1 Wall Thickness Loss (Corrosion) Results of Water Pipes

Table 5-7 presents the wall thickness loss (corrosion) of cast iron pipes in database based on remaining useful life and age of water pipes. The corrosion is divided to five categories based on the low to high wall thickness loss of water pipes. The result shows average remaining useful life of cast Iron pipes between 51-60 years and corrosion is greater than 40% is 14 years and for pipes greater than 60 years is 8 years.

Table 5-7 Corrosion of Cast Iron Pipes in Database

Corrosion (%)	Average Remaining Useful Life				
	21-30 years	31-40 years	41-50 years	51-60 years	> 60 years
10% -20%	46	30	N/A	N/A	N/A
20 % -30%	44	N/A	27	15	N/A
30%-40%	N/A	N/A	25	17	N/A
> 40%	N/A	N/A	22	14	8

Table 5-8 presents the wall thickness loss (corrosion) of ductile iron pipes in database based on remaining useful life and age of water pipes. The corrosion is divided to five categories based on the low to high corrosion of water pipes. The result shows average remaining useful life of ductile iron pipes between 41-50 years and corrosion is from 20% to 30% is 17 years.

Table 5-8 Corrosion of Ductile Iron Pipes in Database

Corrosion (%)	Average Remaining Useful Life			
	1-20 years	21-30 years	31-40 years	41-50 years
< 10%	52	46	N/A	N/A
10% -20%	49	41	37	N/A
20% -30%	N/A	N/A	33	17

Table 5-9 presents the wall thickness loss (corrosion) of asbestos cement pipes in database based on remaining useful life and age of water pipes. The corrosion is divided to five categories based on the low to high corrosion of water pipes. The result shows average remaining useful life of asbestos cement pipes between 51-60 years and corrosion greater than 40% is 12 years.

Table 5-9 Corrosion of AC Pipes in Database

Corrosion (%)	Average Remaining Useful Life			
	21-30 years	31-40 years	41-50 years	51-60 years
10% -20%	44	37	25	N/A
20 % -30%	N/A	32	24	26
30%-40%	N/A	30	19	16
> 40%	N/A	27	16	12

Table 5-10 presents the wall thickness loss (corrosion) of steel pipes in database based on remaining useful life and age of water pipes. The corrosion is divided to five categories based on the low to high corrosion of water pipes. The result shows average remaining useful life of steel pipes greater than 60 years and corrosion greater than 40% is 18 years.

Table 5-10 Corrosion of Steel Pipes in Database

Corrosion (%)	Average Remaining Useful Life			
	1-20 years	31-40 years	41-50 years	> 60 years
< 10%	63	N/A	N/A	N/A
10% -20%	N/A	39	N/A	N/A
20 % -30%	N/A	N/A	N/A	N/A
30%-40%	N/A	N/A	25	N/A
> 40%	N/A	N/A	N/A	18

The deterioration models are determined with most significant variables (age and wall thickness loss). Variables are added into the non-linear multivariable regression (X_1: age, X_2: wall thickness loss and Y: remaining useful life). The non-linear multivariate regression is described in chapter 4 (section 4.4). The regression models are selected based on high correlation with variables started from degree 1 and repeated the process with degree two and three to find the best correlation and high value of coefficient of determination. Table 5-11 presents deterioration models for different types of water pipes. The coefficient of determination (R^2) of these models (0.73 – 0.80) are considered good fit for non-linear regression models.

Table 5-11 Deterioration Models for Different Pipes Material in Database

Pipe Material	Non-linear Regression Models	R^2
Cast Iron	$Y= -0.342A^2 + 0.0548W + 48.163$	0.78
Ductile Iron	$Y= 0.004A^3 -0.025W^2 + 0.11AW + 51$	0.74
Asbestos Cement	$Y= 0.0038A^2 -0.49W + 195.92$	0.80
Steel	$Y= 0.005A^3 -0.012W^2 -0.989AW - 0.012$	0.73
A: Age of Pipes, W: Wall Thickness Loss, Y: Remaining Useful Life		

5.3.2 Contribution to The Water Pipeline Industry

The results of this dissertation can help water utilities to manage and optimize their water distribution system. Furthermore, required actions need to consider based on remaining useful life and corrosion rates. The AC pipe replacement with new main construction are recommended for the project due to the current high cost of rehabilitation (Providence Infrastructure Consultants, 2016). As future advancements are made with the technology, rehabilitation may become a more cost-effective alternative to new pipe construction. Moreover, Cat Iron (CI) pipes needs to be replaced within the next five (5) years since it is reaching the end of its useful life based on the installation year and a design service life CI pipes.

5.4 Chapter Summary

This chapter presented the detailed overview of the development of the neural network and neuro fuzzy inference system models. Numerous structures were tested and the best architecture among them was chosen for further explanation and development. It was observed that the model displayed a good learning trend towards the facts presented. Moreover, the research used the neuro fuzzy inference system and the training data to create the predicted model to forecast the remaining useful life. The model constructed using the neuro fuzzy inference system theory can efficiently forecast the remaining useful life of water pipes. It is concluded that the applications of neural network and Adaptive Neural Fuzzy Inference System (ANFIS) to solve the problem of remaining useful life prediction of water pipes are feasible and the precision of the model depends on obtaining a larger and more comprehensive sample size. Moreover, Pipe age and wall thickness loss are most significant parameters to predict the remaining useful life of the water pipes. Additionally, ductile iron and steel pipes have more remaining useful life compared to cast iron and asbestos-cement pipes.

Chapter 6 Conclusions and Recommendations for Future Research

Due to their low visibility, rehabilitation of underground water system is often ignored until a terrible failure occurs. This, often, results in expensive and difficult rehabilitation due to the crucial nature of confirming that the water system is operational. Most of water pipe projects are not considered a “practical” method for the failure of water pipes. There are two main reasons for this: the first is the inaccessibility of acceptable information regarding the condition of the water system. The second is the uselessness of predicting water scarcity prior to failure or an adverse condition so that inspection and repairs could be performed prior to failure of the system that might lead to an expensive fix and other risks.

The main contribution of this dissertation is the development of neural network model and neuro fuzzy inference system to evaluate the prediction of remaining useful life. This prediction model is developed to improve the objectivity of practical management of water systems. Since all the parameters that were recognized to affect the water deterioration, as recognized in the literature were not readily available to be combined in this model, it is recommended that the model be expanded to include those parameters and retrained. Through this process, the neural network will keep learning the updated information and adjust its hidden weights to ensure the predicting precision.

To accurately quantify the effect of certain input parameters for water deterioration, it will be useful to develop a neural network model, as demonstrated in this dissertation as an initial starting base. However, the models with more descriptive parameters will improve the understanding of the effects of influencing input parameters on water systems. The specific conclusions of this dissertation are:

- Increasing 30% of wall thickness loss for asbestos-cement pipes between 41-50 years old, the remaining useful life decrease 40% approximately.

- Increasing 20% of wall thickness loss for cast iron pipes between 50 to 60 years old, the remaining useful life decrease 25% approximately.
- Increasing 14% of wall thickness loss for ductile iron pipes between 21-30 years, the remaining useful life decrease 20% approximately.
- Increasing 8% of wall thickness loss for steel pipes between 41-50 years old, the remaining useful life decrease 20% approximately.
- On the average, with approximately 10% of wall thickness loss in existing cast iron, ductile iron, asbestos-cement and steel water pipes, the reduction of their remaining useful life will be approximately 50%.

6.1 Limitations of this Research

As indicated previously, this research demonstrated the possibility of using neural network and neuro fuzzy inference system models as a screening tool to predict the remaining useful life of water pipes. The availability of fewer numbers of deterioration parameters and limited data availability posed the primary disadvantage to effective neural network training and caused the main limitation to this dissertation. Environmental parameters affecting the pipe, such as overburden pressure, soil type and properties, underground water-table location and other factors identified in the literature were omitted due to lack of monitoring of the data. Review of literature showed these parameters to be suitable measures of prediction of remaining useful life. Hence, the largest limiting factor for modeling ease and precision was the unavailability of inclusive data.

6.2 Recommendations for Future Research

Since the developed model does not include several parameters thought to be important to water deterioration, the model developed in this exercise is not complete. While it determines the utility of using artificial neural networks and neuro fuzzy inference system models for predicting water condition, further work for data collection and model development is required to confirm that the model is more precise and reliable for future applications. Having made the above conclusions, it is clear more work is required to simplify future use of the model. This dissertation determines the need for the following actions, to facilitate simplicity, and more comprehensive development of artificial neural network models and neuro fuzzy inference system models for water condition prediction:

- Utilization of more data.
- Examination of input parameter importance, and other factors affecting water deterioration.

To further the development of neural network and neuro fuzzy inference system models that are precise and flexible, inclusion of more descriptive data is needed. The model developed in this research required making assumptions that were scope limiting since it required values from some factors that may affect the remaining useful life prediction. The availability of detailed soils parameters, water-table location, physical pipe characteristics, and pipe conditions would be resources to model the deterioration of water and precisely predict the remaining useful life. A list of possible parameters that impact on remaining useful life prediction and can be factored into the model is listed below:

- Soil conditions
- Groundwater location and fluctuation
- Joint condition
- Water Pressure

- Leakage History
- Installation Depth
- Temperature
- Water Corrosive Conditions

The neural network and Adaptive Neural Fuzzy Inference System (ANFIS) water remaining useful life prediction models, can then be combined with an inclusive infrastructure asset management system to aid the municipal agencies in better planning and spending of their limited available budget.

Appendix A

Abbreviations

AC – Asbestos-cement Pipe

ANOVA – Analysis of Variance

ASCE – American Society of Civil Engineers

ASL – Anticipated Service Life

ANN - Artificial Neural Network

ANFIS – Adaptive Neural Fuzzy Inference System

AWWA – American Water Works Association

AWWARF – American Water Works Association Research Foundation

BPNN - Backpropagation Neural Network

CCTV - Closed Circuit Television

CIP – Cast Iron Pipe

CIP – Capital Improvement Projects

COF – Consequence of Failure

CPP - Pressured Concrete Pipe

CUIRE - Center for Underground Infrastructure Research and Education

DIP – Ductile Iron Pipe

DSL – Design Service Life

HDPE - High Density Polyethylene

HFES - Hierarchical Fuzzy Expert System

LOF – Likelihood of Failure

MAE – Mean Absolute Error

MAPE – Mean Absolute Percentage Error

MFL - Magnetic Flux Leakage

NASTT - North American Society of Trenchless Technology

PCCP – Prestressed Concrete Cylinder Pipe

RRSE – Root Relative Square Error

PVC - Polyvinyl Chloride Pipe

RAE – Relative Absolute Error

RCP – Reinforced Concrete

RFEC - Remote Field Eddy Current

R^2 – Coefficient of Determination

RUL – Remaining Useful Life

SP – Steel Pipe

TxDOT - Texas Department of Transportation

UTA - The University of Texas at Arlington

WRF – Water Research Foundation

Appendix B

Case Studies

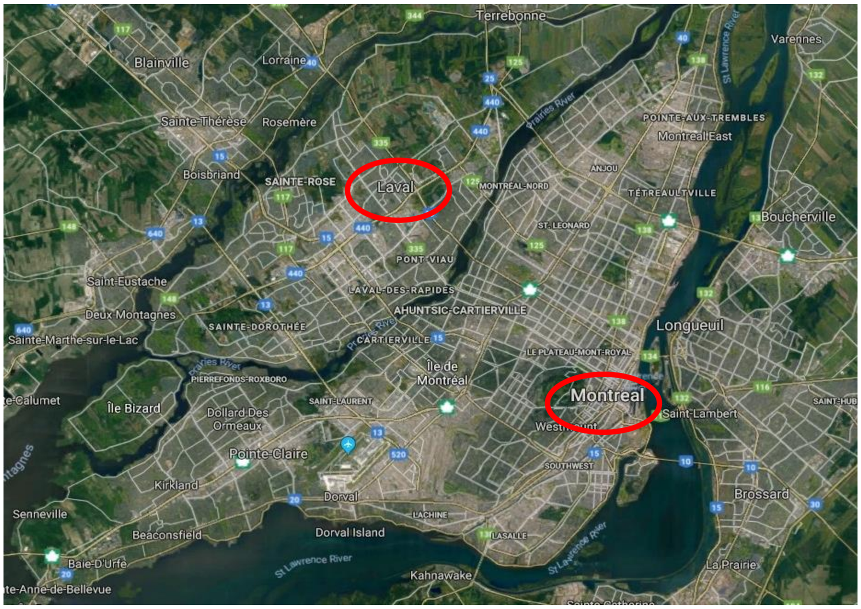


Figure B-1 City of Montreal and Laval, Canada

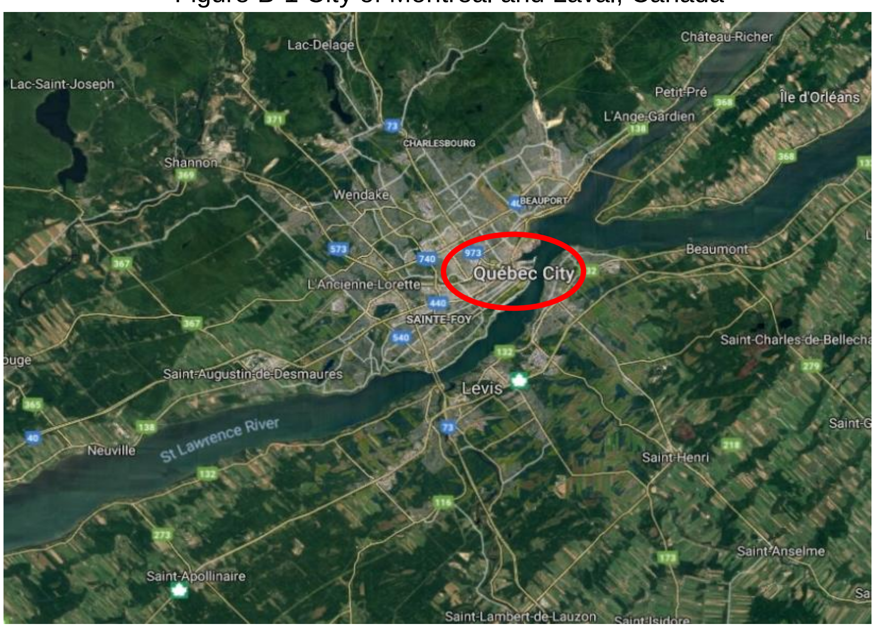


Figure B-2 Quebec City, Canada

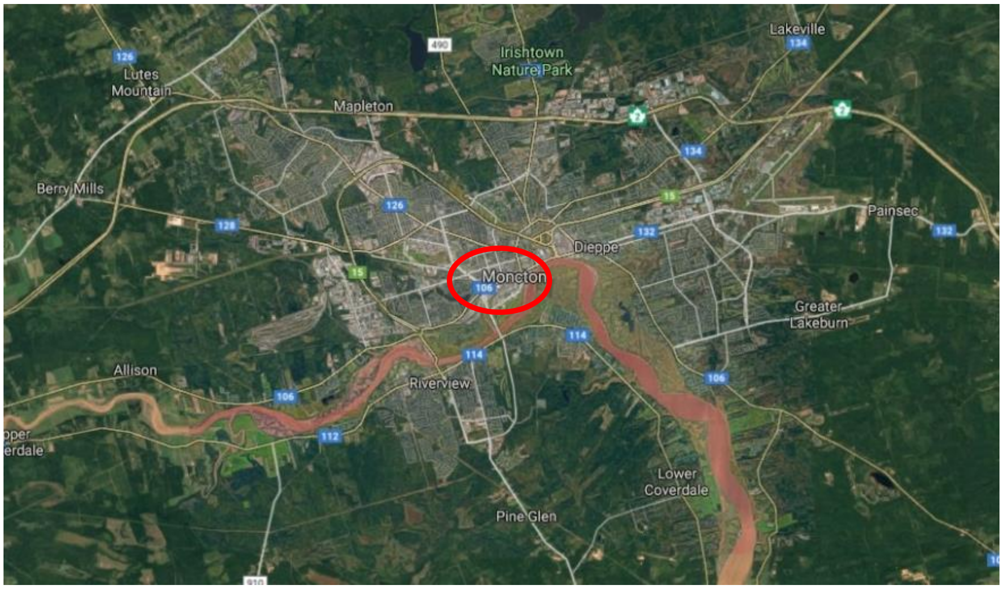


Figure B-3 City of Moncton, Canada

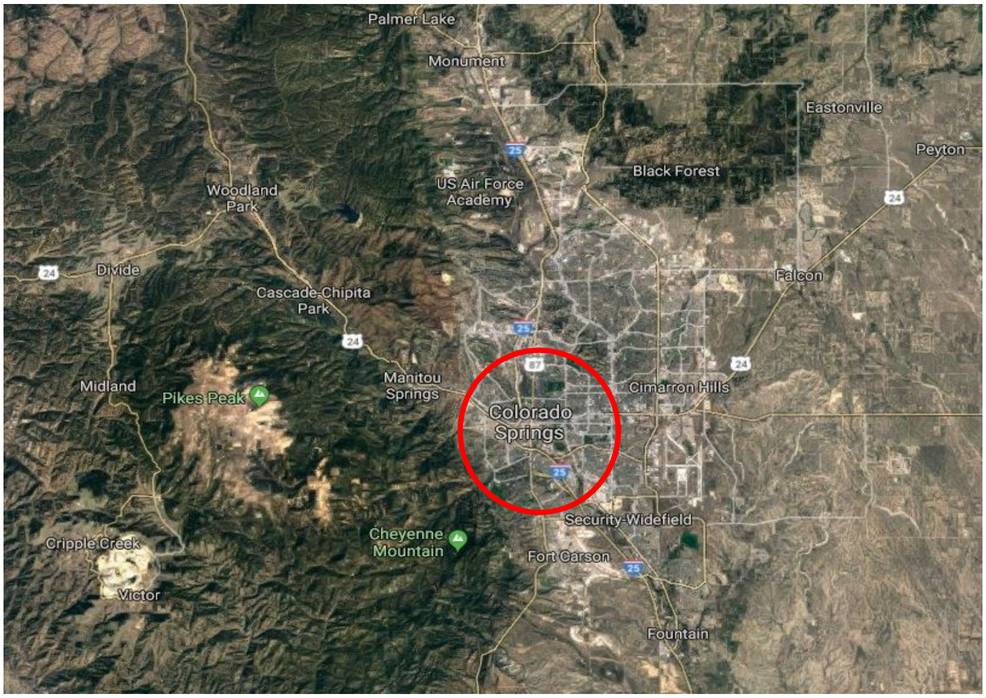


Figure B-4 Colorado Springs, Colorado

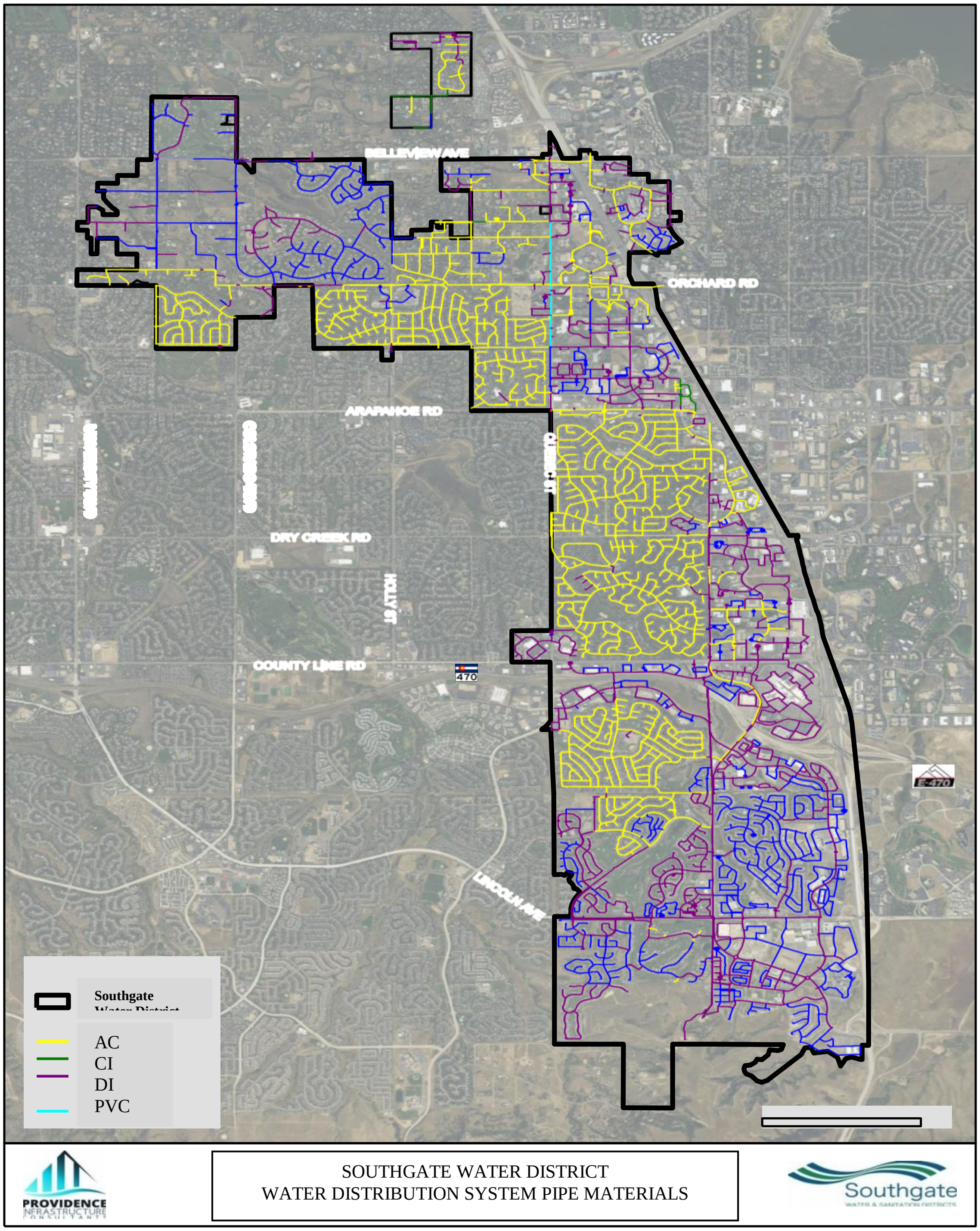
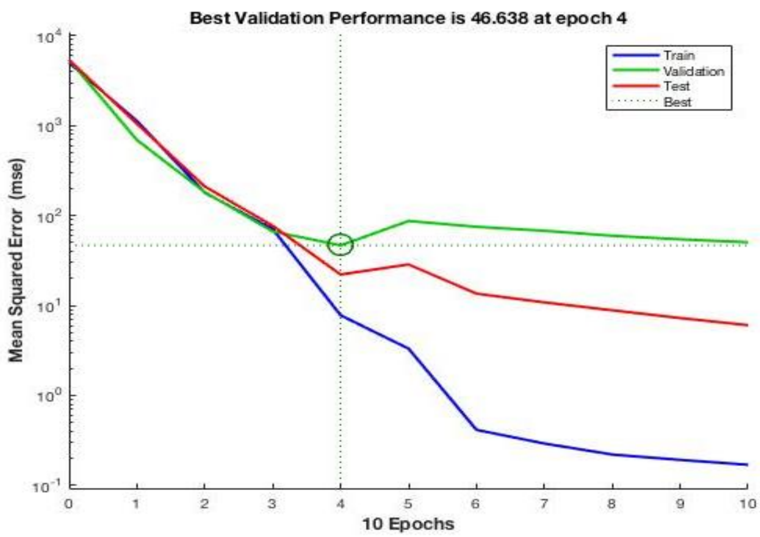


Figure B-5 Southgate Water Distribution System, Denver, Colorado

Appendix C

Neural Network Results

Figure C-1 illustrates the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 4. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.



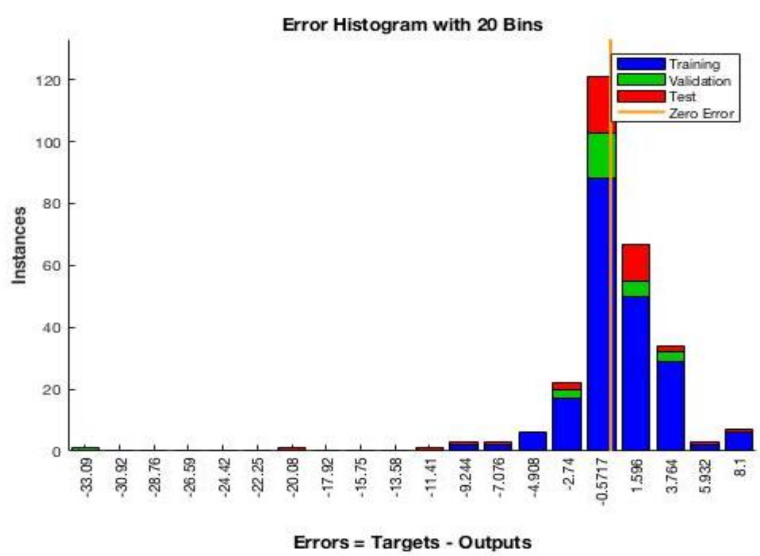


Figure C-1 Performance Chart and Histogram of Errors of ANN Model 1

Figure C-2 shows the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 16. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.

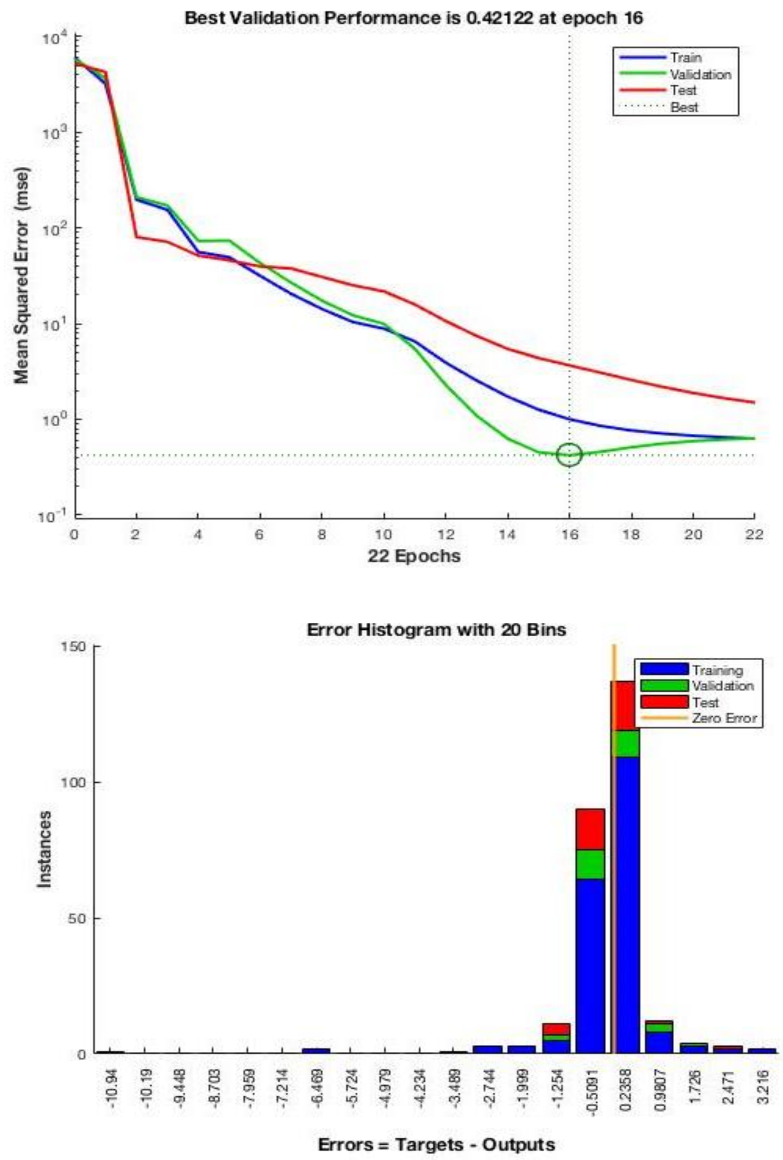


Figure C-2 Performance Chart and Histogram of Errors of ANN Model 2

Figure C-3 shows the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 26. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.

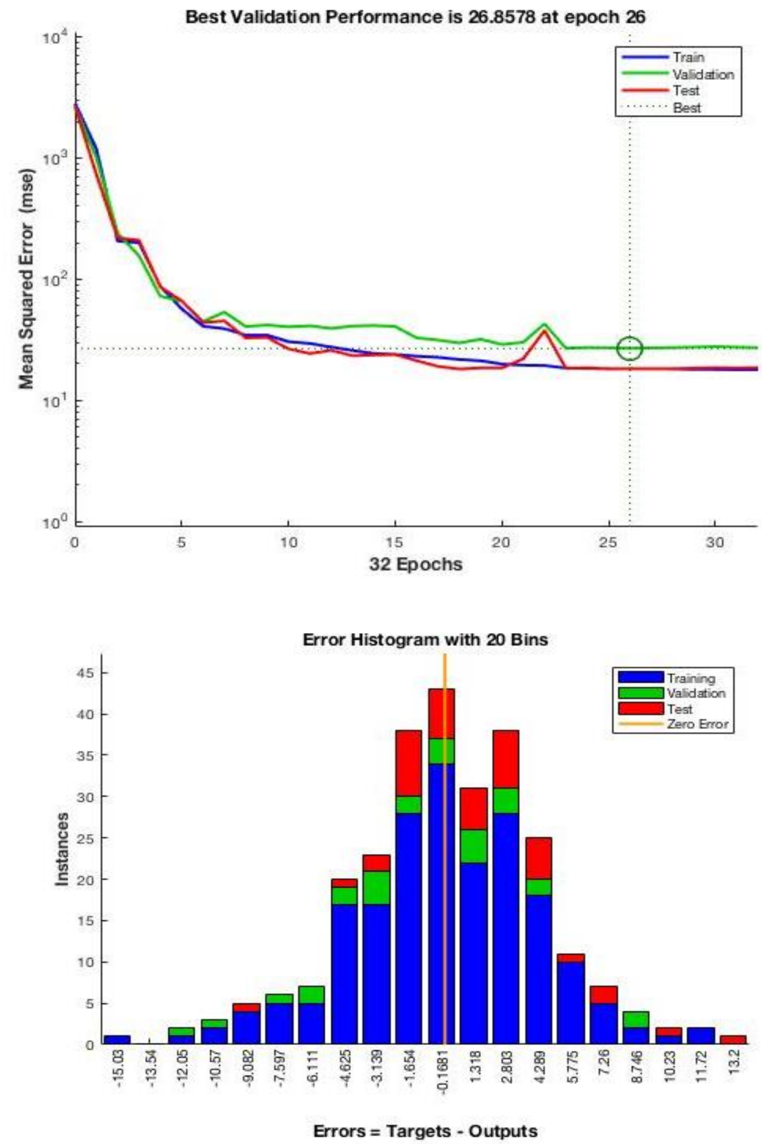
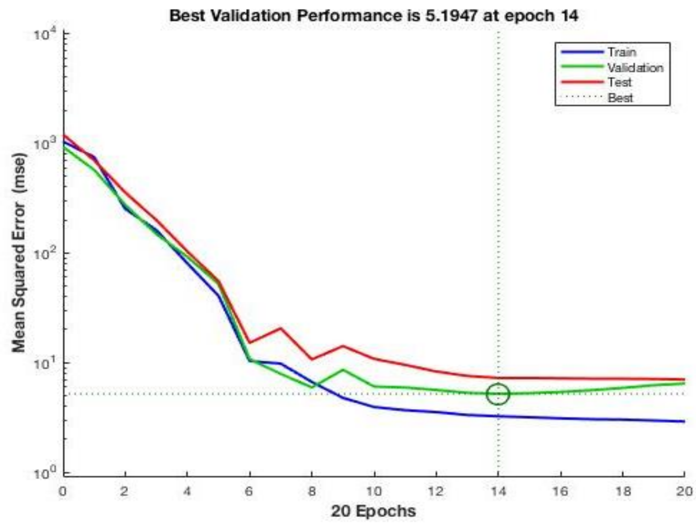


Figure C-3 Performance Chart and Histogram of Errors of ANN Model 3

Figure C-4 shows the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 14. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.



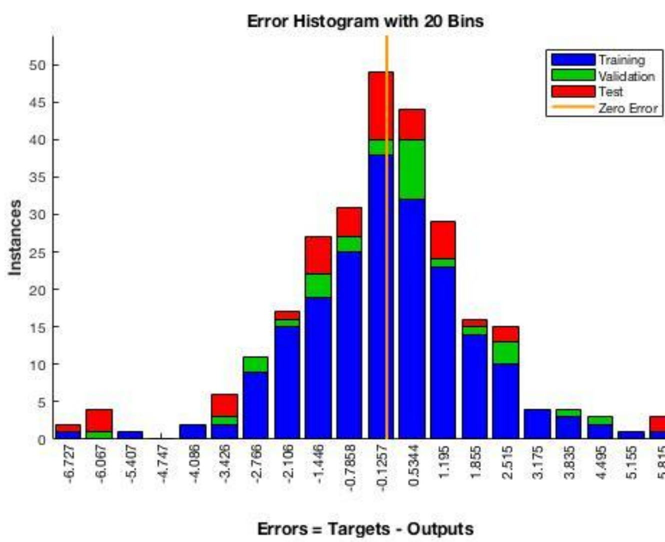
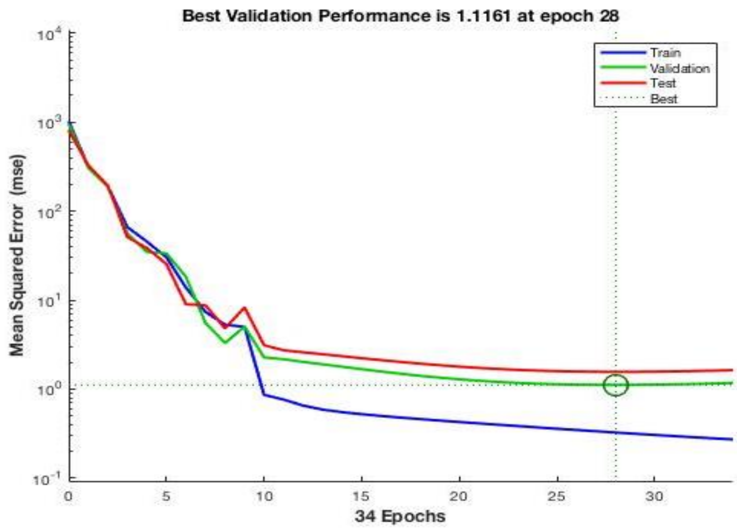


Figure C-4 Performance Chart and Histogram of Errors of ANN Model 4

Figure C-5 shows the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 28. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.



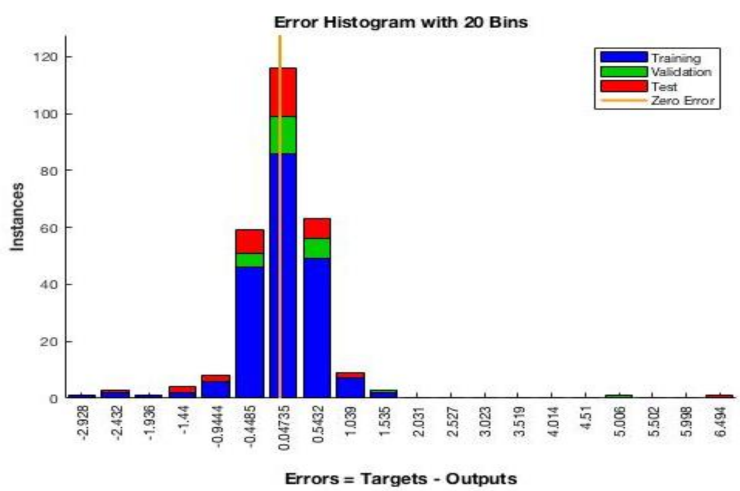


Figure C-5 Performance Chart and Histogram of Errors of ANN Model 5

Figure C-6 shows the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 18. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.

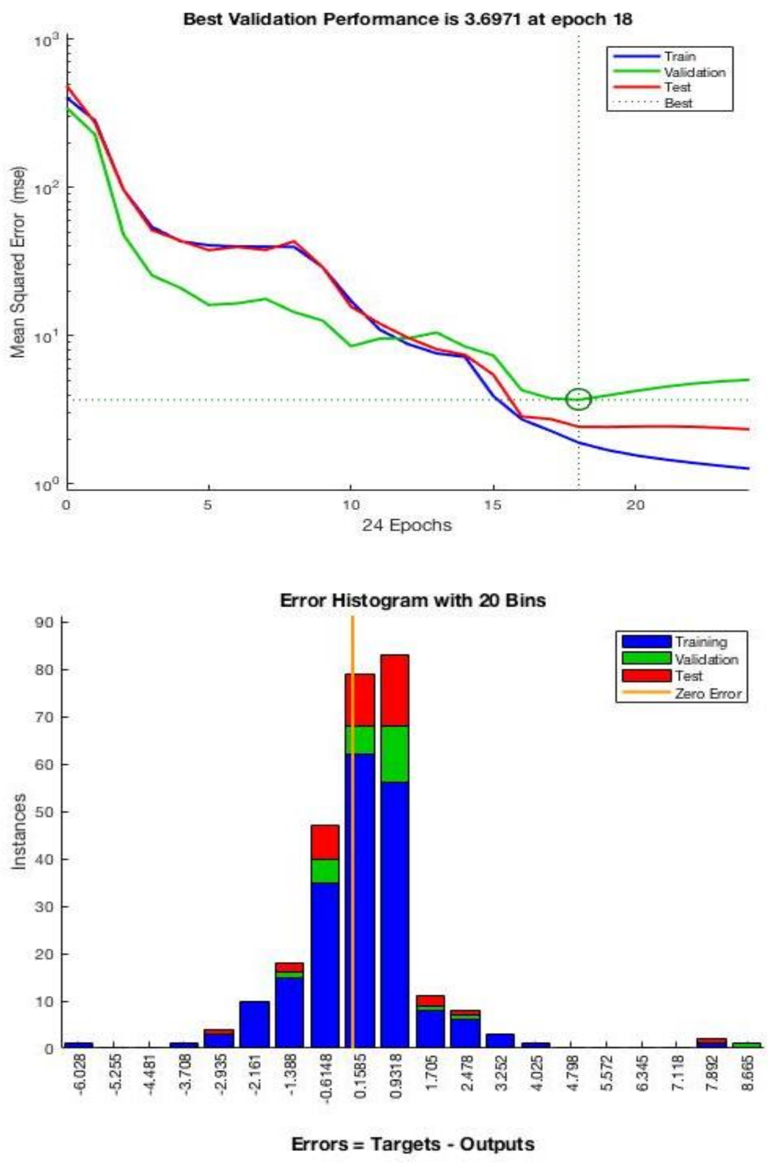


Figure C-6 Performance Chart and Histogram of Errors of ANN Model 6

Figure C-7 shows the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 1. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.

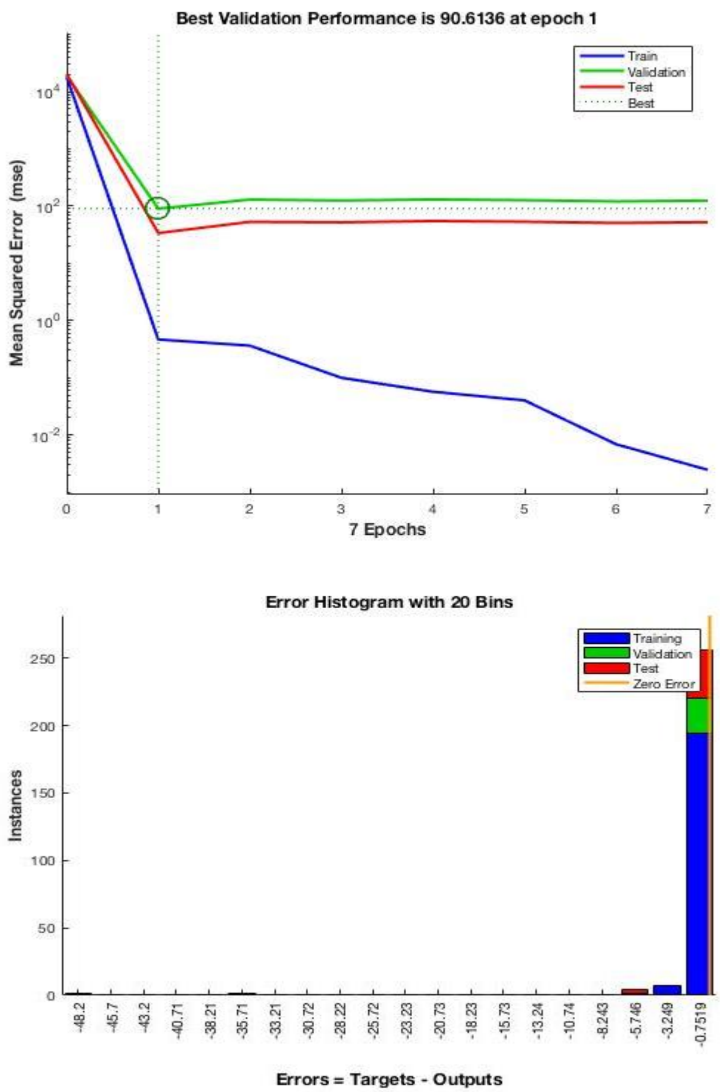
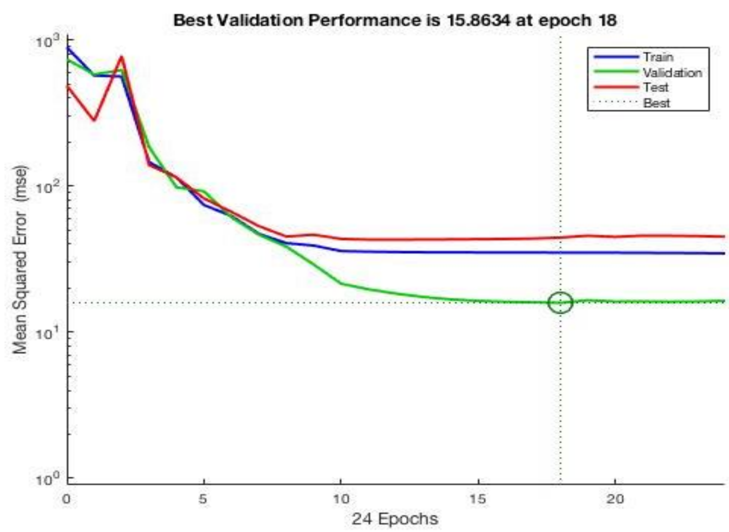


Figure C-7 Performance Chart and Histogram of Errors of ANN Model 7

Figure C-8 shows the performance chart for the trend of validation, training and testing models are similar and best validation happens in epoch 18. The histogram of error presents the number of errors in training are more than validation and testing. The histogram of error and performance chart are described in section 5.1.1 and 5.1.2.



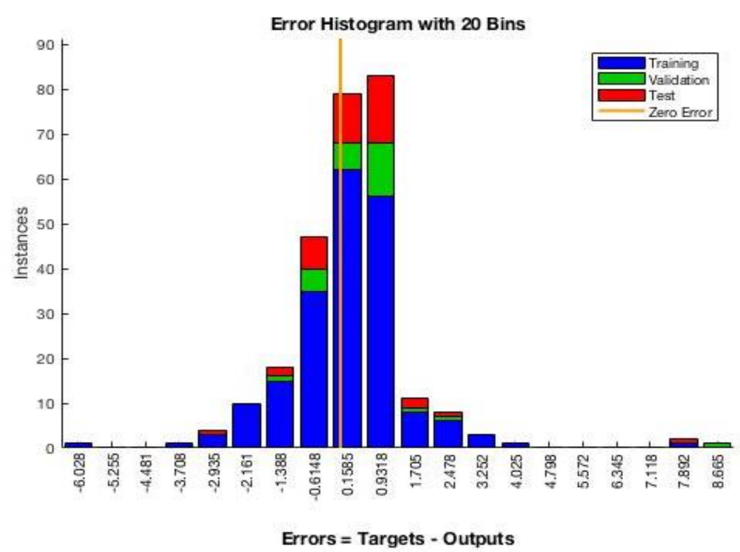


Figure C-8 Performance Chart and Histogram of Errors of ANN Model 8

Figure C-9 illustrates the training error using ANFIS. Input and output data have been entered to ANFIS using hybrid method. The training error has been described in section 5.2.1.

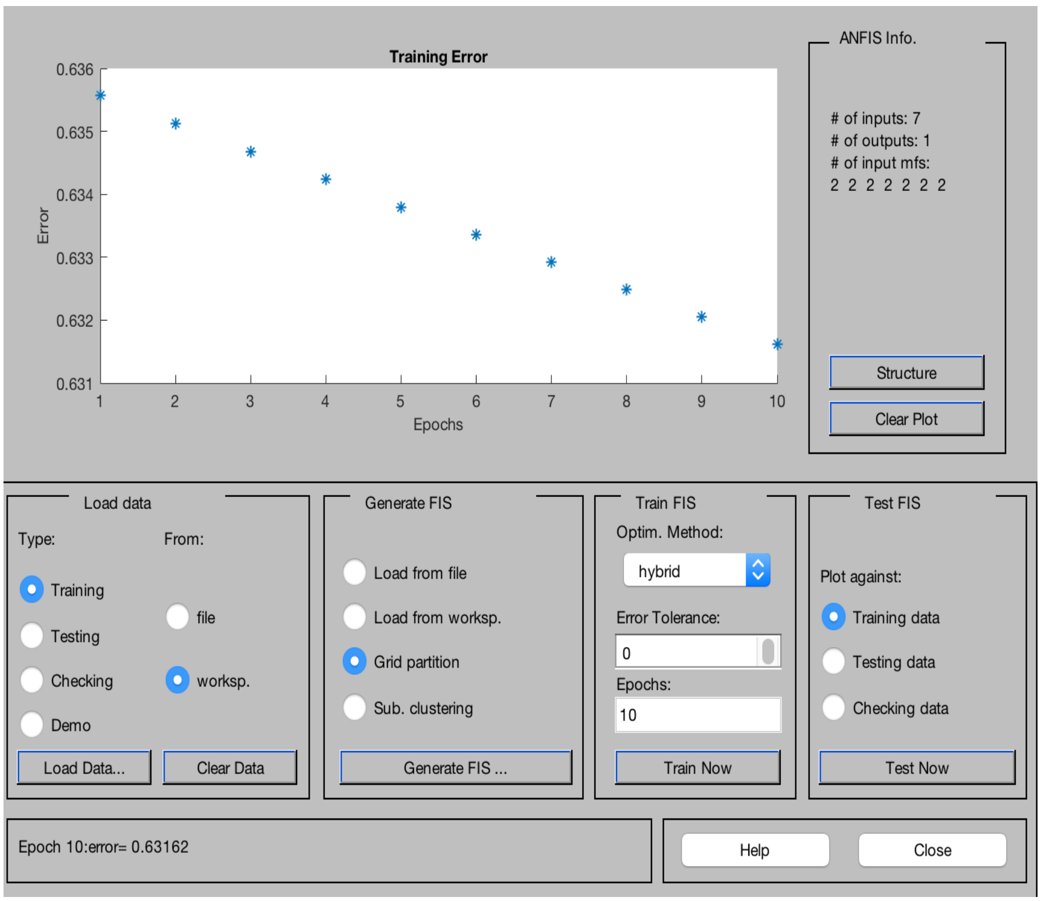


Figure C-9 Error Results of the Model Using ANFIS

MATLAB R2017

Figure C-10 illustrates the results of training data and ANFIS output. Input and output data have been entered to ANFIS using hybrid method. The results of training data and ANFIS output has been described in section 5.2.2.

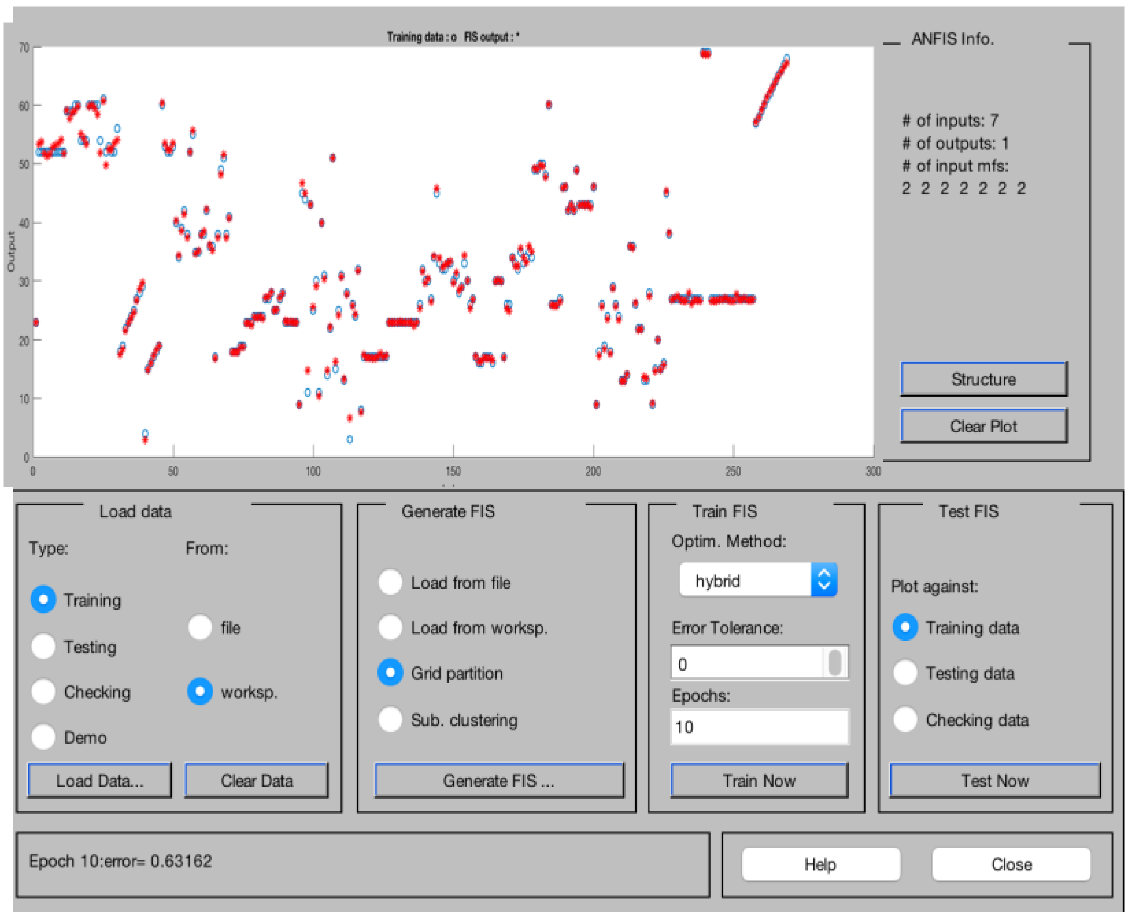


Figure C-10 Results of Training Data and ANFIS Output

MATLAB R2017

Appendix D

Data Samples Used for Modeling

Material*	Age	Length (ft)	Diameter (in.)	# of Break	Installation year	Wall Loss (%)	RUL
CI	93	20.5	6	0	1913	41	23
CI	122	413.4	6	5	1896	9	52
CI	122	374	6	1	1896	8	52
CI	122	341.2	6	2	1896	13	52
CI	122	240.4	6	1	1896	14	52
CI	122	465.8	6	1	1896	13	52
CI	122	94.4	6	1	1896	11	52
CI	122	200.45	6	1	1896	10	52
CI	122	231.95	6	1	1896	9	52
CI	122	175.5	6	1	1896	7	52
CI	122	216.2	6	1	1896	13	52
CI	129	38.7	6	1	1889	8	59
CI	129	439.6	6	1	1889	11	59
CI	129	215.2	6	1	1889	9	59
CI	130	465.8	6	4	1888	9	60
CI	130	649.6	6	3	1888	7	60
CI	124	175.1	6	1	1894	9	54
CI	124	52.4	6	1	1894	11	54
CI	124	278.5	6	2	1894	13	54
CI	130	84.6	6	1	1888	8	60
CI	130	158.4	6	1	1888	7	60
CI	130	40.3	6	1	1888	9	60
CI	130	308	6	4	1888	11	60
CI	124	160	6	2	1894	16	54
CI	131	205.3	6	1	1887	7	61
CI	122	230.3	6	1	1896	17	52
CI	123	363.9	6	1	1895	13	53
CI	122	593.8	6	1	1896	11	52
CI	122	234.5	6	2	1896	9	52
CI	126	31.5	6	1	1892	15	56
CI	52	1492.7	4	3	1954	47	18
CI	51	255.9	4	4	1955	45	19
CI	48	157.4	4	1	1958	39	22
CI	47	324.8	7	0	1959	41	23

*The material converted to numerical value for ANN Models development (Project one).

Material	Age	Length (ft)	Diameter (in.)	# of Break	Installation year	Wall Loss (%)	RUL
CI	46	849.7	4	0	1960	37	24
CI	45	1,886.4	4	2	1961	33	25
CI	43	613.5	4	2	1963	31	27
CI	42	1,000.6	4	3	1964	39	28
CI	41	4,242	4	8	1965	45	29
CI	66	7,992	6	0	1940	47	4
CI	55	2,335.9	6	2	1951	41	15
CI	54	3,6161.4	6	10	1952	43	16
CI	53	13,500.6	6	14	1953	39	17
CI	52	36,161.4	6	3	1954	45	18
CI	51	13,500.6	6	95	1955	49	19
DI	10	326	8	0	2006	4	60
DI	17	399	12	0	1999	6	53
DI	18	299	12	0	1998	7	52
DI	18	320	12	0	1998	3	52
DI	17	230	12	0	1999	5	53
DI	30	273	8	1	1986	11	40
DI	36	375	12	1	1980	21	34
DI	31	498	12	1	1985	17	39
DI	28	384	12	1	1988	10	42
DI	32	322	12	1	1984	13	38
DI	18	321	8	0	1998	7	52
DI	15	357	12	0	2001	4	55
DI	35	287	8	2	1981	23	35
DI	35	324	8	2	1981	21	35
DI	32	287	8	1	1984	17	38
DI	32	287	8	1	1984	14	38
DI	28	373	8	1	1988	9	42
DI	34	360	8	1	1982	19	36
DI	34	226	8	0	1982	23	36
DI	53	317	18	3	1963	25	17
DI	32	351	12	2	1984	15	38
DI	21	353	8	0	1995	13	49
DI	19	352	12	0	1997	14	51

Material	Age	Length (ft)	Diameter (in.)	# of Break	Installation year	Wall Loss (%)	RUL
DI	32	452	12	3	1984	16	38
DI	29	237	12	1	1987	19	41
CI	52	95.1	12	1	1954	43	18
CI	52	39.37	12	1	1954	47	18
CI	52	72.1	12	1	1954	45	18
CI	51	328	12	2	1955	47	19
CI	51	328	12	3	1955	49	19
CI	47	190.28	12	3	1959	39	23
CI	47	190.28	12	2	1959	37	23
CI	47	190.28	12	3	1959	31	23
CI	46	485.5	12	4	1960	39	24
CI	46	485.5	12	1	1960	33	24
CI	46	485.5	12	1	1960	39	24
CI	46	485.5	12	2	1960	31	24
CI	43	324.8	12	1	1963	35	27
CI	43	354.3	12	1	1963	39	27
CI	42	88.5	16	1	1964	33	28
CI	45	482.2	16	2	1961	35	25
CI	45	482.2	16	2	1961	39	25
CI	43	167.3	16	1	1963	36	27
CI	42	295.2	16	1	1964	29	28
CI	47	324.8	24	5	1959	33	23
CI	47	324.8	24	4	1959	29	23
CI	47	324.8	24	3	1959	27	23
CI	47	324.8	24	4	1959	23	23
CI	47	324.8	24	2	1959	23	23
CI	79	5,280	12	10	1927	37	9
CI	25	2,640	6	10	1981	21	45
CI	26	5,280	10	10	1980	23	44
CI	81	2,640	6	10	1925	29	11
AC	27	21,120	6	10	1979	17	43
AC	45	15,840	8	10	1961	21	25
CI	40	5,280	6	10	1966	13	30
CI	59	21,120	8	10	1947	35	11

(Project 2)

Material	Age	Length (ft)	Diameter (in.)	# of Break	Installation year	Wall Loss (%)	RUL
CI	30	5,280	12	10	1976	11	40
CI	39	5,280	6	10	1967	16	31
CI	56	5,280	6	10	1950	29	14
CI	48	5,280	12	10	1958	33	22
DI	19	15,840	8	0	1987	9	51
CI	55	5,280	12	0	1951	28	15
CI	45	5,280	6	0	1961	27	25
AC	39	21,120	8	0	1967	17	31
CI	57	5,280	6	0	1949	33	13
CI	42	5,280	16	0	1964	19	28
CI	73	2,640	6	0	1933	49	3
AC	44	5,280	12	0	1962	27	26
CI	46	21,120	6	0	1960	25	24
CI	102	5,280	6	0	1904	19	32
CI	78	2,640	6	0	1928	45	8
AC	53	5,280	20	0	1960	36	17
AC	53	330	20	0	1960	30.7	17
AC	53	400	20	0	1960	31.3	17
AC	53	370	20	0	1960	35.3	17
AC	53	350	20	0	1960	35.3	17
AC	53	420	20	0	1960	23.3	17
AC	53	520	20	10	1960	23.3	17
AC	53	430	20	0	1960	23.3	17
AC	53	320	20	10	1960	28.7	17
AC	47	360	20	0	1966	14	23
AC	47	410	20	0	1966	15.3	23
AC	47	530	20	0	1966	26	23
AC	47	340	20	0	1966	36.7	23
AC	47	460	20	10	1966	32	23
AC	47	320	20	0	1966	19.3	23
AC	47	528	20	0	1966	28	23
AC	47	450	20	10	1966	30.7	23
AC	47	335	20	0	1966	16.7	23
AC	47	550	20	10	1966	21.3	23

Material	Age	Length (ft)	Diameter (in.)	# of Break	Installation year	Wall Loss (%)	RUL
AC	36	330	8	0	1980	23	34
AC	37	383	8	0	1979	21	33
AC	38	319	12	10	1978	26	32
AC	35	445	8	0	1981	27	35
AC	37	547	12	10	1979	29	33
AC	36	398	8	0	1980	16	34
AC	35	389	8	0	1981	29	35
AC	36	410	8	10	1980	33	34
DI	21	7,920	20	10	1985	11	49
DI	21	7,920	20	10	1985	9	49
DI	20	5,280	6	10	1986	8	50
DI	20	5,280	8	10	1986	9	50
DI	22	5,280	12	10	1984	10	48
DI	10	10,560	12	10	1996	6	60
CI	44	15,840	8	10	1962	33	26
CI	44	10,560	8	10	1962	35	26
CI	44	5,280	6	10	1962	37	26
CI	43	15,840	8	10	1963	34	27
DI	24	10,560	12	10	1982	9	46
DI	24	5,280	12	10	1982	8	46
DI	28	5,280	16	10	1978	10	42
DI	27	7,920	24	10	1979	11	43
DI	28	2,640	6	10	1978	9	42
DI	21	7,920	20	10	1985	5	49
DI	27	7,920	24	10	1979	8	43
DI	27	7,920	24	10	1979	9	43
DI	27	7,920	24	10	1979	6	43
DI	27	7,920	24	10	1979	9	43
DI	27	5,280	12	10	1979	11	43
DI	24	10,560	12	10	1982	7	46
Steel	61	7,920	20	10	1945	47	9
CI	52	5,280	4	10	1954	43	18
CI	44	5,280	6	10	1962	33	26
CI	51	2,640	6	10	1955	37	19

(Project 3)

Material	Age	Length (ft)	Diameter (in.)	# of Break	Installation year	Wall Loss (%)	RUL
CI	46	5,280	6	10	1960	31	24
CI	52	5,280	6	10	1954	35	18
CI	41	7,920	6	10	1965	31	29
CI	44	5,280	6	10	1962	32	26
CI	46	2,640	6	10	1960	31	24
CI	57	5,280	6	10	1949	41	13
CI	57	5,280	6	10	1949	42	13
CI	56	2,640	6	10	1950	39	14
DI	34	21,120	8	20	1972	13	36
CI	34	5,280	10	20	1972	11	36
CI	44	5,280	10	20	1962	21	26
CI	48	2,640	6	20	1958	27	22
CI	48	5,280	6	20	1958	29	22
CI	57	5,280	4	20	1949	35	13
CI	57	2,640	6	20	1949	37	13
CI	98	5,280	10	20	1908	31	28
CI	79	5,280	16	20	1927	49	9
CI	55	5,280	4	20	1951	39	15
CI	50	2,640	6	20	1956	27	20
CI	55	5,280	6	20	1951	38	15
CI	86	2,640	6	20	1920	35	16
DI	25	5,280	10	20	1981	8	45
Steel	32	2,640	6	0	1974	13	38
AC	43	1,998	14	6	1969	29	27
AC	43	7,142.38	14	8	1969	27	27
DI	43	1,948.82	14	3	1969	31	27
AC	43	1,939	14	5	1969	32	27
AC	43	2,877	12	6	1969	35	27
AC	43	2,008	12	8	1969	33	27
DI	43	1,000.66	12	4	1969	31	27
DI	43	1,033.46	12	5	1969	37	27
Steel	43	2,286.75	12	6	1969	29	27
Steel	43	554.46	12	8	1969	27	27
Steel	43	3687.6	12	10	1969	35	27

Material	Age	Length (ft)	Diameter (in.)	# of Break	Installation year	Wall Loss (%)	RUL
Steel	1	236	12	0	2011	1	69
Steel	1	787.4	12	0	2011	4	69
Steel	1	623	12	0	2011	3	69
CI	43	1,424	12	8	1969	32	27
AC	43	2,129	12	5	1969	34	27
AC	43	1,434	12	10	1969	35	27
CI	43	3,281	14	6	1969	28	27
CI	43	1,384.5	14	5	1969	33	27
CI	43	15,403.5	14	8	1969	33	27
CI	43	4,537	14	8	1969	29	27
Steel	43	1,509	15	2	1969	37	27
CI	43	820	14	6	1969	28	27
steel	43	919	15	1	1969	31	27
CI	43	945	14	6	1969	33	27
CI	43	1,503	14	8	1969	31	27
steel	43	853	17	1	1969	37	27
CI	43	656	14	5	1969	29	27
steel	43	2,464	17	2	1969	39	27
CI	43	2,543	14	10	1969	28	27
Steel	13	1,330	12	0	1993	8	57
Steel	12	1,260	12	0	1994	9	58
Steel	11	1,420	12	1	1995	7	59
Steel	10	1,250	12	0	1996	6	60
Steel	9	1,220	12	1	1997	4	61
Steel	8	1,200	12	0	1998	3	62
Steel	7	1,320	12	1	1999	5	63
Steel	6	1,340	12	1	2000	4	64
Steel	5	1,410	12	1	2001	4	65
Steel	4	1,360	12	0	2002	3	66
Steel	3	1,310	12	1	2003	3	67
Steel	2	1,380	12	0	2004	5	68

(Project 4)

References

Agarwal, M., (2010), “Developing a Framework for Selecting Condition Assessment Technologies for Water and Wastewater Pipes.” Master’ Thesis, *Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 2010.*

Al-Barqawi, H., (2006), “Condition Rating Model for Underground Infrastructure Water Mains.” Master’ Thesis, Concordia University, Quebec, Canada.

Al-Barqawi, H., and Zayed, T. (2008) “Infrastructure Management: Integrated AHP/ANN Model To Evaluate Municipal Water Mains’ Performance” *J. Infrastructure Systems*, 14(4), 305 - 318.

Allison, P. D. (1999). Survival Analysis Using SAS®: A Practical Guide, SAS Institute Inc., Cary,

Clark, Robert M., Stafford, C. L., and Goodrich, J. A. (1982). “Water Distribution Systems: A Spatial and Cost Evaluation,” Journal of Water Resources Planning and Management, ASCE, 108(3).

Cooper, N.R., Blakey, G., Sherwin, C., Ta, Whiter, J.T., and Woodward, C.A. (2000), “The Use of GIS to Develop a Probability-based Trunk Main Burst Risk Model,” Urban Water, No.2, pp. 97-103.

Cortz, Hernan (2015), “A Risk Analysis Model for the Maintenance and Rehabilitation of Pipes in a Water Distribution System: A Statistical Approach.” California Polytechnic University, San Luis Obispo, master’s Thesis.

Covilakam, M., (2011), “Evaluation of Structural Monitoring Methods for Large Diameter Water Transmission Pipelines,” Master’ Thesis, the University of Texas at Arlington, (2011).

Daly, C., Zhao, Ch., Smith, M., Walt, G., (2016), “Not All Data Is Created Equal: Impact on Decision Making,” Proceedings of the Pipelines 2016 Conference, Kansas City, Missouri, July 2016.

Devera, Jan C. (2013). “Risk Assessment Model for Pipe Rehabilitation and Replacement in a Water Distribution System.” California Polytechnic State University, San Luis Obispo, Master’s Thesis.

Deshmukh, P., (2012), “Performance of Large Diameter Polyvinyl Chloride (PVC) Pipes in Water Applications,” Master’ Thesis, the University of Texas at Arlington, (2012).

Garrett, J. H., (1992), “Neural Networks and Their Applicability within Civil Engineering,” Computing in Civil Engineering, Proc. Of the Eight Conference, Dallas, Texas, 1992, pp. 1155-1162.

EPA, (2013), “Primer on Condition Curves for Water mains.” December 2013, <www. Epa.gov/research>.

Fahmy, M., (2009), “Integrated Multiple-Sensor Methodology for Condition Assessment of Water Mains.” Ph.D. Dissertation, Concordia University, Montreal, Quebec, Canada, (2009).

Fahmy, M., and Moslehi, O., (2009), “Forecasting the Remaining Useful Life of Cast Iron Water Mains,” Journal of Performance of Construction Facilities, 2009

Fares, H., and Zayed, T. (2008), “Hierarchical fuzzy Expert System for Risk of Failure of Water Mains.” Journal of Pipeline Systems Engineering and Practice, 1(1), 53 – 62.

Fares, H., and Zayed, T. (2010). “Hierarchical Fuzzy Expert System for Risk of Failure of Water Mains.” Journal of Pipeline Systems Engineering and Practice, 1(1).

Gershteyn, Y., and Perman L., (2003), MATLAB: ANFIS Toolbox, The Math works, available at: <http://www.mathworks.com/access/helpdesk/help/toolbox/fuzzy/fuzzy.shtml >

Haykin, S., (1999), "Neural Networks: a Comprehensive Foundation," Second Edition, Prentice-Ha, NJ.

Habibian, A., (2016), "Achieving Cost Efficiency through a Programmatic Approach to Condition Assessment," Proceedings of the Pipelines 2016 Conference, July 17-20, 2016.

Jang, J.S.R., (1993), "ANFIS: Adaptive Network-based Fuzzy Inference Systems," IEEE Transactions on Systems, Man, and Cybernetics, Vol 23, No.3, May/June 1993.

Jenkins, L., (2014), "Optimizing Maintenance and Replacement Activities for Water Distribution Pipelines," Ph.D. Dissertation, Vanderbilt University, Nashville, Tennessee, (2014).

Joshi, T., (2012), "An Evaluation of Large Diameter Steel Water Pipelines," Master' Thesis, the University of Texas at Arlington, (2012).

Karn, B., (2016), "A Quick Introduction to Neural Networks," available at:

< http://bbs.pinggu.org/thread-4933567-1-1.html>

Kettler, A.J., and Goulter, I.C. (1985). "An Analysis of Pipe Breakage in Urban Water Distribution Networks," Canadian Journal of Civil Engineering, 12, pp. 286-293.

Kleiner, Y. and Rajani, B. (2001). "Comprehensive Review of Structural Deterioration of Water Mains: Statistical Models," NRCC-42586. A version of this paper is published in Urban Water, v. 3 (3), Oct. 2001.

Kroon, D., 2001, "Corrosion Control for Water Systems-What You Need to Know but Were Afraid to Ask," Pipe Line International Conference, York, UK, pp.1-10

Kruse, R., (2008), “Fuzzy Neural Network,” available at: <http://www.scholarpedia.org/article/Fuzzy_neural_network>

Kulandaivel, G., (2004), “Sewer Pipeline Condition Prediction Using Neural Network Models,” Master’ Thesis, Michigan State University, 2004.

Lawrence, J. (1994), “Introduction to Neural Networks: Design, Theory and Applications,” California Scientific Software Press, Nevada City, CA.

Lin, B., and Huang, K., (2013), “an Adaptive-Network-Based Fuzzy Inference System for Predicting Springback of U-Bending,” Transactions of the Canadian Society for Mechanical Engineering, Vol. 37, No. 3, 2013.

Lou, Z., Lu, J. J. and Gunaratne, M. (1999). “Road Surface Crack Condition Forecasting Using Neural Network Models.” Florida Department of Transportation, Report WPI# 0510816, October 1999.

Makar, J.M. (2001). “Failure Modes and Mechanisms in Gray Cast Iron Pipe,” Presented at Underground Infrastructure Research 2001. Waterloo, Ontario. Natural Research Council Canada.

Manda, R., (2012), “Performance of Prestressed Concrete Cylinder Pipe (PCCP) In Water Applications,” Master’s Thesis, the University of Texas at Arlington, (2012).

MATLAB [Computer Software], (2017), MathWorks, Natick, MA.

Montgomery, D., Runger, G., (1994), “Applied Statistics and Probability for Engineers,” Wiley publication, Fourth edition.

Najafi, M. (2010), “Trenchless Technology Piping: Installation and Inspection,” McGraw-Hill, New York.

Najafi M., and Kulandaivel (2005), “Pipeline Condition Prediction Using Neural Network Models,” Pipeline Division Specialty Conference 2005, August 21-24, 2005 Houston, Texas, United States.

Najafi, M., and Gokhale, S. (2005), "Trenchless Technology: Pipeline and Utility Design, Construction and Renewal," McGraw-Hill, New York.

Najafi, M., (2016), "Pipeline Infrastructure Renewal and Asset management," McGraw-Hill, New York.

Nemeth, L., (2016), "A Comparison of Risk Assessment Model for Pipe Replacement and Rehabilitation in water Distribution System," Master' Thesis, California Polytechnic State University, San Luis Obispo, 2016.

Nishiyama, M., (2013), "Forecasting Water Main Failures in the City of Kingston Using Artificial Neural Networks," Master' Dissertation, Queen's University, Kingston, Ontario, Canada, October 2013, North Carolina, USA.

Paradkar, A., (2012), "An Evaluation of Failure Modes for Cast Iron and Ductile Iron Water Pipes," Master Thesis, the University of Texas at Arlington, (2012).

Parisher, R. A., and Rhea, R. A. (2012). "Chapter 2- Steel Pipe - Pipe Drafting and Design." Gulf Publishing Company, Elsevier Science Ltd., Houston, TX.

Providence infrastructure Consultant, (2016), "Southgate Water & Sewer System Master Plan," Southgate Water & Sanitation Districts, 2016.

Rajani, B.; Kleiner, Y., (2004). "Non-destructive Inspection Techniques to Determine Structural Distress Indicators in Water Mains," NRCC-47068, Evaluation and Control of Water Loss in Urban Water Networks, Valencia, Spain, June 21-25, pp. 1-20.

Rogers, P., (2011), "Prioritizing Water Main Renewals: Case Study of the Denver Water System," Journal of Pipeline Systems Engineering and Practice, ASCE, 2011.

Sacluti, F. R., (1999), "Modeling Water Distribution Failures Using Artificial Neural Networks," Master of Science Dissertation, Department of Civil and Environmental Engineering, Edmonton, Alberta, spring 1999.

Salman, B., (2010), “Infrastructure management and Deterioration Risk Assessment of Wastewater Collection Systems,” Ph.D. Dissertation, University of Cincinnati, May 2010.

Skipworth, P. Engelhardt, M. Cashman, A. Savic, D. Saul, A. Walters, G., (2002). “Whole Life Costing for Water Distribution Network Management.” Thomas Telford Publishing, London, UK.

Storm, T. S. and Rasmussen, S. C. (2011). “100+ Years of Plastic.” Leo Baekeland and Beyond, American Chemical Society, printed by Oxford University Press, Inc.

Suparta, W., Alhasa. K.M. (2016), “Modeling of Tropospheric Delays Using ANFIS,” Springer Briefs in Meteorology DOI 10.1007/978-3-319-28437.

Syachrani, S., (2010), “Advancement Sewer Asset Management Using Dynamic Deterioration Models,” Ph.D. Dissertation, Oklahoma State University, December 2010.

Tuhovcak, L., and Mika, P., (2013), “Indirect Condition Assessment of Water Mains.” 12th International Conference on Computing and Control for the Water Industry,” CCW12013, 2014.

Wade, M., (2016), “Strategic Integration of Risk-Based Asset Management to Improve Large- Diameter Pipelines,” Proceedings of the Pipelines 2016 Conference, July 17-20, 2016.

Walski, T.M., and Pellicia, A. (1982), “Economic Analysis of Water Main Breaks,” Journal of AWWA, 74, pp.140-147.

Wang, Y., (2006), “Deterioration and Condition Rating Analysis of Water Mains,” Ph.D. Dissertation, Concordia University, Montreal, Quebec, Canada, (2006).

Water Research Foundation (WRF), (2013), “Water Distribution System Risk Tool for Investment Planning.” Web Report # 4332. 2013.

Wilson, D., (2014), “Investigation of The Failure of Large-Diameter Cast Iron Water Mains Using A Stochastic,” Physical Model, Queen’s University, Kingstone, Ontario, Canada, October 2014.

Wurst, D., (2016), “A Risk-Based Condition Assessment of Process Piping for the San Jose-Santa Clara Regional Wastewater Facility,” Proceedings of the Pipelines 2016 Conference, July 17-20, 2016.

Zangenemadar, Z., Moslehi, O., (2016a), “Application of Neural Networks in Predicting the Remaining Useful Life of Water Pipelines,” Proceedings of the Pipelines 2016 Conference, July 17-20, 2016.

Zangenemadar, Z., Moslehi, O., (2016b), “Prioritizing Deterioration Factors of Water Pipelines Using Delphi Method,” 2016 Elsevier, 90 (2016) 491-499.

Zou, J. Han, Y, and So, S-S. (2008). “Overview of Artificial Neural Networks, Artificial Neural Networks: Methods and Applications,” D. J. Livingstone, ed., Humana Press, Totowa, New Jersey.

Biographical Information

Razieh Tavakoli graduated in 2009 with a Bachelor of Science in Architectural Engineering from Guilan University, Rasht, Iran. She completed her Master's Degree in Urban Design from Tehran Azad University, Tehran, Iran, in 2013. During her Doctoral studies at the University of Texas at Arlington (UTA), she was a teaching assistant for Construction Cost Estimating, Construction Contracts and Specifications, and Construction Management. She taught Timberline construction cost estimating software. Also during her time as a Ph.D. student, she was appointed as a graduate research assistant for a research project granted to the Center for Underground Infrastructure Research and Education (CUIRE) by the Texas Department of Transportation (TxDOT). She was fully funded throughout her Doctoral Program. Tavakoli served as an officer of the UTA student chapter of the North American Society of Trenchless Technology (NASTT) and has been an active ASCE student member.