



University of Nairobi

School of Engineering

Modelling Pipe Failure using Statistical models

By

Eliab Kimutai

F56/70101/2011

BSc: Soil, Water and Environmental Engineering, JKUAT

A thesis submitted in partial/fulfillment for the Degree of Master of Science in
Civil Engineering (Water Resources Option) in the Department of Civil and
Construction Engineering of the University of Nairobi

Thesis Supervisors:

Dr. Zablon Isaboke Oonge

Dr. Solomon Tesfamariam

June 2015

Declaration/ Approval

Declaration:

This thesis is my original work and has not been presented for a degree in any university

Approval:

This thesis has been submitted for examination with our approval as university supervisors.

i) First Supervisor

Dr. Zablon Isaboke Oonge

University of Nairobi

Signature.....

Date.....

ii) Second Supervisor

Dr. Solomon Tesfamariam

University of British Columbia

Signature.....

Date.....

Dedication

I am dedicating this Thesis to my wife, Abigail Bett, my sons Ryan Kipkoech and Jason Kiptoo for their love and constant encouragement through the bad and good times. Without the sacrifices they made, this study would not have been done.

Also, I am dedicating this thesis to my parents, Mr. and Mrs. Gideon Korir, my brother Erastus, my sisters Joan, Ascar and Orpa and my nephews Harrison, and Roy who believed and stood by me through the good and worst moments.

Acknowledgment

My sincere gratitude goes to Gandhi Smarak Nidhi Fund for awarding me scholarship to do my master's degree at the University of Nairobi.

I am also very grateful to the Canadian Commonwealth Scholarship Program (CCSP) for awarding me the scholarship to carry out this thesis in Canada. My sincere gratitude goes to the University of British Columbia, Okanagan Campus, for hosting me during this study period.

I owe my uttermost felt gratitude to my supervisors, Dr. Zablon Oonge and Dr. Solomon Tesfamariam, for their guidance and encouragement throughout the development of this thesis.

I sincerely express gratitude my friend, Getnet Betrie Dubale who offered me advice, motivation and encouragement to carry on with this thesis. May the Lord God bless you and your family abundantly.

I am indebted to Tanja Hogg for the healthy discussions we had on the implementation of R software. I am really grateful and my God continue to bless you.

I acknowledge the City of Calgary water utility for availing the pipe failure data for use in this study, for without it, this study would not have been complete.

Finally, I would like to thank God for the love and care that he has given me throughout this study period.

Table of contents

Declaration/ Approval	i
Dedication	iii
Acknowledgment.....	iv
Table of contents	v
List of Tables.....	viii
List of Figures	ix
List of Plates.....	x
List of Abbreviations.....	xi
List of Symbols	xii
Abstract.....	xiii
1 Introduction	1
1.1 Background.....	1
1.2 Problem statement.....	3
1.3 Objectives	4
1.4 Thesis structure	4
2 Literature Review.....	6
2.1 Introduction.....	6
2.2 Causes of failure.....	6
2.2.1 Pipe material.....	7
2.2.2 Pipe age.....	8
2.2.3 Pipe diameter.....	8
2.2.4 Seasonal variation.....	9
2.2.5 Temperature.....	10
2.2.6 Soil factors.....	10
2.2.7 Water Pressure.....	11
2.2.8 Construction and Repairs practices	12
2.2.9 Corrosion.....	12
2.3 Modes of pipe failures	13
2.3.1 Circumferential (Circular) breaks.....	13
2.3.2 Corrosion through holes.....	13
2.3.3 Leaks from joints and fittings.....	13
2.3.4 Bell splits.....	14

2.3.5	Blow-outs	14
2.3.6	Longitudinal breaks	14
2.4	Statistical models used in failure prediction	15
2.4.1	Physical analysis	15
2.4.2	Descriptive analysis	15
2.4.3	Predictive modeling	16
2.5	Survival Analysis	20
2.5.1	Overview	20
2.5.2	Observation Time and Censoring in Survival Analysis	21
2.6	Application of Statistical Models to Water Networks	22
3	Statistical Models	30
3.1.1	Cox-Proportional Hazard Model (Cox-PHM)	30
3.1.2	Weibull Proportional Hazard Model (WPHM)	34
3.1.3	Poisson process	36
3.2	Model Development	37
3.2.1	Input Processing	38
3.2.2	Model Variable Selection	39
3.2.3	Model Training and Validation	40
3.2.4	Model Evaluation	40
4	Results and Modelling Approach	42
4.1	Study Area Description	42
4.1.1	Weather Data	43
4.2	Data Preparation for Modelling	44
4.2.1	Data Extraction	45
4.2.2	Data Cleaning	46
4.3	Pipe Data Analysis	49
4.3.1	Distribution of Pipe Materials for the CCWN	49
4.3.2	Pipe Laying for CCWN	51
4.3.3	Break Rates for Different Pipe Materials	53
4.3.4	Effects of weather parameters on breaks	59
4.3.5	Pipe Diameter in Relation with Material	62
4.3.6	Spatial Distance and Break Occurrence	67
4.3.7	Influence of Pipe Previous Failure History	69

4.3.8	Influence of Type of Service Area on Number of Breaks	71
4.3.9	Cathodic Protection.....	74
5	Discussion.....	76
5.1	Model Variable Selection	76
5.2	Model Training.....	77
5.2.1	Effects of Covariates on Pipe Failure	77
5.2.2	Model Break Prediction	81
5.2.3	Checking for Model Appropriateness for Cox-PHM and WPHM.....	86
5.3	Baseline Hazard Estimation	89
5.4	Estimation of Time to Failure of a Pipe.....	90
5.4.1	Survival Curves	92
5.4.2	Mean Time to Failure (MTF)	96
5.5	Sensitivity Analysis for different Break Types.....	97
5.6	Model Verification	100
5.7	Model Evaluation	105
6	Conclusion and Recommendations.....	107
6.1	Conclusion	107
6.2	Recommendation.....	109
7	References	110
8	Annexes.....	117
8.1	A-1: Climatic Data	117
8.2	A-2: Cox Proportional Hazard Assumption.....	121

List of Tables

Table 2-1: Factors affecting failure rate of pipes.....	7
Table 2-2: Summary of Survival Models.....	27
Table 2-3: Summary of statistical model applications to water networks	28
Table 4-1: Number of breaks experienced by a pipe with respect to material type	70
Table 5-1: Number of pipes for training and testing of the models in each data set.....	76
Table 5-2: Selected covariates and corresponding coefficients Cox-PHM and WPHMfor DI strata	77
Table 5-3: Selected covariates and corresponding coefficients for Cox-PHM and WPHM for CI strata	78
Table 5-4: Selected covariates and corresponding coefficients for Cox-PHM and WPHM for PVC strata.....	78
Table 5-5: Parameter values for different covariates from past studies.....	81
Table 5-6: Estimated cumulative hazards equations for DI, CI and PVC pipe strata from Cox-PHM	90
Table 5-7: Number of pipes in each of the defined strata in pipe data set.....	91
Table 5-8: Significant covariates and corresponding coefficients for CI pipe strata.....	91
Table 5-9: Significant covariates and corresponding coefficients for DI pipe strata	91
Table 5-10: Mean Time to Failure for DI and CI pipes ranked on previous failures	97
Table 5-11: Sensitivity analysis for ALL break-types and corrosion break-type on CI pipe .	98
Table 5-12: Sensitivity analysis for Circular and Leak break-type on CI pipe	98
Table 5-13: Sensitivity analysis for ALL and Corrosion break-type for DI pipe.....	99
Table 5-14: Performance of Cox-PHM, WPHM, and Poisson on Break Prediction for DI, CI and PVC strata	105

List of Figures

Figure 1-1: Thesis Structure	5
Figure 2-1: Failure rate per kilometre per Year.....	9
Figure 2-2: Occurrence of pipe failures in a year	10
Figure 2-3: weather-soil Type failure rate interaction-Minimum API.	11
Figure 2-4: Observation Time for a Pipe	22
Figure 3-1: Schematization of model development.....	37
Figure 4-1: Water distribution system of the City of Calgary.....	42
Figure 4-2: Monthly temperature pattern for the City of Calgary	43
Figure 4-3: Monthly precipitation for the City of Calgary.....	44
Figure 4-4: Data Extraction Flowchart	46
Figure 4-5: Multiple break records of a single pipe occurring in one day	47
Figure 4-6: Conflicting records of lengths, diameters, installation year and material type for one pipe	48
Figure 4-7: Proportion of pipe material for CCWN.....	50
Figure 4-8: Pipe laying in different Pipe Manufacturing Periods.....	51
Figure 4-9: Annual break rate for CI pipe material	53
Figure 4-10: Annual break rate for DI_A Pipe Material.....	55
Figure 4-11: Annual break rate for PVC pipe material.....	56
Figure 4-12: Annual Break Rate for CCWN	58
Figure 4-13: Monthly break-rate for CCWN utility	59
Figure 4-14: The effects of temperature on pipe break rate of CCWN utility	61
Figure 4-15: Pipe material in relation to diameters in the utility.....	63
Figure 4-16: Relationship between pipe material and break-type	64
Figure 4-17: relationship between pipe diameter and break-type	67
Figure 4-18: Spatial distance in relation to pipe breaks.....	69
Figure 4-19: Pipe breaks observed in commercial units	72
Figure 4-20: Pipe Breaks Observed in Residential Units.....	73
Figure 4-21: Relationship between retrofitted pipes and material type	75
Figure 5-1: Prediction of Breaks for CI Strata from CI Data Training Set.....	82
Figure 5-2: Prediction of Breaks for DI Strata from DI Data Training Set	83
Figure 5-3: Prediction of Breaks for PVC Strata from PVC Data Training Set.....	85
Figure 5-4: Schoenfield’s residuals for the CI pipe strata with no previous breaks.....	87
Figure 5-5: Schoenfield’s residuals for the CI pipe strata with more than one break	87
Figure 5-6: WPHM appropriateness using graphical method for pipes with more than one break.....	88
Figure 5-7: WPHM appropriateness using graphical method for pipes with no previous breaks	88

Figure 5-8: Observed and predicted CH for pipes with PRB =0 for CI, DI_A and PVC pipe strata	89
Figure 5-9: Observed and predicted CH for pipes with PRB >0 for CI and DI_A pipe strata	90
Figure 5-10: Survival curves for YDI and DI pipe	93
Figure 5-11: Survival curves for thick and thin CI pipe	93
Figure 5-12: Survival curves for CI pipe with different ordered breaks.....	95
Figure 5-13: Survival curves for DI pipe with different ordered breaks.....	96
Figure 5-14: Observed and Predicted number of breaks for DI pipe strata	102
Figure 5-15: Observed and Predicted number of breaks for CI pipe strata.....	103
Figure 5-16: Observed and Predicted number of breaks for PVC pipe strata.....	104
Figure 8-1: Schoenfield’s residuals for the DI pipe strata with less than one break	121
Figure 8-2: Schoenfield’s residuals for the DI pipe strata with more than one break	121
Figure 8-3: Schoenfield’s residuals for the DI pipe strata with more than one break	122
Figure 8-4: Schoenfield’s residuals for the PVC pipe strata.....	122

List of Plates

Plate 4-1: Cement Lining (a) and Anode Protection (b) for the YDI Pipe.....	52
Plate 4-1: Corrosion band located from 3’5” from end 9A.....	65
Plate 4-2: Corrosion band located from 3’4” from end 41A.....	66

List of Abbreviations

AIC- Akaike Information Criterion
CCSP-Canadian Commonwealth Scholarship Program
CCWN- City of Calgary Water Network
CI- thick and thin Cast Iron pipe
CI_TN- Thin Cast Iron pipe
CI_TK- Thick Cast Iron pipe
Cox-PHM- Cox Proportional Hazard Model
CORR- Corrosion type of break
CIRC- Circular type of break
DI_A- all ductile iron pipes
DI- wrapped and unwrapped Ductile Iron pipe
FI- Freezing Index
GIS- Geographical Information System
MAE- Mean Absolute Error
MAT- Pipe material
MLE- Maximum Likelihood Estimator
MTF-Mean Time to Failure
PHM-Proportional Hazard Model
PRB- Previous Number of Failures
RAE-Relative Absolute Error
RD- Rain Deficiency
REST- Soil Resistivity
RMSE-Root Mean Square Error
RRSE -Root Relative Squared Error
WPHM- Weibull Proportional Hazard Model
YDI- Yellow-Jacketed Ductile Iron pipe

List of Symbols

z - Pipe covariate

$h_o(t)$ - Baseline hazard function

δ_i - Event indicator, either an occurrence of a break or not

L_p - Partial likelihood function

S - Survival function

W - Error term

T - Time to failure of a pipe

σ - Weibull shape parameter

γ - Weibull intercept parameter

β - Coefficient of covariates

h - Hazard function

\mathbb{R} - Risk set of pipes

N - Number of pipe failures for Poisson model

λ - Rate of Poisson process

$E[N_i(t)]$ - Poisson expected value

Δ - Cumulative baseline hazard function

L - Maximum likelihood function

Abstract

The degradation of pipe factor of safety upon pipe installation has increased interest in buried infrastructure asset management. Several modelling approaches using statistical models have been proposed to explain the effects of covariates in the failure of water pipes.

In this thesis, Cox Proportional Hazard Model (Cox-PHM) was used in the prediction of number of break. Curve fitting techniques were proposed for estimation of baseline hazard function and the resulting equation applied in break prediction. The results from the model were compared to the results from Weibull Proportional Hazard Model (WPHM) and Poisson Model. Further, Cox-PHM was used to determine the time to failure of metallic pipes.

Results indicated that physical factors e.g. diameter, were the critical factors impacting pipe failure and the occurrence of a particular break-type. Further, results indicated that the effects of covariates differ according to material type with PVC and DI pipe showing low and high vulnerability to breaks, respectively. Additionally, when mean time to failure (MTF) of a pipe was analyzed, it revealed that after the occurrence of a failure, time to failure for DI decreases significantly compared to CI pipe. Results indicated that when a pipe has had four breaks, it should be considered for replacement.

The prediction results from the models revealed that different models depending on pipe material can be used to model the evolution of breaks for the City of Calgary Water Network (CCWN). Therefore, Poisson Model and WPHM performed best in the prediction of the PVC, and both DI and CI pipes, respectively. Results from Cox-PHM indicate that the estimation of the baseline hazard function using curve fitting techniques captures the trend of metallic pipes especially for the young water networks.

It is therefore recommended that a combination of models should be used based on the rate of deterioration and material type of the system rather than a single model.

Chapter 1

1 Introduction

1.1 Background

Water mains form an essential part of the lifeline systems of modern urban towns and cities (Rajani et al. 1996). In the early 1900's cast iron was used as the main distribution system until the introduction of ductile iron in 1970's. PVC was introduced in 1970's but has seen the introduction of medium and high polyethylene in 1990's (MDPE & HDPE) which have become alternatives for replacing the existing water mains (Brander 2001; Rajani et al. 1996). Asbestos cement pipes were used extensively in 1930's through 1960's (Hu et al. 2008) but were later discontinued in 1970's due to health concerns associated with the manufacturing processes and its release of asbestos fibres as it deteriorates (Gibbons et al. 2008; Hu et al. 2008).

Nonetheless, occurrences of the bursts/ breaks from these water networks are observed regularly (Berardi et al. 2008). These occurrences have stimulated development of buried infrastructure asset management strategies (Clair and Sinha 2012; Rogers 2011) to explain these breaks (Hu and Hubble 2007; Rajani et al. 1996). To fully implement these management strategies, a complete understanding of the asset condition and performance over time is required (Clair and Sinha 2012).

Pipe breaks are caused by environmental conditions, physical and operational factors (Berardi et al. 2008; Rajani and Kleiner 2001) that combine in degrading the pipe factor of safety (FS) (Rajani and Kleiner 2001). It occurs when FS critical limit (usually 1) (Rajani and Tesfamariam 2005) has been exceeded. The consequence of these failures are financial burdens, water service interruptions, traffic delays, loss of water and energy among others (Berardi et al. 2008).

To successfully manage the pipe assets, time dependent deterioration and failure of a pipe needs to be understood and estimated (Davis et al. 2008). As a result, many modelling approaches that provide support in planning and renewal decisions of the water systems have been

developed (Rogers 2011) and published (e.g. Le Gat and Eisenbeis 2000; Gowlter and Kazemi 1989; Mailhot et al. 2000; Pelletier et al. 2003; among others).

Kleiner and Rajani (2001) and Rajani and Kleiner (2001) classified models into physical/mechanistic and statistical models. Physical models predict failure by assessing the load the pipe is subjected to as well as the ability to resist the loads (Rajani and Kleiner 2001; Rajani and Tesfamariam 2004). Physical models require physical, environmental and operational data that are often difficult to obtain due to time and financial constraints (Kleiner and Rajani 2001; Marlow et al. 2010). Examples of physical models include simplified Winkler model (Rajani and Tesfamariam 2004; Rajani et al. 1996), corrosion models (Rajani and Makar 2001; Rajani and Tesfamariam 2005), and time-to-failure (Rajani and Tesfamariam 2004, 2005, 2007; Tesfamariam and Rajani 2004).

On the other hand, statistical models offer an alternative that is both economical and effective for analyzing structural state of pipes (Kleiner and Rajani 2002). They model pipe failure by utilizing information from past failure data to identify the breakage patterns which are assumed to continue into the future (Kleiner and Rajani 2001). Statistical models are further classified as deterministic and probabilistic models. Deterministic models use two to three variables based on pipe age or history to predict break rates (Kleiner and Rajani 2001). In this thesis the terms “covariates”, “explanatory variables”, “variables”, and “risk factors” are used interchangeably to mean factors that increase the likelihood of a pipe failure.

Probabilistic models, either single or multi-variate models, focus on time-to-failure of a pipe by estimating the probability that a break will occur at some future time (Andreou 1986). Single-variate probabilistic models (e.g. Poisson) are used in; i) predicting failure patterns of pipe groupings or cohorts (Fuchs-Hanusch et al. 2011; Martins et al. 2013), ii) deriving probabilities in life expectancy, and iii) probabilistic analysis of break clustering (Kleiner and Rajani 2001). Multi-variate models are used in predicting individual pipes breaks rather than cohorts (Kleiner and Rajani 2001).

The widely used multi-variate approach is survival analysis which utilizes right censored data (Pelletier et al. 2003) with Proportional Hazard Model (PHM) developed by Cox (1972) being the commonly used. PHM can either be parametric or semi-parametric models. Semi-parametric models assumes the baseline hazard function is left undefined and therefore not

estimated while for parametric models, the baseline hazard function is allowed to follow a certain distribution e.g. Weibull (Røstum 2000).

Therefore, through application of these statistical models and making use of all the explanatory variables from the data, the expected number of future breaks events can be estimated with reasonable accuracy for each individual pipe (Andreou et al. 1987). The output from estimation of future breaks will be very important in; i) determining the budget needs for future repairs; ii) establish the optimum time of replacement of the breaking mains, and; iii) obtain the reliability estimates of the individual pipe (Andreou 1986; Andreou et al. 1987a; Boxall et al. 2007).

1.2 Problem statement

As pipe ages, its structural integrity diminishes and become susceptible to breaks and leaks (Toumbou et al. 2012) leading to loss of water, increased cost of system maintenance, reduced the quality of service; loss of water quality and increased traffic congestion (Berardi et al. 2008). However, by maintaining the records of these breaks and leaks, the current and past structural state of the utility can be analysed (Toumbou et al. 2012).

Planning and renewal decisions in the water systems have been made by applying statistical models (Rogers 2011). Some modelling approaches have applied PHM in making these decisions with the focus being on the parametric forms of the model (Le Gat and Eisenbeis 2000; Martins et al. 2013; Røstum 2000). However, semi-parametric models have been applied in evaluating the effects of covariates on the hazard (Christodoulou 2010; Kumar and Klefsjö 1994; Park et al. 2008) with limited use in prediction of number of breaks. This limitation is as a result of baseline hazard function which is left undefined (Cox 1972). The estimate of this function is of great importance in validation of the model using independent data (Royston and Altman 2013; Royston 2011).

In this thesis, the use of semi-parametric model, Cox Proportional Hazard Model (Cox-PHM), is applied in prediction of number of pipe breaks. The proposed methodology allows for the estimation of baseline hazard function using curve smoothing techniques and the resulting equation applied in break prediction. Further, water managers often are in need of a simple model that can combine the ease of use as well as predict the failure pattern of a water network

accurately by using all the risk factors in the data (Andreou et al. 1987a). Therefore, the prediction results from the semi-parametric model are compared with output from the Weibull Proportional Hazard Model (WPHM) and the Poisson Model (PM) to in order to determine a simple and robust model for use in predicting the evolution of pipe breaks.

1.3 Objectives

The main objective of this study is to evaluate the suitability of statistical models in prediction of failures of each individual pipe in a water distribution network.

Specific objectives to:

1. Establish what covariates are important in the failure of a pipe
2. Reveal which covariates influence occurrence of a particular break-type
3. Determine the robust model that best represent a failure pattern of the city of Calgary
4. Determine the suitability of Cox-PHM in break prediction

1.4 Thesis structure

This thesis is organized as given in Figure 1-1. Section 2 presents the literature review of pipe failure giving details of causes and mechanism of failure is presented and a review of modelling strategies. In Section3, a review of the statistical background of the models used in this study is presented. Additionally, parameter estimation and failure prediction of each model is outlined and a section of model development is also presented. A case study and results are given in Section 4, while discussion of results is presented in Section 5. Finally, Section 6 presents the concluding remarks and recommendations

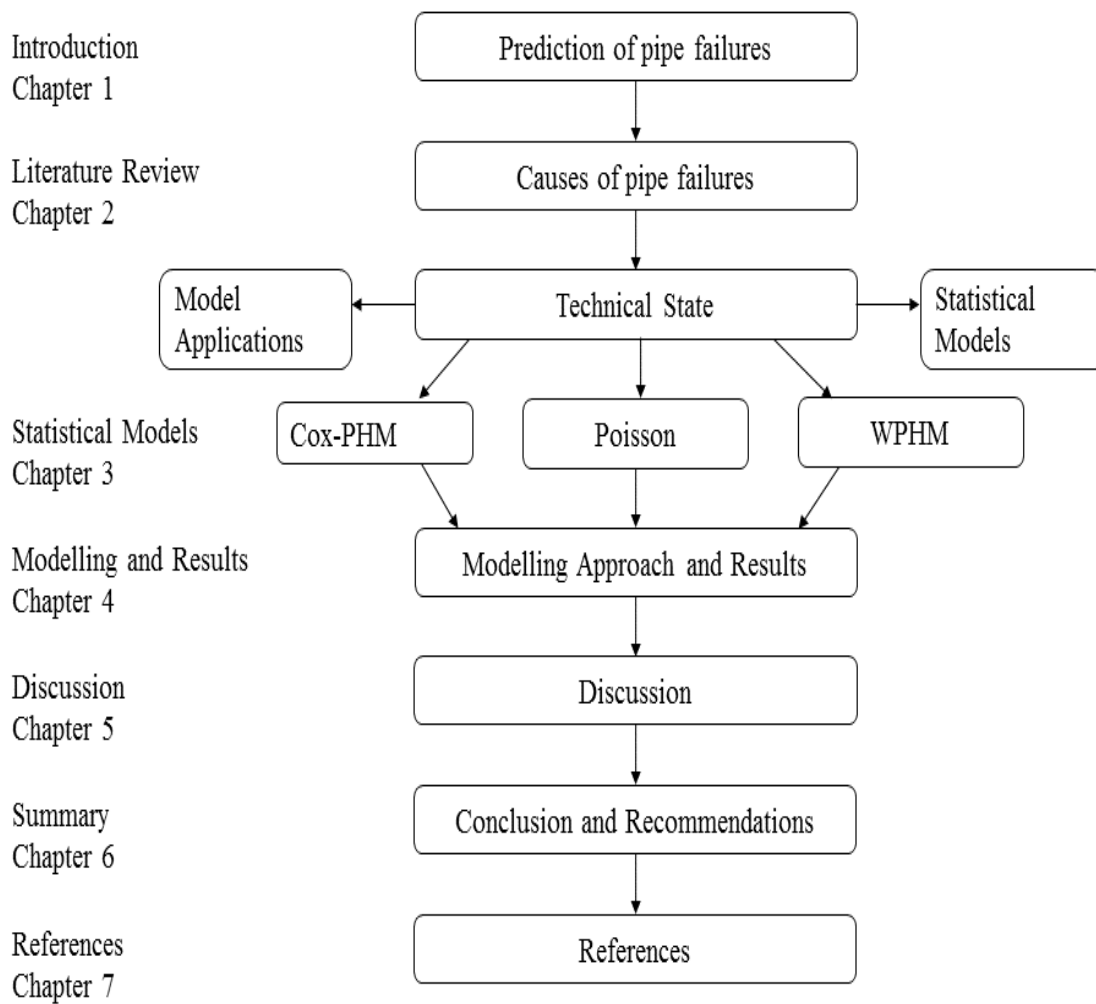


Figure 1-1: Thesis Structure

Chapter 2

2 Literature Review

2.1 Introduction

The purpose of this section is to give an insight and document critical factors responsible for the deterioration of pipes and how these factors contribute in planning for the maintenance of an efficient pipe system. In Sections 2.2 and 2.3, discussion on the causes of pipe failures and their mechanism is presented. The review of models used to describe the state of pipes is presented in section 2.4 and finally in Section 2.5, a brief overview of applications of statistical models to pipe failure is presented.

2.2 Causes of failure

Pipe failures are a regular occurrence in water distribution systems (Berardi et al. 2008). These failures are caused by a number of factors classified by Rajani and Kleiner (2001) as:

- (i) Physical characteristics of pipes
- (ii) Environmental characteristics and,
- (iii) Operational characteristics

The combination of these factors determines the pipe failure processes and modes, though the roles played by each factor vary from water main to water main owing to site-specific conditions (Hu and Hubble 2007). Table **2-1** gives a summary of the factors failure rates of the pipe.

Table 2-1: Factors affecting failure rate of pipes

Physical characteristics	Operational factors	Environmental characteristics
Pipe diameter	Maintenance and repair	Soil factors
Age	Pressure surge	Groundwater level
Length	Water pressure	Seasonal climatic variations
Material type	Water quality	Stray current
Depth	Water velocity	Ground movement
Pressure class	Previous failure history	Road surface
Pipe wall thickness	Cathodic protection	Traffic and surface loads
		Leakage rate

(source: Røstum (2000))

The following subsections describe the factors which are commonly assumed to have the greatest impact on the pipe failure.

2.2.1 Pipe material

Several pipe materials exist, ranging from concrete pipes (e.g. asbestos cement), metallic pipes (e.g. cast iron, ductile iron), PVC, medium density polyethylene (MDPE) and high density polyethylene (HDPE) (Rajani et al. 1996). Studies have been carried out by many researchers (e.g. Hu and Hubble (2007); Kettler and Goulter (1985), among others) to determine the influence of these pipe materials in pipe failure rates.

A study by Pelletier et al. (2003) revealed a close dependency exists between pipe material and year it was laid. Material strength of pipe determines the breaking patterns of pipes with cast iron displaying high breaking trend compared to ductile iron (Kuraoka and Rainer 1996) because of its low strength (Makar et al. 2001). Additionally, age has been found to contribute to low material strength of pipe, with increase in number of breaks being observed as the pipe ages (Kettler and Goulter 1985; Mordak and Wheeler 1988). This trend seems to affect the metallic pipes and asbestos cement pipes with plastic pipes exhibiting different patterns. PVC have been reported to withstand aggressive environments, abrasion and strong acids, with decrease in failure rate having been reported in one case (Davis et al. 2004; Knight 2002).

2.2.2 Pipe age

Disagreements exist in literature on the influence of age in pipe break. Some authors have observed a linear relationship between age and pipe breaks (Harris and Radlinski 2013; Hu and Hubble 2007; Hu et al. 2010; Kettler and Goulter 1985) while others have observed a non-linear relationship (Andreou 1986; Boxall et al. 2007; Gowlter and Kazemi 1989; O'Day 1982). Those who argue non-linear relationship agree that age should contribute to pipe failure later but in some circumstances, high number of early breaks is observed due to quality of materials used and localized unfavourable environmental pipe conditions. (Boxall et al. 2007) recommended that age should be combined with the knowledge of the pipe network condition and weak points to allow accurate assessment.

2.2.3 Pipe diameter

Diameter of the pipe has been found to contribute significantly to the number of breaks observed. An inverse relationship between the breaking rate of the pipe and the diameter (Figure 2-1) has been reported by many authors (Andreou et al. 1987b; Hu and Hubble 2007; Kettler and Goulter 1985; Mordak and Wheeler 1988; White 1985). Most of the breaks have been observed to occur to small diameter pipes with 150-200mm diameter contributing to up to 94% of all the failures (Hu and Hubble 2007). The vulnerability of these small diameter pipes is thought to result from thinner walls, low bending inertia and less reliable joints (Kettler and Goulter 1985).

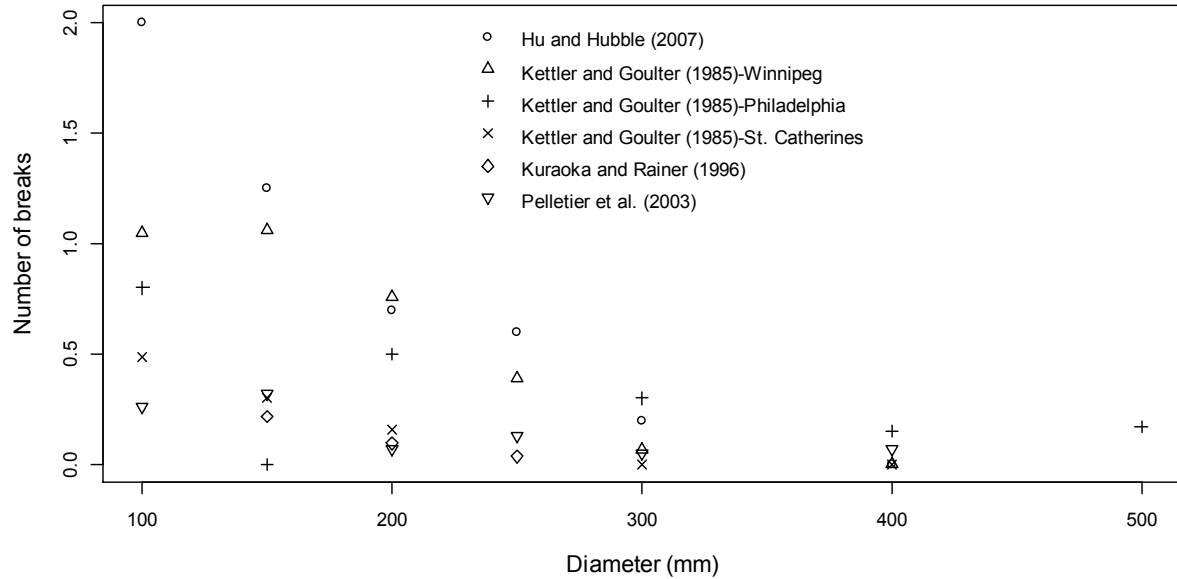


Figure 2-1: Failure rate per kilometre per Year

2.2.4 Seasonal variation

Pipe failures have been observed to follow seasonal patterns and have been reported to either occur in summer (Hu and Hubble 2007; Mordak and Wheeler 1988; White 1985) or in winter (Friedl et al. 2012; Harvey et al. 2013; Hu and Hubble 2007) as shown in Figure 2-2.

During summer, the rainfall deficit increases and reaches its maximum has it ends, causing a high shrinkage of clay soil (Hu and Vu 2011). This situation result in low attenuation of imposed loads causing higher loads to be transmitted to the pipes (Hu and Hubble 2007; Mordak and Wheeler 1988). During winter period, cold temperatures causes freezing and expansion of water in the ground, thereby increasing earth loads on the buried pipes (Harvey et al. 2013; Hu and Hubble 2007).

Further, frozen soils above the pipe put an extra load on the pipe bedding soil support causing it to weaken (Hu and Vu 2011). Harvey et al. (2013) reported that about 60% of the failures reported in Scarborough are strongly related to winter season.

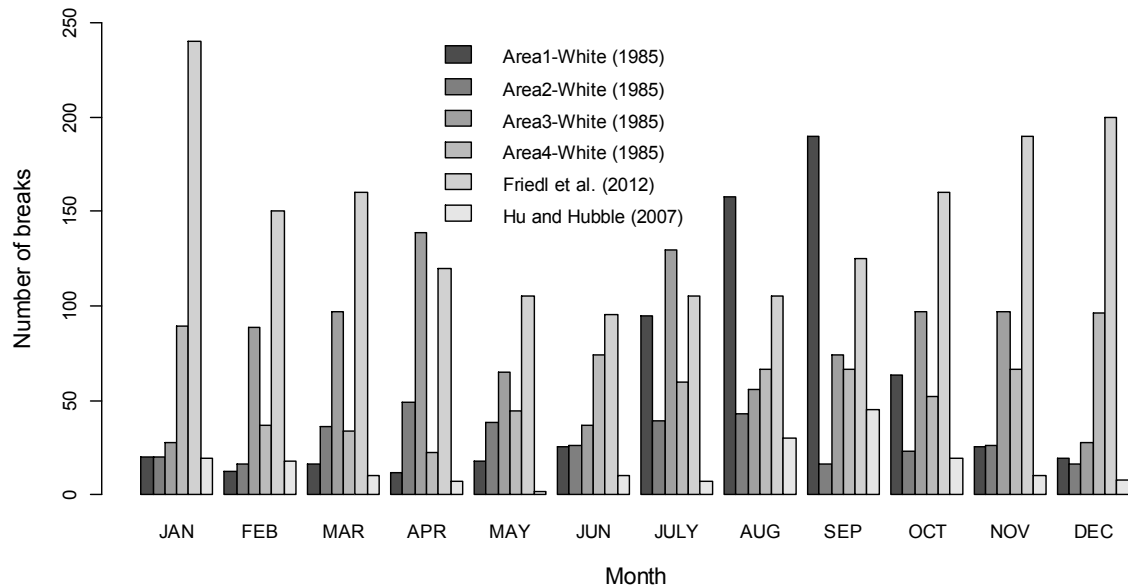


Figure 2-2: Occurrence of pipe failures in a year

2.2.5 Temperature

The importance of temperature in influencing pipe failure has been documented by many authors (e.g. (Habibian 1994; Rajani et al. 1996, 2012) among others). The influence of temperature differs depending on the type of pipe material (Rajani et al. 2012) with PVC, though with high coefficient of thermal expansion, showing little effect due to their flexibility (American Water Works Service Co. 2002).

However, the thermal differences between metallic pipe and water results in high breaks (American Water Works Service Co. 2002). For example as the temperature falls, and at approximately 4°C, water begins to expand while the pipes continue to contract. As a result, the expanding water exerts pressure on the pipe resulting in breaks (American Water Works Service Co. 2002; Hu et al. 2010). Additionally, when temperature hits a new low, an increase of breaks occurs affecting the weaker pipes that were not affected in the last period (Habibian 1994).

2.2.6 Soil factors

The interaction of soil type and weather elements has been documented in many studies (e.g. (Gould et al. 2009; Hu and Vu 2006; Mordak and Wheeler 1988) to result in soil movement leading to transmission of loads to the pipe. This trend is commonly observed in very expansive

soil (clay) as compared to stable soils (Mordak and Wheeler 1988) as can be seen in Figure 2-3. Most of the failures associated with soil movement are circumferential breaks (Rajani et al. 1996) and have been observed to occur during dry months (Gould et al. 2009; Hu and Hubble 2007; Mordak and Wheeler 1988)

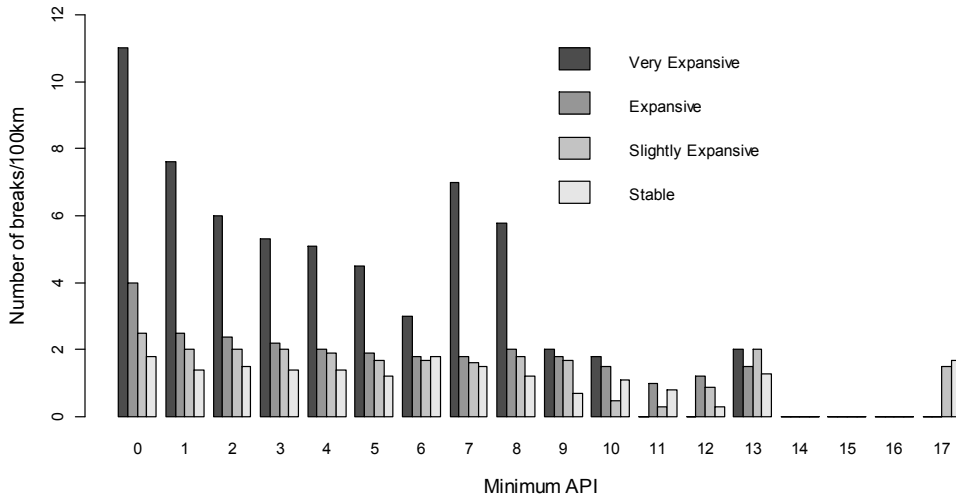


Figure 2-3: weather-soil Type failure rate interaction-Minimum API.

Source: (Gould et al. 2009)

2.2.7 Water Pressure

A pipe is designed with a wall thickness in order to withstand standard internal pressure and external loadings (Hu et al. 2010). However, the load bearing capability decreases as the pipe deteriorates since the degraded part of the pipe becomes much weaker than the intact portions (Hu et al. 2008).

A linear correlation between water pressure and pipe breakage rates have been observed (Hu et al. 2010) and is thought to contribute to occurrence of circumferential breaks (Kleiner and Rajani 2000) and blowout of the weakest portion of the mains wall (O’Day 1982). Andreou et al. (1987b) found that high internal pressure imposes a loading on a pipe especially on corroded pipe thus increasing the likelihood of occurrence of multiple breaks.

2.2.8 Construction and Repairs practices

2.2.8.1 Construction practices

Construction practices, such as inadequate compaction of backfill has been identified as the cause of subsidence of the backfill after reinstatement (Hu and Hubble 2007). Differential soil movement and uneven bedding caused by differential subsidence between the backfill and the surrounding native soil which produces shear and bending loads on the buried pipe (Hu and Hubble 2007; O'Day 1982). Bedding material may indicate a particular construction practice that induces physical stress in a pipe ultimately causing its failure (Wood and Lence 2009). Hu and Hubble (2007) found that about 5.1% of the failures were attributed to inadequate compaction of backfill especially in pipes connected to other structures e.g. hydrants

2.2.8.2 Repairs practices

A linear relationship between pipe break and pipe repairs has been observed by some authors e.g. (Goulter and Kazemi 1989; Hu and Hubble 2007) with about 21.6% of all breaks occurring from zero distance of the previous failure. This trend has been attributed to disturbance of surrounding soil as a result of failure itself and the repair practices. Additionally, a leakage from the first breaks can cause erosion of the bedding material or change the moisture conditions of the surrounding environment, therefore influencing expansion and contraction of the soil (Goulter and Kazemi 1988; Hu and Hubble 2007; O'Day 1982). Furthermore, in all failures associated with re-breaks, almost 40% were from previous repairs during winter when the temperatures were below 0°C (Hu and Hubble 2007).

2.2.9 Corrosion

Upon installation, a buried pipe is expected to maintain its original factor of safety for the rest of its life (Rajani and Makar 2001) but, this is not the case as it diminishes due to corrosion that initiates randomly and grows over time (Rajani and Tesfamariam 2005). Corrosion in pipes is thought to be enhanced by, among other factors, the interaction between the soil properties (e.g. moisture, redox potential, pH, oxygen and microbial activity) and the environmental conditions (Liu et al. 2010).

Karaa (1984) and Makar (2000) noted that corrosion was a major contributor to pipe failures and becomes serious as the pipe ages, causing decrease in pipe wall thickness and pipe strength thereby, increasing the likelihood of pipe breakage. This observation was also documented in Mordak and Wheeler (1988) who found that, 90% of failures of AC pipes with 50 or more years were attributable to corrosion. In contrast, other studies have suggested that over time, the rate of corrosion is slowed due to production of red oxides that inhibits the rate of corrosion (Rajani and Tesfamariam 2005).

2.3 Modes of pipe failures

This subsection gives a brief summary on the types of break that are common in water pipes.

2.3.1 Circumferential (Circular) breaks

Literature has pointed out that the major causes of circular break-type are soil movement (Gould et al. 2009; Rajani et al. 1996) and temperature changes (Habibian 1994), and have been reported to occur mostly when there is decrease in soil moisture content (Gould et al. 2009). This break-type causes an average of 70% of all pipe failures (Rajani et al. 1996) and occurs mostly in small diameter pipes with CI pipe being the most affected (Makar et al. 2001). The vulnerability of the small diameter pipes to breaks is contributed by their thin walls and the small moments of inertia (Kettler and Goulter 1985; Mordak and Wheeler 1988).

2.3.2 Corrosion through holes

Corrosion break-type has been identified to be a major cause of breaks of water pipes (Makar 2000). This break-type is thought to result from continuous damaging reactions that reduces the pipe factor of safety (Rajani and Tesfamariam 2007). Apart from majorly affecting the CI pipe (Makar 2000), literature has shown that unprotected DI pipe is very vulnerable to corrosion and can corrode at a rate that equals that of CI pipe (Rajani and Kleiner 2003). Some have noted that DI pipe is fails majorly as a result of corrosion (Makar 2000). A linear relationship between this break type and pipe age has been noted and to some extent an inverse relationship being observed (Mordak and Wheeler 1988; Rajani and Tesfamariam 2005).

2.3.3 Leaks from joints and fittings

Leaks have been also been identified as the causes of failures in a water network and is caused by leaking joints, line valves, hydrant branches and services (Karaa and Marks 1990). Joint

leaks result from separation of the main sections caused by expansion and contraction of the pipe while other leaks may be a resultant of corrosion and stress impacts (Karaa and Marks 1990). Literature have indicated that the occurrence of leaks in the water network could lead to erosion of the bedding material, cause difference in moisture conditions in pipe environment (Gowltter and Kazemi 1989; Hu and Hubble 2007) which may result in pipe failing as a result of circumferential (circular) failure (Karaa and Marks 1990).

2.3.4 Bell splits

This type of break, associated with smaller diameter pipes (Makar et al. 2001), is thought to result from stress that develops when there is pipe temperature is higher or lower than when it was installed (Makar 2000). Some studies have indicated that the occurrence of this break-type is preceded by pipe corrosion which initiates a point of weakness in a pipe (Makar 2000).

2.3.5 Blow-outs

This type of break has been taken as an indicator of the occurrence of corrosion in a water network (Makar et al. 2001; O'Day 1982). It occurs when pipe strength has been weakened by corrosion (Makar 2000; Makar et al. 2001; O'Day 1982). As a result, a sudden increase in water pressure results in hoop stress causing blow-outs at the weakest sections of the pipe (O'Day 1982).

2.3.6 Longitudinal breaks

Longitudinal break-type is associated with large diameter pipes is caused by excessive internal pressure of the water, crushing and compression forces acting on the pipe (Makar et al. 2001). Additionally, pipe thinning is thought to cause this break (Makar 2000). The result is the crack that may extends the entire length of the pipe (Makar 2000; Makar et al. 2001)

2.4 Statistical models used in failure prediction

Since early 1980's, several attempts have been made to model statistically the failure risk of water mains (Le Gat and Eisenbeis 2000). Techniques employed in assessment of pipe condition can be classified into three areas: descriptive analysis, physical analysis, and predictive models (Karaa and Marks 1990). Predictive modelling uses statistical models to predict future system failures (Kleiner and Rajani 2001; Pelletier et al. 2003; Røstum 2000).

2.4.1 Physical analysis

Physical analysis employs engineering based equations to develop structurally based estimates of water main conditions (Karaa and Marks 1990). They are used to predict the service lifetime of an individual asset rather than the failure rate for the cohorts using the load bearing capacity of an asset (Marlow et al. 2010). The prediction involves understanding the degradation mechanism that influences the load bearing capacity of an asset and its service lifetime. Majority of the models have focused on the corrosion and structural failure of buried cast iron pipes (Karaa and Marks 1990; Marlow et al. 2010; Pelletier et al. 2003). The models require data of the surrounding situation and the internal and external loads that affect the pipe (Marlow et al. 2010)

2.4.2 Descriptive analysis

Descriptive analysis organizes and summarizes system inventory and condition, leak/break repair data to determine the physical characteristics and repair patterns of the system (Karaa and Marks 1990). It focuses on overall analysis of the system rather than the pipe individuals and is suitable for cities that have comprehensive databases on characteristics of their pipes and the pipe breaks (Hu and Hubble 2007). These analyses tend to obscure the high variability of failure patterns that exist among different pipes in the system (Andreou 1986). Such studies have been conducted by researchers (O'Day 1982)(Andreou 1986), (Karaa and Marks 1990), (Pelletier et al. 2003); among others), where several systems were analyzed and break patterns identified. Weaknesses attributed to this type of analysis include(Andreou 1986);

- a) Failure to capture complex interaction between break causing factors
- b) They don't provide information on the failure of individual pipes

- c) Difficulty in revealing non-linearity between break and factors contributing to break
- d) Often provides overwhelming information which is not clear on how it can be used for repair/rehabilitation/replacement decision at the individual pipe level.

It is however recommended that every technical state of a system should begin from this analysis (Røstum 2000).

2.4.3 Predictive modeling

The need to develop a quantitative decision tool for repair and replacement decisions for deteriorating pipes led to the development of predictive models for pipe break failures (Andreou 1986). Predictive modeling is done by modeling the past pipe breakage behavior in order to predict the future (Pelletier et al. 2003). Three categories have been investigated (Karaa and Marks 1990) and are: i) aggregate models; ii) regression models; and, iii) probabilistic models.

i. Aggregate models

Linear and exponential models described by Shamir and Howard (1979), are examples of the examples of aggregate models. They predict pipe failure using the two or three parameters based on pipe age, and breakage history (Kleiner and Rajani 2001). The pipes are grouped into uniform and homogeneous groups according to their attributes (WERF, 2009). The model requires relatively small amount of data for analysis. The major disadvantage of these models is failure to deal with the right censored data (Martins 2011). They also, do not provide information about the goodness of fit and the statistical significances of the coefficients used, and the individual pipe level in the system, as they do not account for various pipe characteristics, previous break history and environmental factors (Karaa and Marks 1990).

ii. Regression models

A model of this type was proposed by Clark *et al.* (1982) and consists of two equations; a first-event equation to predict time elapsed between original time of installation and the time of first break; and, accumulated event equation for predicting subsequent failures (Rogers and Grigg 2009) using a set of explanatory variables, such as pipe age and pipe diameter (Martins et al. 2013). Therefore, the model can provide greater insights into break-causing factors in water

pipes, and is better suited for repair-versus replacement decisions than the aggregate-type models (Karaa and Marks 1990). Some of the weakness of the model are: i) The model does not capture the effects of aging on the pipe satisfactorily; ii) it requires model identification; and, iii) it does not use the information of pipes that have not broken effectively in predicting the pipe breaks (Andreou 1986).

iii. Probabilistic modeling

Two types of predictive models have investigated and are; i) single-variate; and, ii) multivariate probabilistic models (Martins 2011). They focus on time-to-failure and estimated the probability that a break will occur at some future time (Andreou 1986).

Single-variate probabilistic models, for example Cohort survival model described by Herz (1996), are used in; i) predicting failure patterns of pipe groupings or cohorts (Fuchs-Hanusch et al. 2011), (Martins et al. 2013); ii) deriving probabilities in life expectancy, and; iii) probabilistic analysis of break clustering (Kleiner and Rajani 2001). Other models that apply the attributes of the pipes as a grouping criteria, rather than covariates include Bayesian diagnostic model proposed by Kulkarni et al. (1986) and Poisson distribution as applied by (Martins 2011). The major issue of single-variate models is the assumption that the failure rate is the same for each pipe cohort. This can lead to biased predictions when large number of groups exists leading to shortness of data (Martins 2011).

Multi-variate probabilistic models are used in predicting individual pipes rather than the cohorts (Kleiner and Rajani 2001). The most common and widely used approach is survival analysis and has been applied by researchers in predicting pipe failure for more than two decades (Pelletier et al. 2003). Description of survival analysis is given in the following subsection.

A few of the break estimation models are discussed here below;

a) Time replacement models

Shamir and Howard (1978) developed a methodology to determine when to replace the pipes in a homogeneous group. Based on the pipe break data, the number of pipe breaks expected is given by Equation 2-1

$$C(t) = C(t_o)e^{A(t-t_o)} \quad 2-1$$

where, $C(t)$ = number of breaks/km in year t , A = growth rate per year and ranges from 0.01-0.15, t_o = base year for the analysis. The economic data were combined with break forecast to determine pipe replacement time. The model is easy to use and requires less data in analysis. However, it does not consider right censored data and physical, operational and environmental attributes of the data (Karaa and Marks 1990; Martins 2011)

Walski and Pelliccia (1982) proposed an almost similar model to that developed by Shamir and Howard (1978) for pipe break prediction of homogeneous pipe groups. The model considered pipe previous break failures and diameter and is defined by Equation 2-2

$$C(t) = c_1 c_2 a e^{b(t-k)} \quad 2-2$$

where, a = regression coefficient, b = regression coefficient/year, $C(t)$ = number of breaks/km in year t , c_1 = correction factor for previous breaks, c_2 = correction factor for pipe size and k = pipe installation year.

b) Time linear model

Kettler and Goulter (1985) proposed a simple linear model that establishes a relationship between pipe age and number of breaks. They found that the number of breaks increased with age. The number of breaks per year C is given by

$$C = k_o \text{ Age} \quad 2-3$$

where, k_o = regression parameter. A low coefficient of correlation of 0.562 and 0.1 for asbestos cement and cast iron pipes respectively was found which increased to 0.88 and 0.67 asbestos cement and cast iron pipes respectively when an outlier was omitted. The models is easy to use and requires less data but does not consider physical, operational and environmental attributes of the data (Kleiner and Rajani 2001)

c) Break clustering

Goulter et al. (1993) proposed model defined by Equation 2-4 to predict the number of breaks as a function of time and space.

$$P(x) = \frac{m^x e^{-m}}{x!}, \quad m = \int_0^s \int_0^t r(s, t) dt, ds \quad 2-4$$

where, m = mean number of subsequent failures occurring in the cluster domain, x = number of subsequent failures, and $r(s, t)$ = failure rate as a function of time t and distance s . Cross-referencing is done to determine mean number of breaks that occur after the first break. The model is limited to establishing the recurrence of breaks and cannot be used in predicting the occurrence of the first break (Kleiner and Rajani 2001). Data required in this model is the break data with precise location and time of the break.

d) Threshold Break Rate

Loganathan et al. (2002) developed a model to determine when to replace or repair a pipe. The model implies that repairs have been done $(n-1)$ times but ultimately replacement will be done at the n^{th} break. Pipe replacement is done when the cost of repairing the current break equals or exceeds the threshold break rate given by:

$$Brk_{th} = \frac{1}{t_n - t_i} = \frac{\ln(1-R)}{\ln\left(\frac{C_{n+1} + A_{n+1}}{A_n}\right)} \quad 2-5$$

where, Brk_{th} = threshold break rate, C_i = the repair cost of the i^{th} break, A_n = replacement cost at time t_n , R = discount rate, t_i = time of i^{th} break from installation date. The data used in the model are the failure history of the pipe, cost of repair and replacement, the time of installation of the pipe, the discounting rate.

2.5 Survival Analysis

2.5.1 Overview

Survival analysis incorporates the fact that while some pipes break, others do not and this information has a strong impact on impact of analysis (Mailhot et al. 2000). The models use covariates to differentiate the pipe failure distributions without splitting the failure data, thereby giving a better understanding of how covariates influence the failure of the pipe (Martins 2011; Martins et al. 2013).

The ability to quantitatively and explicitly consider most of the covariates in the analyses makes them powerful and general in predicting the future breakage rates of water mains (Kleiner and Rajani 2001). Survival and hazard functions are the main fundamental elements in survival analysis (Røstum 2000). A brief description of them is given here below:

2.5.1.1 Survival Function

Survival function, S refers to the probability of a pipe surviving beyond time t (Røstum 2000) and often describes time to the occurrence of an event which can either be break or censorship (Cox 1972).

$$S(t) = P(T > t) \quad 2-6$$

where, T is the random variable representing time-to-failure

S is non-decreasing function and has a value of one at the origin and zero at infinity (Røstum 2000). When T is a continuous random variable, the survival function is a complement of the cumulative density function F given as

$$F(t) = 1 - S(t) = P(T < t) \quad 2-7$$

2.5.1.2 Hazard Function

This is the instantaneous failure rate of a pipe and is defined by:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad 2-8$$

where, $h(t)$ is the hazard function and Δt is a small change in time. When T is a continuous random variable representing time to the next failure, then

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \ln[S(t)] \quad 2-9$$

where, $f(t)$ is the probability density function (pdf). Baseline cumulative hazard can then be calculated from Equation 2-10 as

$$H(t) = -\int_0^t h(t) \quad 2-10$$

Therefore, baseline survival function can be obtained from cumulative hazard as:

$$S_0(t) = \exp[H(t)] = \exp\left\{-\int_0^t h(t)\right\} \quad 2-11$$

Therefore, the failure time distribution of the pipes in water distribution network may be investigated through survival function $S(t)$, or the hazard function $h(t)$.

2.5.2 Observation Time and Censoring in Survival Analysis

Observation time of a pipe is the time from the beginning of study until when failure occurs or when the study ends- whichever comes first (Smith et al. 2003). Normally, for pipes installed before the study began, observation time starts of at the time of study, while for pipes installed later, their observation time begin during the study period. Figure 2-4 shows a simple design where observation time for a pipe starts at a consistent point in time (t_0).

If the subject under study fails (event has occurred) within the observation time (t_i), then the subject is known as uncensored subject (indicated by solid line with round bullet in the figure below). However, if the event does not occur until the end of the study, then the subject is known as censored subject as the information is lost on the subject. This type of censoring is called right censored.

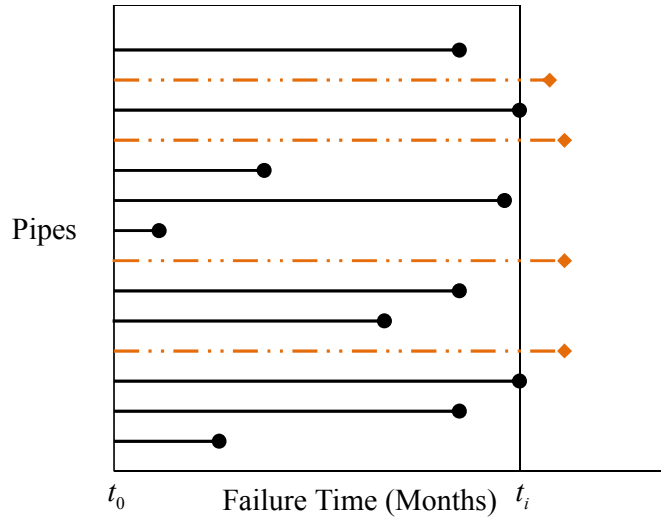


Figure 2-4: Observation Time for a Pipe

Left censoring occurs when the pipe has already failed by the time the study begins and no data were recorded (Røstum 2000). Of great interest in survival analysis is the right censored data (Smith et al. 2003).

2.6 Application of Statistical Models to Water Networks

Many statistical models have been applied in many water networks in the world. A brief review of some of the applications of Cox-PHM, WPHM and PM is presented here below.

Andreou et al. (1987b) applied the PHM to describe the slow and fast breaking phases of pipe break. Data were stratified according to the number of pipe breaks with covariates such as pipe length, diameter, installation period, material type, soil corrosivity and land development being included model development. The results indicated that after the third break pipes had a constant break rate. The coefficients of determination (R^2) for the predictive models were found to be 0.34 and 0.46 for pipes with more than three breaks and more than six breaks, respectively. The authors concluded that PHM can only best described the slow breaking phase while the fast breaking phase is modelled as Poisson arrivals.

Vanrenterghem-Raven et al. (2004) applied Cox-PHM and WPHM models to determine pipe ranking based prioritization and optimization techniques using 20yr data from the New York City. Covariates included were number of pipe breaks, diameter, length, and installation dates, pressure, soil type, material with stratification done basing on number of previous breaks. An index was developed to prioritize pipes and optimization was done by estimating the number of expected breaks using WPHM.

Park et al. (2008) applied Cox-PHM in modeling hazard rates for consecutive pipe (cast iron) failures for a water distribution system with 94yrs of pipe data. The covariates analyzed included pipe material, installation time, diameter, length, land development, and pipe internal pressure. Pipes were categorized into seven ordered survival time groups according to minimum number of breaks recorded. The results indicated that the hazard rate for the first break increased with time while for the second and more breaks, the hazards decreased with time. Further, the time between a pair of consecutive breaks generally decreased as the number of breaks increased. Additionally, longer pipe lengths and small diameter pipes were found to increase the hazard of the pipes breaking but the hazards decreased as the number of breaks increased.

Christodoulou (2010) presented a framework to determine the time-to-failure of a pipe using a combination of ANN, parametric and non-parametric survival analysis. The 5yr data from City of Limassol, Cyprus, were analyzed using Poisson and Cox-PHM regression. ANN model was used for sensitivity analysis and sensitive covariates were found to be number of previous breaks, material, diameter, traffic loading the vicinity of the pipe, pipe type (house connection or water main), incident type (pipe deterioration, corrosion, interference by others, tree roots, connection hose, other). Data stratification was done using diameter and number of previous breaks. Analysis from Cox-PHM revealed high hazard ratios for MDPE pipe material, pipe deterioration, pipes laid closer to trees, and smaller diameter pipes. Further, it was found that the hazard ratio increases by 20.49% with each increase in break for pipes with 4-8 breaks.

Fuchs-Hanusch et al. (2012) applied Cox-PHM in pipe failure prediction and rehabilitation to water networks in order to describe how the system reacted on rehabilitation of specific pipes in three Austrian utilities. Covariates used in the study were diameter, length, number of previous failures, soil type, material, and vintage. The break data were stratified based on the number of previous breaks and the results showed that the steepness of the survival curve

increased with the amount of previous failures. Also, validation of the model was done using data from one of the cities under study and the model predicted good results.

Le Gat and Eisenbeis (2000) used WPHM to calculate and forecast the water main breaks for water utilities (Charente-Maritime and Lausanne) with short and long-term records. Covariates considered for Charente-Maritime were pipe length, age, diameter, types of pipe assembling, soil type, level of traffic and the kind of supply, while for Lausanne, the factors considered included; length, diameter, age and pressure. The data were stratified according to the material type and the number of observed breaks. The model reproduced observed data from Charente-Maritime but underestimated the prediction for Lausanne. The authors therefore recommended the use of the model for use in water networks with short data records.

Mailhot et al. (2000) applied a two parameter Weibull (W) distribution and one parameter exponential (E) distribution to three municipalities in Canada with 21yrs of data records. The covariates considered for the models were pipe material, diameter, installation year, pipe length, soil type and land use above the pipe. Data were stratified according to the year of pipe installation. Four different combinations of W and E distributions were modelled with increasing number of parameters. The models were trained using the 21yr data and validated with synthetic data generated from the developed models. The authors reported a strong relationship between number of breaks, installation year and material type. Further, higher risk was associated with pipes installed in 1960s than all other periods. The authors concluded that after the first break, the occurrence of subsequent breaks did not vary with time therefore they concluded that Weibull distribution can be used to described time-to-first break while exponential distribution modeled the subsequent breaks. Additionally, they concluded that the models were appropriate for modeling water networks with less data.

Toumbou et al. (2012) applied Weibull-exponential model to describe the time of breaks for a small city in Quebec. The Weibull model was used to describe time to first break, while the exponential model was used to model the subsequent breaks. Data used included number of pipe breaks, diameter, length, and installation dates. Comparison between the model with and without covariates was done and though the results were similar, the model with covariates predicted slightly low values. The results from the model indicated that the proposed model with the covariates provided a good evolution of breaks.

Duchesne et al. (2012) applied both Weibull and exponential distribution in a study of different deterioration states of a sewer system in Quebec City using camera inspection data. Data which included inspection data containing, diameter, age of the pipes, installation year, sewer type, material type were used to categorize the pipes into different states. A calibration procedure that captured both the left and right censored data was developed and applied. The models were trained using the data from pipes 50yrs or less at inspection time while data from pipes installed from the year 1900 were used for validation. Both models yielded almost similar results but the ease of working with exponential distribution led the authors recommend exponential distribution to be used in modeling global sewer system. Some of the model limitations include uncertainties in image interpretation, inability of the model to capture recently replaced pipes and the model cannot consider explanatory variables.

Martins (2011) applied three failure prediction models which are; Weibull accelerated lifetime model (WAFT), single-variate Poisson process, and linear extended Yule process to a Portuguese urban water supply network. Data used for all the models were pipe diameter, material type and length number of previous failures, pipe material, and pipe age. WAFT model produced the best prediction results as compared to the other models and had ability to detect pipes with high likelihood of failure. Poisson model performed poorly in the prediction of pipe breaks due to the assumption of constant failure rate for all pipes in a category. Additionally, the results indicated no significant difference between the models when predicting pipes with no failure history.

Rajani et al. (2012) applied Non-Homogeneous Poisson model to three sets of data from US and Canadian municipalities to study the effects of temperature based covariates on pipe break of different materials (DI, CI and galvanized steel). Data used in the study included water temperature and air temperature. The model was applied only to homogeneous pipe groups and individual water mains within the group were not addressed. Coefficient of determination (R^2) and likelihood ratio test (LR) were used in determination of best covariates. The authors found that water-based covariates had significant impact on pipe break than the air-based covariates and the impacts of these covariates on pipe differed according to type of pipe material.

Kleiner and Rajani (2010) described a Non-Homogeneous Poisson for analysis and forecasting of break patterns in individual water mains that includes dynamic and static factors. This model was named Individual water main renewal planner (I-WARP). The model was applied to data

from a water utility in western Canada with 40 years of pipe break data. The performance of the model was tested using the pipe coefficient of determination (pR^2) and time coefficient of determination (tR^2). The model predicted well the cumulative yearly breaks ($tR^2=0.61$) but the prediction per pipe varied depending on the length of the data ($pR^2=0.43$) with overestimation and underestimation being observed with short and long data sets respectively.

Asnaashari et al. (2009) used two regression models- multiple regression model and Poisson model to model pipe break for the City of Sanandaj, Iran with 10 years of recorded data. The data were categorized depending on four material types, asbestos cement (AC), cast iron (CI), ductile iron (DI) and polyethylene (PE). Covariates used in model development included pipe diameter, length, age, cover depth, pressure, pipe wall thickness, pipe location and failure history. Equations for these pipe materials were generated. Though satisfactory results were obtained from the regression model with R^2 range of 0.52-0.88, analysis of the model residuals indicated a poor fit. Poisson models therefore provided a better fit of the data (R^2 range of 0.71-0.95) with an overall R^2 of 0.9 fit for the data.

Boxall et al. (2007) used theory of generalized linear models and applied Poisson model in studying the burst rates of two material types; cast iron and asbestos cement pipes, using two data sets from water utilities in UK. Data used in developing the model are pipe age, length, diameter, and soil characteristics. The derived relationships between the burst rate and the covariates were explained using the parametric models and validation performed using the non-parametric forms of developed models. The results indicated an exponential relationship between burst rate and diameter, reduction in failure rates as length increases and a complex relationship between age and failure rates. Further, no relationship was found between soil and burst rates.

The summary of the above survival models are given in Table 2-2 while the data requirements are as shown in Table 2-3.

Table 2-2: Summary of Survival Models

Reference	Model	Pipe applied	Data requirements
Andrauet. al. (1986a and b)	Early stage $h(t) = h_o(t)e^{zb}$ Late Failure stage $P(x) = \frac{(\lambda t)e^{-\lambda t}}{x!}$	Metallic pipes	<ul style="list-style-type: none"> • Pipe length, diameter, installation period, material type, soil corrosivity and land development type
Le Gat and Eisenbeis (2000)	$\ln(T) = X\beta + \sigma W$	AC, steel, ductile iron, cast iron, PVC,	<ul style="list-style-type: none"> • Pipelength, age of the pipe, pipe diameter, type of pipeassembling, soil type, level of traffic and the supplymethods (e.g. gravity and pumping)
Pelletier <i>et. al.</i> (2003)	First failure $\lambda(t) = kp(kt)^{p-1}$ Subsequent failures $\lambda(t) = k_2$	Cast iron, ductile iron and PVC	<ul style="list-style-type: none"> • Pipediameter, length, type of material, year of installation, type of soil and land use above the soil
Davies <i>et al.</i> (2008)	Weibull extreme value distribution $F(S_R) = 1 - \exp(-S_R/\alpha)^n$ Herz distribution	Asbestos cement pipes in Australia	<ul style="list-style-type: none"> • pipe diameter, wall thickness, installationyear, internal pressure, burial depth, unit weight of thesurrounding soil, live surface load from traffic anddegradation rate
Park <i>et al.</i> (2008)	Proportional hazard model $h(t) = h_o(t)e^{zb}$	Cast iron	<ul style="list-style-type: none"> • pipe material, installation time, diameter, length, land development, and pipe internal pressure
Rogers and Grigg (2009) and Park <i>et al.</i> (2007)	Non-Homogenous Poisson process (NHPP) $M(t) = \lambda t^\beta$	Cast iron, ductile iron, PVC	<ul style="list-style-type: none"> • pipe material, diameter, pipe break type, soil type, pipe installation date
Martins (2011)	$P(x) = \frac{(\lambda t)e^{-\lambda t}}{x!}$ $\ln(T) = X\beta + \sigma W$ Linear extended YULE (YELP)	AC, PVC, ductile iron, and HDPE	<ul style="list-style-type: none"> • For Poisson: pipe diameter, material type and length • For Weibull: pipe length, diameter and number of previous failures • For YELP: Pipe material, length, diameter, pipe age and number of previous failures

Table 2-3: Summary of statistical model applications to water networks

Model	Authors	Covariates considered			Prediction type
		Physical	Environmenta l	Operation al	
Cox-PHM	Fuchs-Hanusch et al. (2012)	D, L, M, A	S	PF	Break prediction
Weibull Exponential model	Toumbou et al. (2012)	D, L			Break prediction
Weibull Exponential model	Duchesne et al. (2012)	D, M, IY			State of sewer deterioration
NHPP	Rajani et al. (2012)		wTm, aTm		Break prediction
SVPP	Martins (2011)	D, L, M, IY		PF	
Cox-PHM, Poisson	Christodoulou (2010)	M, D	TL	PF, IT	Break prediction
NHPP (I- WARP)	Kleiner and Rajani (2010)	D, L, M	Tm, Pr		Break prediction
Poisson and Multiple regression	Asnaashari et al. (2009)	D, L, A, M, H, W	PL	P, PF	Break prediction
Cox-PHM Poison	Park et al. (2008) Boxall et al. (2007)	M, IY, D, L D, L, M, IY	LD S	P	
Cox-PHM and WPHM	Vanrenterghem-Raven et al. (2004)	D, L, IY	S	P, PF	Pipe replacement
Weibull and Exponential	Mailhot et al. (2000)	M, IY, L	S, LD		Weibull-1 st break prediction, Exponential-subsequent breaks
WPHM	Le Gat and Eisenbeis (2000)	L, A, D, As, SL	S, TL	P	Break prediction
Cox-PHM	Karaa and Marks (1990)	L, D, IY, M	SC, LD		Probability of failure
Cox-PHM	Andreou et al. (1987b)	L, D, IY, M	SC, LD		Probability of failure

D=diameter, L=length, M=material, S=soil type, SC=soil corrosivity, LD=land development, IY=installation year, A=age, P=water pressure, SL=supply level, TL=level of traffic, PF=previous failure, Tm=temperature, Pr=precipitation, wTm=water temperature, aTm=air

temperature, IT=incident type, PL=pipe location, W=wall thickness, H=depth, As=assembling type, NHPP= non-homogeneous Poisson process and SVPP= single-variate Poisson process.

Chapter 3

3 Statistical Models

This section presents an overview of survival analysis, and then a review of Cox-Proportional Hazard Model (Cox-PHM), Weibull Proportional Hazard Model (WPHM) and PM.

3.1.1 Cox-Proportional Hazard Model (Cox-PHM)

Cox-Proportional Hazard Model (Cox-PHM) was developed by (Cox 1972) in order to examine the effects of different covariates on the time-to-failure of a system and is of the form:

$$h(t|z) = h_0(t) \exp(z\beta) \quad 3-1$$

where $h(t|z)$ is the hazard function, t is the elapsed time from the last failure, $z = [z_1, z_2, \dots, z_q]$ is the covariates vector, $\beta = [\beta_1, \beta_2, \dots, \beta_q]$ is the vector of coefficients, and $h_0(t)$ is the baseline hazard function. The $h_0(t)$ represents aging process that goes on without the influence of covariates (Andreou et al. 1987a).

The basic assumptions made on this model are (Andreou 1986; Karaa and Marks 1990):

- i) Explanatory variables have multiplicative effect on the break rate; and,
- ii) Log-linear relation between the break rate and the break causing factors.

The distinctive characteristics of the model include (Andreou 1986; Karaa 1984; Karaa and Marks 1990):

- (i) It gives probability of failure for each individual pipe
- (ii) Utilizes effectively the information obtained from censored data
- (iii) Pipes can be grouped and modelled according to the cause of breaks factors, and as a result failure trends amongst pipe groups are easily identifiable.

The assumption of multiplicative effect of the covariates on the baseline hazard rate imply that the ratio of two hazard rates observed at any time t with set of covariates z_1 and z_2 should be constant and proportional (Equation 3-2) to each other (Kumar and Klefsjö 1994).

That is

$$\frac{h(t | z_1)}{h(t | z_2)} = \frac{h_0(t) \exp(z_1 \beta)}{h_0(t) \exp(z_2 \beta)} = \exp \beta (z_1 - z_2) \quad 3-2$$

The survival function S , is given by:

$$S(t | z) = S_o(t)^{\exp(\beta z)} \quad 3-3$$

where, the baseline survivor function $S_o(t)$ is given as given in Equation 3-3 ,

3.1.1.1 Estimation of Cox PHM regression parameters

Coefficients $\beta = [\beta_0, \beta_1, \dots, \beta_q]$, are obtained by maximizing partial likelihood estimate, L_p function.

Let time $t_i, i = 1, 2, \dots, n$ represents time-to-failure for the i^{th} subject for n pipes and let $\mathbb{R}(t_i)$ be the risk set of all pipes, k at time t . If δ_i represents an event indicator at each failure point at time t_i , where $\delta_i = 1$ when event has occurred, otherwise $\delta_i = 0$, and subject i^{th} has a vector of covariates $z = [z_0, z_1, \dots, z_q]$, then the partial likelihood estimate L_p of $\beta = [\beta_0, \beta_1, \dots, \beta_q]$ is (Kumar and Klefsjö 1994):

$$L_p(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta z_i)}{\sum_{k \in \mathbb{R}(t_i)} \exp(\beta z_k)} \right]^{\delta_i} \quad 3-4$$

This procedure is undertaken using *coxph* function in R software (Team 2013).

3.1.1.2 Baseline hazard estimation

The baseline hazard function captures the effects of time on the survival process, and it is to be estimated in order to use the Cox-PHM for failure prediction and evaluating the effects of aging on the pipe (Røstum 2000). It represents the hazard rate that a pipe would experience if the effects of covariates are equal to zero (Kumar and Klefsjö 1994).

In this thesis, the baseline hazard function is approximated by curve smoothing techniques such as linear and non-linear least squares (*nls*) relations (Rossiter 2009). Linear relations technique is used to fit relations that are thought to be linear using ordinary linear squares (OLS) to minimize the sum of squares of the residuals while the *nls* technique is used to fit equations that cannot be linearized (Rossiter 2009). The baseline hazard estimation involves plotting the baseline hazard curve using Equation 3-5 and using the linear and *nls* techniques, a fit of the curve is approximated and any competing equations can be determined using analysis of variance (ANOVA).

$$H(t) = -\ln[S_o(t)] \quad 3-5$$

The assumptions made on this proposed methodology are; i) fitted equation will result in less fitting noise (Rossiter 2009); and, ii) the representative equation will apply to the future failure rates (Kleiner and Rajani 2001).

3.1.1.3 Proportional Hazard Assumption (PHA) checking

The main assumption of the PHM is the proportionality of the hazards of any two individuals i.e. the hazard ratios is constant over time (Qi 2009). The methods used to verify that the model satisfies PHA assumption are discussed here below:

3.1.1.3.1 Graphical method

The PHA holds if two plotted curves are parallel and implies the hazards are constant and proportional (Kumar et al. 1992). This test is performed by plotting the log-minus log survival function of the function (Andreou 1986) given by Equation 3-6

$$\ln[-\ln S(t|z)] = \sum_{i=1}^q z\beta + \ln[-\ln S_o(t)] \quad 3-6$$

If the curves are not parallel, then the PHA is deemed to have been violated. In his thesis, (Qi 2009) reported that graphical method is not suitable for determining proportionality of the PHM if the model has a categorical covariate with several levels.

3.1.1.3.2 Schoenfeld Residuals

After fitting the model, it is vital to check the residuals of the fitted model. For each covariate in the model, Schoenfeld residuals U , are defined for every subject with an event (Abeysekera and Sooriyarachchi 2009). This is given by

$$U_i(t) = Z_i(t_i) - \bar{Z}_i(t_i) \quad 3-7$$

where, Z_i is the observed covariate and \bar{Z}_i is the weighted average of the covariate over the risk time t_i .

It is expected that, if the model satisfy the PHA, the effects of covariates should be constant over time (Abeysekera and Sooriyarachchi 2009). Therefore, a plot of the residuals should be clustered around zero, otherwise the PHA is deemed to have been violated (Kumar and Klefsjö 1994).

3.1.1.4 Stratification of the PHM

Stratification is done only on the covariates that do not comply with PHA, and it involves fitting a model that has a different baseline function in each stratum. The model is given by

$$h_g(t|z) = h_{og}(t) \exp(z\beta) \quad 3-8$$

where, g represents stratum and β and z are as defined in the preceding sections.

The drawback of this method is that the effect of the covariate used in stratification on the failure of the pipe cannot be determined (Kumar and Klefsjö 1994).

3.1.1.5 Failure prediction

Failure prediction of the Cox-PHM involves estimation of the baseline hazard function using the methodology presented in Section 3.1.1.2. The resulting equation is fitted to Equation 3-3 and failure probabilities of each survival time (Jenkins 2005; Zhang 2005) are generated for each individual pipe in the data set. The sum of all the probabilities of each survival time gives the sum of the breaks for that survival time (Toumbou et al. 2012).

3.1.2 Weibull Proportional Hazard Model (WPHM)

This is a parametric versions of Cox-PHM but the difference is that the baseline function is assumed to follow a specific distribution when the model is fitted with data (Qi 2009). It models the distribution of times to failure and to link the explanatory variables to these inter-arrival times (Le Gat and Eisenbeis 2000) rather than modelling the number of failures. In Weibull Proportional Hazard Model WPHM, a set of vector of covariates $z = [z_0, z_1, \dots, z_q]$ and error term W , is assumed to be linearly related to logarithm of time, T via a linear model (Le Gat and Eisenbeis 2000) given by Equation 3-9

$$\log T = \sigma W - \beta z \quad 3-9$$

where, $\beta = \frac{\gamma}{\sigma}$ with $\gamma = [\gamma_0, \gamma_1, \dots, \gamma_q]$ and σ being is intercept and shape parameters respectively. The hazard rate increases when $\sigma > 1$ and decreases when $\sigma < 1$. When $\sigma = 1$, the hazard rate remains constant and therefore, the model becomes exponential proportional hazard model which is a special case of WPHM.

If Equation 3-9 follows extreme value distribution (Le Gat and Eisenbeis 2000), then survival function S is given by Equation 3-10

$$S(t | \gamma, \sigma) = \exp \left\{ -t^{\frac{1}{\sigma}} \exp(-z\beta) \right\} \quad 3-10$$

The Weibull proportional hazard model follows the basic assumptions of the PHA and thus the ratio of two covariates should be proportional if the model is valid (Kumar and Klefsjö 1994). Therefore, the same principles as discussed for Cox-PHM are applicable.

3.1.2.1 Estimation of regression parameters in Weibull PHM

The regression parameters σ , and β are estimated using maximum likelihood estimator (MLE).

Consider a Weibull distribution with density $f(t)$, survivor $S(t)$ and hazard functions $h(t)$ and parameters σ and β , the likelihood contribution of the pipe failing at time t_i is given by its density written as the product of the hazard and the survivor function. If δ_i is the death indicator where $\delta_i = 1$ if pipe fails otherwise, $\delta_i = 0$ and if the failures are independent then the contribution of m failures (Crowther and Lambert 2013) is

$$\log f(t_i | \beta, \sigma) = \prod_{i=1}^m \delta_i \log[h_0(t_i | \sigma)] + \delta_i \beta z + \exp(\beta z) \log(S_0(t_i | \sigma)) \quad 3-11$$

where, $h_0(t_i)$ and $S_0(t_i)$ is baseline hazard and baseline survival function for the WPHM, respectively.

In the R software, the MLE values are obtained by maximizing the logarithm Equation 3-11 using *survreg* function (Team 2013).

3.1.2.2 Failure prediction

By representing Equation 3-10 as a function of time, random failure times were generated for all the pipes using Equation 3-12, and then the failure times within the prediction time are then calculated for each pipe and recorded by applying methodology developed by Le Gat and Eisenbeis (2000).

$$t = \left\{ \ln \left[\frac{1}{S} \right] \exp \left(\frac{-\gamma z}{\sigma} \right) \right\}^{\sigma}$$

This process is done 1000 time and therefore the number of failures is the mean of the simulated failures (Le Gat and Eisenbeis 2000)

3.1.3 Poisson process

Poisson Model (PM) is used to count the number of failures $N(t)$ of repairable systems within a time interval $(0, t]$ (Røstum 2000). Poisson process is said to be a counting process $\{N(t), t \geq 0\}$ if $N(t)$ satisfies the following conditions (Røstum 2000)

1. $N(t) > 0$
2. $N(t)$ is an integer
3. If $t_1 < t_2$ then $N(t_1) \leq N(t_2)$
4. For $t_1 < t_2$, $[N(t_2) - N(t_1)]$ represent the number of failures that have occurred in the time interval $[t_1, t_2]$

It follows that if the number of random events $N(t)$ is Poisson distributed (Greene 2003) then

$$P\{N(t) = n\} = \frac{e^{-\beta t} (\beta t)^n}{n!} \quad 3-13$$

where $P\{N(t) = n\}$ is probability of occurrence of n failures for $n = 1, 2, \dots$, and β is the coefficients of covariates which for Poisson represents the rate of Poisson process.

3.1.3.1 Parameter Estimation

The β is estimated using the Maximum Likelihood Estimator (MLE) (Greene 2003). Consider pipe i with number of failures n_i for $i = 1, 2, \dots, m$, occurring during the observation time t_i . The MLE, L is then given by

$$L(\beta | t, n) = \prod_{i=1}^m \frac{e^{-\beta t_i} (\beta t_i)^{n_i}}{n_i!}$$

where, $t_i = 1, 2, \dots, m$.

This procedure is undertaken in R software (Team 2013).

3.1.3.2 Prediction of Failures

The estimated failure rate is assumed to be the same for all the pipes in a stratum. Using estimate $\hat{\beta}$ of β , the number of failures can be obtained using Equation 3-15 by calculating the expected value $E[N_i(t)]$ for a given pipe in a given time period (Martins et al. 2013).

$$E[N_i(t)] = \hat{\beta}t$$

3.2 Model Development

The procedure followed in model development is depicted in Figure 3-1. The details of each step are provided in the subsections below:

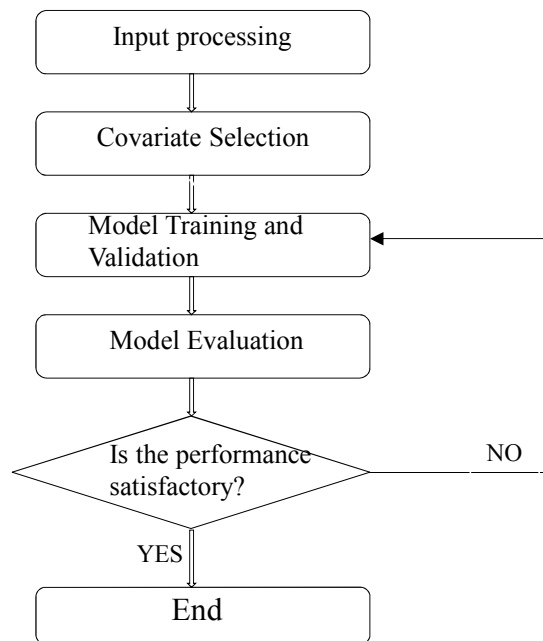


Figure 3-1: Schematization of model development

3.2.1 Input Processing

3.2.1.1 Pipe Data

Data was divided into three main categories according to material type giving rise to Cast Iron (CI) (both thin-wall and thick-wall), Ductile Iron (DI) that combine both unwrapped and wrapped DI and YDI), and PVC strata. The basis of this stratification is that each material type exhibits different break pattern e.g. temperature will not act in the same way in DI and PVC. For each defined strata except for PVC, two strata were defined according to the number of previous breaks (PRB); one with no PRB and the other with more than one PRB. For PVC strata, the full data set are used. Similar stratification has been applied by Le Gat and Eisenbeis (2000). The above stratification was applied in break prediction

Further, for estimation of Mean Time to Failure (MTF) of pipe materials, pipes were stratified according to PRB where a total of seven ordered strata were obtained. The seventh group contains all pipes with seven and more breaks

3.2.1.2 Weather data

Weather data such as temperature and precipitation used in this study was acquired the Environment Canada (Environment Canada 2013). Data from Calgary International Airport weather station was used as this station was within the study area and had the longest record panning to 1894. The summary of the weather data of how weather data are applied in this thesis is presented in the subsections below:

3.2.1.2.1 Freezing Index

The number of breaks has been observed to increase whenever there is an increase or a decrease in the temperature. In continental climates, most of the breaks are observed to occur in winter (Hu and Hubble 2007; Kleiner and Rajani 2000, 2002), while in temperate climates, most of the breaks occur during summer (Mordak and Wheeler 1988). Temperature was applied as Freezing Index (FI) which is the cumulative number of days below 0 degrees during a given period (Kleiner and Rajani 2002) given as

$$FI_p = \sum_{i \in DD_p} \phi - X_i \quad 3-16$$

where, FI_p is the freezing index for day i , DD_p is all the days below the threshold temperature ϕ and X is the daily temperature for day i

3.2.1.2.2 Rain deficit (RD)

The interaction between expansive soil and precipitation results in soil movement (Gould et al. 2009; Mordak and Wheeler 1988) that leads to transmission of loads (Hu and Vu 2006) to the pipes thereby causing circumferential failure type (Rajani et al. 1996). Rain deficit was selected to quantify the depletion of soil moisture in the ground. Evapotranspiration was used as a surrogate of water loss in the ground and was calculated using Thornthwaite (1948) method. The difference between precipitation P and evapotranspiration (Markovic et al. 2012) gives the rainfall deficit. Negative values indicate a precipitation deficit while positive values indicate surplus of precipitation. The Thornthwaite equation is given as (Thornthwaite 1948)

$$e = 1.6(10T / I)^a \quad 3-17$$

where, e is the monthly evapotranspiration in cm, T is mean monthly temperature, °C, I is a constant representing heat index and ranges from 0-160, a is a constant ranging from 0-4.25 and is a function of I . Therefore, RD is given as (Markovic et al. 2012)

$$RD = P - e \quad 3-18$$

3.2.2 Model Variable Selection

Although there are several methods for variable selection, the common is the use of stepwise selection method (Gupta 2013). This method involves addition of the best covariate step-at-a-time into the model from those that are currently excluded from the model (Demyanov et al. 2012). Example of the method is Akaike's Information Criterion (AIC) developed by Akaike (1973) and is defined by Equation 3-19

$$AIC = -2(\log \text{likelihood}) + kp$$

where, p is the number of parameters in the model, k is 2. The log-likelihood indicates the overall fit of the model with the smaller likelihood values indicate the worst fit and lower AIC values indicate the best variable or model. The kp term penalizes variables and the overly complex models (Demeyanov et al. 2012). This methodology is applied in selecting the variables to be included in model development.

3.2.3 Model Training and Validation

Model training is the use of numerical techniques to fit the available data into the model in order to obtain the model coefficients that accurately fit the data and to question the inclusion of some variables in the fitted model is justified (Jones 1983).

Model validation is an essential part of model building and refers to the stability of the regression coefficients, usability of the regression function and the ability to generate inferences drawn from regression analysis (Gupta 2013). This step is important in checking the model performance before being used in the decision making (Gupta 2013).

To train and validate the models, data are split into two parts (training and test sets) where the model is trained using the training set and validated from the test set. Some of the methods used in data splitting include random sampling, temporal techniques, k-fold sampling technique (Betrie et al. 2013; Martins 2011; Richardson et al. 1995). In this study, random sampling was employed and because of large data sets, two-thirds of the data was used in model training while a third was used in model testing. Dividing the data into training and test sets using random sampling method makes it very unlikely that the division is engineered to favour any one model class (Richardson et al. 1995).

3.2.4 Model Evaluation

Evaluation of the model helps in assessing the overall match between observed and predicted values for each of the models studied (Betrie et al. 2013). The techniques that will be applied when evaluating the models are the Root Mean Square Error (RMSE), Mean Absolute Error

(MAE), Root Relative Squared Error (RRSE) and Relative Absolute Error (RAE). These techniques are given by the following equations.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O - P)^2}{n}} \quad 3-20$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (O - P)^2}{\sum_{i=1}^n (O - \bar{P})^2}} \quad 3-21$$

$$MAE = \frac{\sum_{i=1}^n |O - P|}{n} \quad 3-22$$

$$RAE = \frac{\sum_{i=1}^n |O - P|}{\sum_{i=1}^n |O - \bar{P}|} \quad 3-23$$

where, O is the observed value, P the predicted value \bar{P} is the mean of the predicted value and n is the number of the observed values.

Chapter 4

4 Results and Modelling Approach

This section presents the description of the study area and descriptive statistics to give insight into the impact of different covariates on the structural deterioration of the pipes for the city of Calgary Water Network

4.1 Study Area Description

A case study was conducted for the City of Calgary Water Distribution Network (CCWN) which is shown in Figure 4-1.

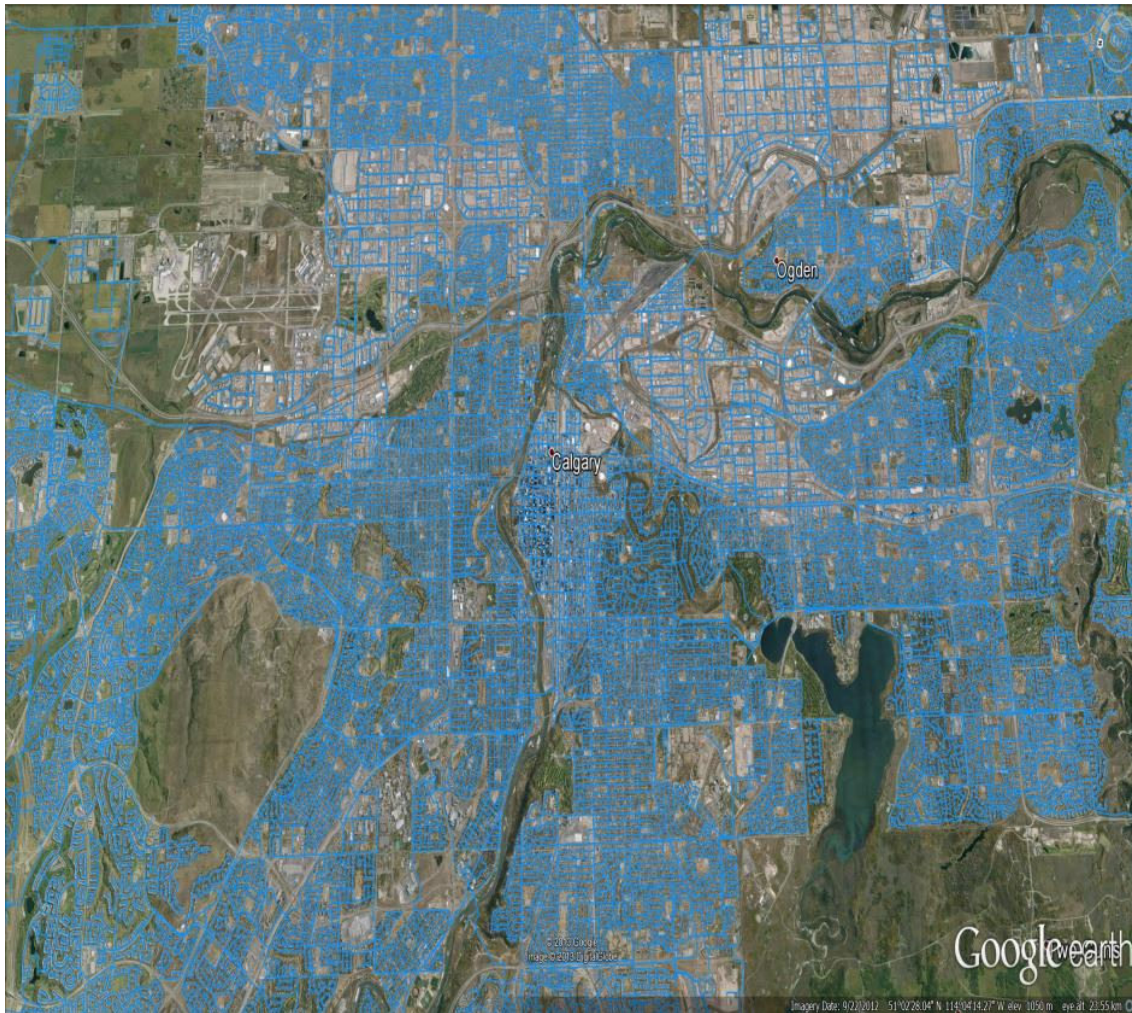


Figure 4-1: Water distribution system of the City of Calgary

The City of Calgary is located in Alberta, Canada and had a population of 1,149,552 people (in year 2013, Census 2013). It is located at approximately 1048m above the sea level and has a humid continental climate (Peel et al. 2007).

4.1.1 Weather Data

Weather data such as temperature and precipitation used in this study was acquired from Environment Canada (Environment Canada 2013) (Appendix A-1). Data from the Calgary International Airport weather station were used since it resides in the city and has over 100 years of recorded data. The following subsections describes the temperature and precipitation for the City of Calgary.

4.1.1.1 Temperature

Figure 4-2 presents the monthly temperature for the city of Calgary from the year 1956-2012. The time period for this data has been chosen to coincide with the time when recording of pipe breaks began. Additionally, monthly data records were used because the records of pipe breaks are in monthly format.

From the figure, temperature pattern varies from season to season with high and low temperatures being observed in summer and winter, respectively.

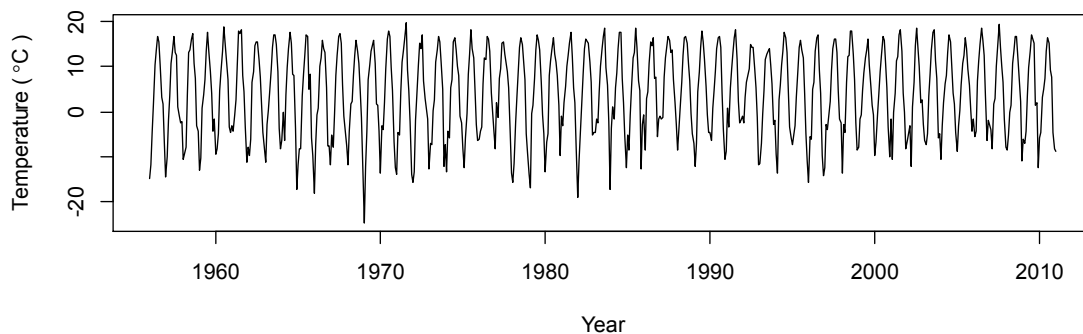


Figure 4-2: Monthly temperature pattern for the City of Calgary

The temperature trend follows a cyclic nature with the lowest temperature being observed during the winter period and the highest temperatures being in summer. The lowest temperature that has been observed in the city is -24.6 °C, whereas the highest is 19.6 °C with an overall

average of 4.0 °C. In winter period, the city has an average of -3.7 °C while in summer the average is 11.7 °C. The coldest month in the city is January with an average temperature of -9.1 °C. Further the above figure reveals that the lowest temperature ever recorded was below -20 °C and was recorded in late 1960's. According to Environment Canada (2013), average daytime high temperature ranges from 26°C in July to -3°C in mid-January.

4.1.1.2 Precipitation

Figure 4-3 presents precipitation pattern for the city of Calgary. From the figure, it can be seen that the precipitation pattern of the city is of a cyclic with most of the precipitation being observed in summer period especially in the month of June. The city receives annual precipitation of 412.6mm with 320.6mm occurring as rain (Environment Canada 2013).

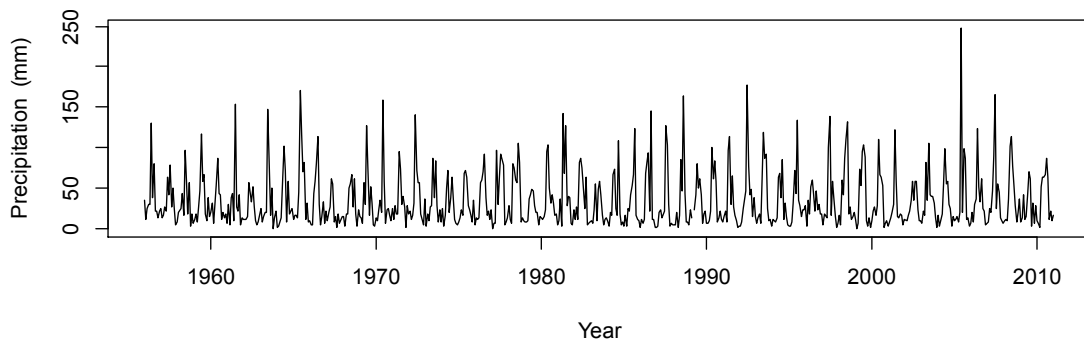


Figure 4-3: Monthly precipitation for the City of Calgary

The lowest, highest, and average precipitation that has been received in the city is 0 mm, 248 mm, and 33 mm, respectively. Moreover, it can be seen that there exist highs in the precipitation trend for the city. It seems that after a period of time, high rainfall is experienced in the city. However, the data suggest that each year the city receives high rainfall which is recorded in the month of June. The highest precipitation over the years was recorded in 2007.

4.2 Data Preparation for Modelling

This section describes the procedure applied in preparing the pipe data and the climatic information for analyses in this thesis. Methodology of data extraction and cleaning is therefore described in the following sections

4.2.1 Data Extraction

The City of Calgary Water Network has been storing their pipe data information in Geographic Information Systems (GIS) software. The utility stored the data in three main files namely; Water Pipes Current (WPC), Valve Circuit (VCIRC) and Water Main Breaks (WMB) files.

The WPC has information on the current pipes that are currently in service while VCIRC contains information all pipes in the system both replaced and those still in service. Data contained in these files include Length of pipe material, year of installation of the pipes, year of retrofitting of pipes, type of pipe material, soil resistivity, count of resistivity reading, number of residential units served, number of commercial services served, diameter of pipe and total breaks on the pipes. However, the WPC contained more information on the location of the pipe with the type of soil the pipe is laid on.

On the other hand, information contained in the WMB file included: the break-type, break-data, active and non-active pipes in the city, type of pipe material, pipe diameter, age of pipe at the time of break, cause of break, sequence of break, distance between breaks on the same pipe, number of commercial units and residential units served by the pipe.

To access these three files, GIS (Geographic Information System) software was used where they were exported to Microsoft Access Database files where they were converted to tables. The Primary key linking these files were identified, however, there was no primary link between WPC file and WMB file. Therefore, the VCIRC file and WMB file were linked together to form one single pipe data file.

Further, the monthly climatic data comprising of mean monthly temperature and monthly rainfall data were downloaded from Environment Canada (2013). The monthly time scale was chosen since the format of recording break data for a long period of time was done on a monthly basis. The key assumption to this is that the monthly record will significantly represent the weather pattern for the whole month. The limitation to this time scale is that the weather pattern within the month cannot be captured entirely by one monthly value. There could be cases where temperature could be very low or high where many breaks could have occurred as a result of this variability. By assuming the mean monthly temperature, these special cases will not be entirely captured and will then be masked in the monthly average. Thereby, this time scale might not truly represent the true information regarding the impact of the climatic parameters on the breaking trend of the pipes.

Nonetheless, temperature parameter was applied as Freezing Index following the methodology given in Section 3.2.1.2.1 of this thesis. Also, the rainfall parameter was applied as the Rainfall Deficit following the methodology described in Section 3.2.1.2.2 of this thesis. These data files were linked together with the pipe data information using the date when break occurred to the pipe and the monthly record of the climatic data. The choice of the date when break occurred as a link between the pipe data and climatic variables is to ascertain the influence of these variables in the pipe failure. It is assumed that these climatic variables have an important role in the breaks of these pipes.

The file was then exported to Microsoft Excel Spreadsheet for further processing. The procedure for extracting these files is described in the Figure 4-4.

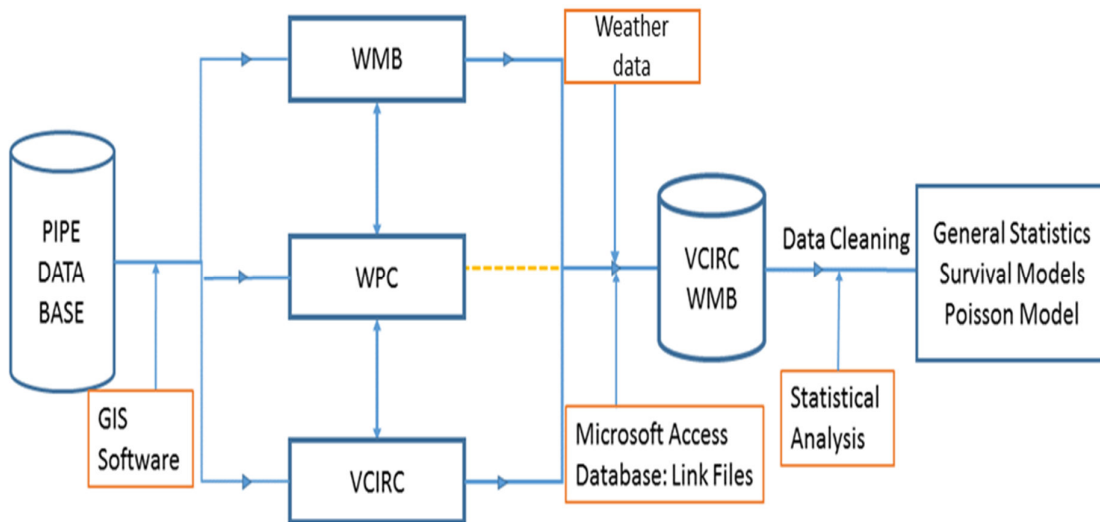


Figure 4-4: Data Extraction Flowchart

4.2.2 Data Cleaning

While cleaning the files before performing statistical analysis in R Software, some redundancies were noted. Noted from the data was the case of redundancy where multiple failures occurring on a single pipe on the same day were recorded. A sample of the affected pipes are given in the highlighted pipes shown in Figure 4-5.

It is very difficult to have the same pipe failing twice or more in one day except if the failures occur in two different locations and all these breaks are recorded. Further, since records of pipe

break data before the year 2000 involved recording monthly records, the option of multiple records happening on one pipe in a month does not hold. For the records of pipe after the year 2000, records give the date and month of occurrence, but there is no record of the exact location of occurrence; therefore, it is still difficult to note two failures in one pipe in a day.

Therefore, these redundancies were removed, which then allows the pipe to have only one record in a day, for the case of records with day, month and year of the break while those with only the month and year, only one break was kept for each month.

1	VCIRC	BREAK_T	BREAK_DATE	RF_AGE	MAT	M	DIAM	YEAR_B	AGE	SEQ	SPACE	SINCE	RSEPR
740	A11001-01				YDI	Y	200						
741	A11002-01				CI	N	200						
742	A11002-02D		3/1/1979	0	CI	N	200	1979	17	1	0	0	
743	A11002-02G		9/11/2001	0	CI	N	200	2001	39	2	35	270	
744	A11003-01 E		2/1/1966	0	CI	N	150	1966	4	2	114	0	
745	A11003-01 E		2/1/1966	0	CI	N	150	1966	4	1	0	0	
746	A11003-02				CI	N	150						
778	A11028				YDI	Y	300						
801	A11048-01 B		5/1/1959	0	CI	K	150	1959	8	1	0	0	
804	A11048-01 B		10/1/1990	0	CI	K	150	1990	39	5	94	56	
842	A11095-01 G		12/1/1982	0	PCI	N	150	1982	15	2	6	32	
843	A11095-01 AC		8/1/1987	0	PCI	N	150	1987	20	3	174	56	
844	A11095-01 G		1/1/1990	0	PCI	N	150	1990	23	4	187	29	
845	A11095-01 G2		7/29/2008	0	PCI	N	150	2008	41	6	2	0	
846	A11095-01 G2		7/29/2008	0	PCI	N	150	2008	41	5	211	223	
847	A11098-01				PVC	P	150						
848	A11099-01				CI	N	250						
878	A11133-02 G		11/1/1994	0	CI	K	150	1994	63	1	0	0	
879	A11133-02 CG		9/30/2010	0	CI	N	150	2010	49	2	3	242	
880	A11133-03				CI	K	150						
881	A11134-01				PDI	D	150						
882	A11136-01				YDI	Y	300						
883	A11137-01 G		1/1/1991	0	PDI	D	250	1991	22	1	0	0	
884	A11137-01 G		1/1/1991	0	PDI	D	250	1991	22	2	8	0	
885	A11137-01 G		8/1/1992	0	PDI	D	250	1992	23	3	17	19	

Figure 4-5: Multiple break records of a single pipe occurring in one day

Another common case of redundancy noted in the data is the existence of different lengths, diameters, installation year and material type for the pipe data recorded. The snapshot of these redundancies is given in Figure 4-6.

As can be seen in this figure, there exist two different diameters for some pipes. For example, for the case of pipe N7001 (with the font red in the figure), the record gives a diameter of 500mm under the heading DIAM_1 and diameter 600mm under the heading D2. Similarly, the

material for this pipe also is conflicting. Further as can be seen, two materials types namely concrete and CI is recorded for the pipe.

Further cases of different installation years for the pipes were noted. For example, the case of pipe no A5704-03 among others. Other special cases included a record of a break occurring even before the pipe was installed in the ground. For example, a pipe had broken in, say 2000 was recorded to have been installed in, say 2004.

All these cases may point to the fact that may be there were replacements made on the pipe and therefore, the records of the pipe was not deleted from the system.

1	VCIRC_1	LENGTH	MATERIA M_1	DIAM_1	YEAR_1	LONGEST	MAT2	M2	D2	Y2	LEN2	
12	A21044	322	CI	K	300	1950	282	CI	K	150	1950	40
13	A5358-01	101	CI	K	150	1950	95	PVC	P	150	2008	6
14	A5806-02	148	CI	K	150	1950	147	PVC	P	150	2002	1
15	N7001	287	CI	K	500	1930	268	CON	O	600	1994	14
16	N7001	287	CI	K	500	1930	268	CON	O	600	1994	14
17	A6041-02	146	CI	K	150	1950	146	CI	N	150	1960	0
18	A5704-04	46	CI	K	150	1951	29	CI	K	150	1950	17
19	A21031	130	CI	N	150	1960	108	CI	K	150	1950	20
25	A5704-03	107	CI	K	150	1951	68	CI	K	150	1950	39
26	A21031	130	CI	N	150	1960	108	CI	K	150	1950	20
27	A6041-01	173	CI	N	150	1960	173			0	0	0
28	A5892-03	152	CI	K	100	1942	152			0	0	0
31	A5184-03	121	CI	K	150	1950	121			0	0	0
32	N5853	196	CI	N	200	1962	196			0	0	0
33	M21015	271	CI	N	200	1962	164	CI	N	150	1962	108
34	N13356	87	CI	N	150	1955	79	CI	N	150	1960	8
36	A20438	136	AC	O	400	1959	97	CI	N	300	1959	15
41	N11747	173	CI	N	250	1961	161	PVC	P	150	1986	6
42	A8480-01	117	CI	K	150	1954	117			0	0	0
43	A8643-01	100	CI	K	100	1954	95	PVC	P	150	1996	5
45	A16177	21	CI	K	400	1953	14	PVC	P	150	2001	4
48	A5384-01	191	CI	K	150	1953	149	CI	K	150	1951	42
49	M24634	657	AC	O	200	1974	271	AC	O	150	1974	241

Figure 4-6: Conflicting records of lengths, diameters, installation year and material type for one pipe

Since redundancies that appear in Figure 4-6 affected most of the pipes, each of the record that had the full information on the pipe were only considered for analyses e.g. Year_1 was considered as it had no missing entries, therefore information from Y2 was disregarded.

After cleaning of the data, the pipe data were divided into three main groups according to the type of material giving rise to PVC, Ductile Iron, and Cast Iron pipe strata. Further subdivisions were done on the metallic pipes i.e. DI and CI where DI was divided into wrapped and unwrapped DI, and yellow jacketed DI commonly known as YDI. Further, CI was divided into thick CI (CI_TK) and thin CI (CI_TN) sub strata.

After processing of data, the file was converted to text file (.txt) for use in statistical analysis using R software.

4.3 Pipe Data Analysis

The water distribution network (in year 2012) consisted of 4,281km length of pipe with a total of 31,662 individual pipes. Systematic recording of the data on pipe breaks began in 1956 and the database contains 13,692 individual pipe breaks.

Presentation of the results of the pipe data for the CCWN follows the categorization of the pipe according to pipe materials as described in Section 3.2.1. This method was applied so as to appreciate the pertinent issues that affects each pipe material.

4.3.1 Distribution of Pipe Materials for the CCWN

CCWN has a variety of pipe materials that are installed all over their network that includes all Ductile Iron (DI_A) pipes (Yellow Jacketed Ductile Iron, and wrapped and unwrapped ductile Iron), Cast Iron (CI) pipes, steel pipes, Asbestos Cement (AC) pipes, and PVC pipes. However, steel pipes and the AC pipes, though in existence, are very few in the system and therefore, they are excluded from further analysis in this thesis.

Figure 4-7 depicts the proportion of pipe materials for the CCWN. It comprises of 11% Thick-Wall Cast Iron (CI_TK), 9.5% Thin-Wall Cast Iron (CI_TN), 10.8% Ductile Iron (DI), 14.5% Yellow-Jacketed Ductile iron (YDI), and 54% plastic (PVC) pipes.

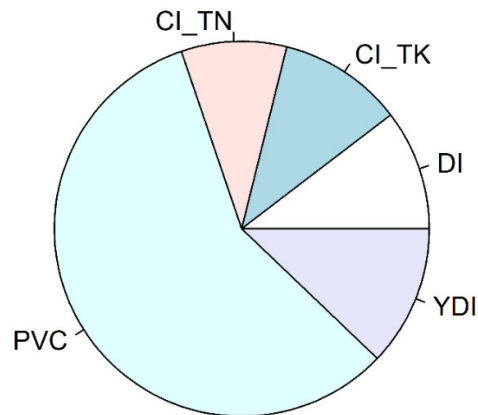


Figure 4-7: Proportion of pipe material for CCWN

CI_TN represents the mains installed from the year 1955 and after with a wall thickness of 10 mm and 12.5 mm for the 150 mm and 200 mm diameters, respectively (Brander and Ng 2000). On the other hand, CI_TK represents main material that was installed from the year 1954 and before and were 20% thicker than CI_TN (Brander and Ng 2000). YDI is known only in Canada and a 40 mm high-density polyethylene strongly bonded to the pipe representing main materials that are well protected against corrosion both with coatings and anode protection (Brander 2001).

Conversely, DI is main material that represents wrapped and unwrapped pipes, and there are various DI poly wrappings that protect the pipe from corrosion (Brander 2001). These wrappings include coatings composed of polythene wrap, zinc galvanization, thermoset coatings, and thermoplastic coatings, and cathodic protection (impressed current, and galvanic anodes) (Rajani and Kleiner 2003). From Figure 4-7, it is clear that the CCWN is tending towards rehabilitating their whole metallic system with non-metallic pipes. This can be seen from the proportion of the non-metallic pipe that has increased significantly since its inception in 1970's and has become the single largest component of the system as well as the most favoured material for pipe replacement (Brander 2001, 2004).

4.3.2 Pipe Laying for CCWN

Figure 4-8 presents the number of different pipes installed during different pipe installation periods. It can be observed that pipe materials usage in the city has since changed from metallic to PVC pipe over the years. Different pipe laying periods depict the ages of pipes, peculiarities associated with manufacturing and installation techniques, and types of material used (Makar et al., 2001).

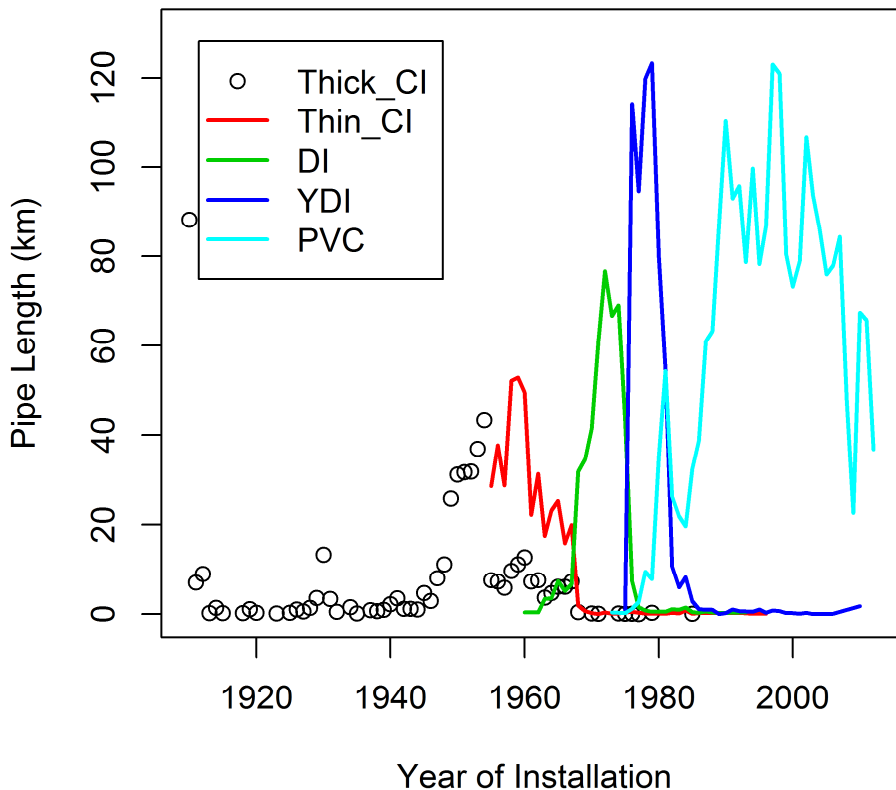


Figure 4-8: Pipe laying in different Pipe Manufacturing Periods

The first pipes installed in the network were the thick-walled CI pipes made by pit cast process and were later replaced by thin-wall CI made by spun cast process (Brander and Ng 2000; Brander 2001). It can be seen that the DI pipes became operational late 1960's. Literature material states that these pipes were manufactured when it was realized that the CI pipe was experiencing rapid breaks (Makar et al., 2001). It is further stated that the CI pipe, a brittle metal in nature, can record as high as 300 breaks in a year (Makar et al., 2001). To reverse this trend, DI pipe was introduced which was made of low phosphorous and sulphur with magnesium added and spun cast manufacturing process used so as to increase its strength

(Rajani and Kleiner 2003). However, it was later realized this pipe material breaks at high rate that can compare to the breaks from the CI pipe (Rajani and Kleiner 2003) which is as a result of its thinner wall thickness.

The upsurge of breaks from the DI pipes in the 70's due to corrosion related breaks when the pipe was between 7 and 10years old (Brander 2001) led to the introduction of YDI pipes, commonly known as the Yellow Jacketed Ductile Iron Pipe. As noted in the preceding sections, the YDI pipe is a heavily coated pipe with a 40-mm high-density polyethylene that is strongly bonded to the pipe. The YDI pipe is protected against corrosion by coatings (using cement lining, wrapping and oil painting) (Plate 4-1 (a)) and anode protection (Plate 4-1 (b)).

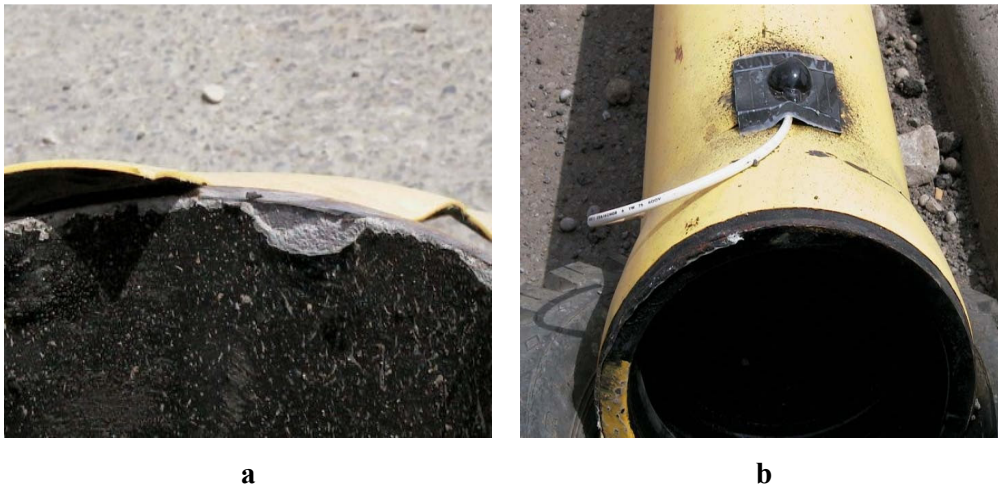


Plate 4-1: Cement Lining (a) and Anode Protection (b) for the YDI Pipe

The cathodic protection shown in figure (b) above necessitates placing of anodes after every 100m length of pipe laid and electrically bonding the connected pipes. As a result it provides service for about 20years before anode replacement is made (Brander 2004). The placement of anodes and critical avoidance of scratches on the pipe coating mean a costly exercise in laying these pipes.

The use of the PVC pipes to replace mains started in late 1970's and peaked towards the end of 1990's and since then the replacement program has been on the decrease. The shift was made because the city was still having corrosion problems caused by the heterogeneity of the soils of city of Calgary especially in the corrosive soils (resistivity less than 2000ohm-cm).

4.3.3 Break Rates for Different Pipe Materials

This section presents the break patterns for the pipe materials in the CCWN. Discussion of CI, DI, PVC pipes and the overall pattern for the city is presented in the following sub-sections

4.3.3.1 Cast Iron Pipe

Figure 4-9 presents the yearly break rate for the CI material type for the City of Calgary water networks system. The number of breaks experienced by this material increased significantly over a period of about 20 years. The breaks for the material was normalized over the length of the pipe. As can be seen in the figure below, the breaks from this material peaked at about 35 breaks/100km in 1985. In 1980's, it is observed that the rate of breaks was over 20 breaks/100km. Similarly, most of the breaks were observed to be in the range of 15 and 20 breaks/100km during the 1970's and post 1985.

However in the post 1985 era, it can be seen that there has been a reduction in the number of breaks observed for this pipe material. From late 1970's, the city began a massive replacement program that targeted metallic water mains prone to failure. Over this period several lengths of the CI pipes were pulled out of the system mainly in areas with highly corrosive soil and replaced by the PVC pipe material (Brander and Ng 2000; Brander 2001, 2004). The significance of this replacement program can be seen from Figure 4-9 where low breaks were observed.

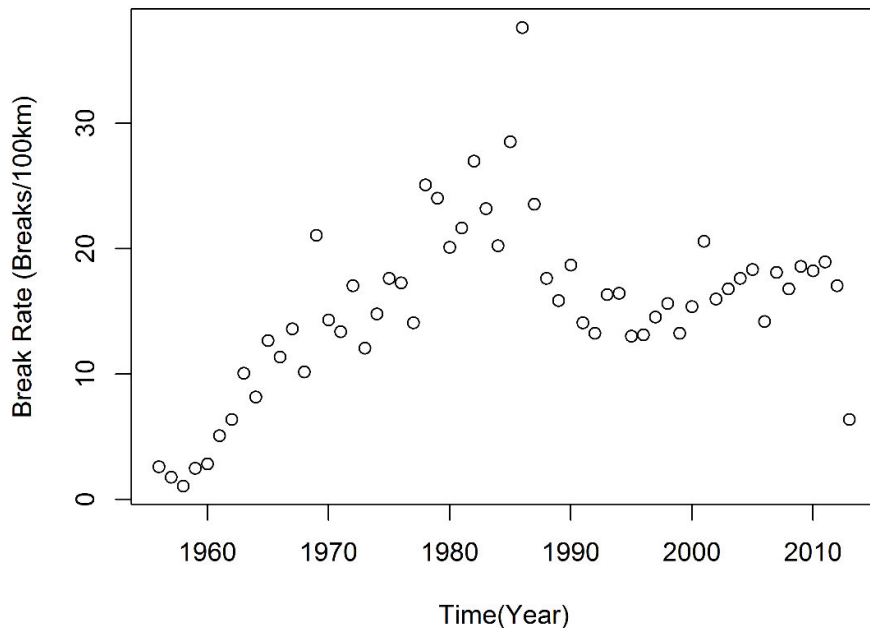


Figure 4-9: Annual break rate for CI pipe material

With the passage of time, and in the late 1990's to date, the breaks from this pipe has been steadily increasing. This resurgence of breaks for the CI pipe material is thought to be as a result of the aging pipe. It should be noted that the some of the pipes were laid in the early 1910's and the installation of this pipe material ended in 1960's. This finding may seem to point to the contribution of pipe age to failure of pipe at its final periods of its life as found by Harris and Radlinski (2013); Mordak and Wheeler (1988); White (1985). However, as can be seen from the figure below, CI pipe experienced many breaks at 'young' pipe age. Therefore, there seems to be a nonlinear relationship between age of the pipe and its break rate. This finding is seem to be in agreement with the findings by Andreou (1986); Boxall et al. (2007); Gowlter and Kazemi (1989); O'Day (1982).

Generally, the rapid breaks observed for CI material is thought to be due to the thinner pipe wall thickness and low material strength as compared to other pipe materials. The pipe was designed to only withstand internal water pressure and crushing forces (Makar et al. 2001) with the latter forces assumed to be only from the loading above the pipe. It has however been found that ground movement due to expansive soils and frost loading may also cause this loading. This omission of loading forces in the design of the CI pipe may be the cause for high failure rate observed in this pipe.

4.3.3.2 All Ductile Iron Pipes

Figure 4-10 presents the breaks rate of the for the All Ductile Iron pipe (DI_A) for the City of Calgary. As can be seen from the figure, the number of breaks for this pipe increased considerably over a period of less than 20years. As in the case of CI pipe, the breaks of this pipe material was normalized over its length.

As can be seen from the figure, the break rate for the pipe began considerably at a low rate for about 8years, then increased significantly over a period of 10yrs and peaked at about 30breaks/100km in the year 1985. Between the year 1980 and 1985, the break rate for the pipe material was between 15breaks/100km and 20 breaks/100km. These years represent the time when most of the breaks in the city was observed for this pipe material. On average, most of the breaks observed for this pipe has been around 15breaks/100km. As noted in the Figure

4-10, there is sharp increase in the number of breaks for this pipe in the early 1980's up to mid-1980. This increase was deemed to have been caused massive number of breaks observed in pipes that were installed in highly corrosive soils (Brander 2001) bearing in mind that DI pipe is highly prone to corrosion.

Further from the figure, it can be seen there was a drop in the number of breaks for the pipe after the year 1985 which began to rise in the early 1990's. The massive rehabilitation program carried out in the city that targeted the pipes prone to failure in the network contributed significantly to this drop. Also, before the introduction of the PVC pipe, the pipes prone to breaks were replaced with the YDI pipes (Brander and Ng 2000). Moreover, cathodic protection was done for most of the DI pipes and this may further explain the reduction of the breaks. Therefore, these activities might explain the reason to why the breaking trend of this pipe material has been on the decline for last decade.

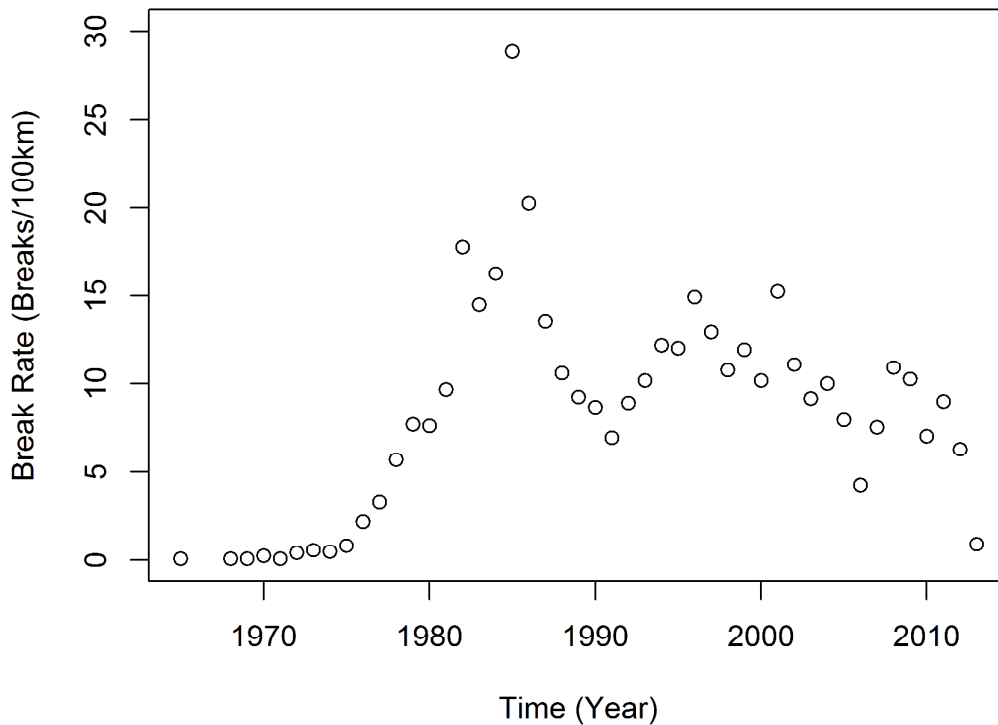


Figure 4-10: Annual break rate for DI_A Pipe Material

Further, the figure above reveals an interesting trend in the failure of the DI_A pipes. It can be observed that the failure rate of the DI pipe is almost comparable to the break rate for the CI pipe with break rates for both pipes peaking almost at the same time. The introduction of the

DI pipe meant that the dominance of the CI pipe came to an end. Made to withstand mechanical loads observed as a weakness in CI pipe, the DI pipe was made with smaller wall thickness than its equivalent CI pipe (Rajani and Kleiner 2003) thereby causing the pipe to record high number of breaks. The pipe is mostly prone to corrosion related breaks and thereby graphitization and pits can be observed in young pipes.

The increase in the number of breaks observed from early 1990's is due to reduction of the replacement program for this pipe in favour of the CI pipe which was having more breaks (Brander and Ng 2000).

Similar to the breaking pattern for the CI pipe, age of the DI pipe seem not to contribute linearly to the failure of the pipe. This result was seen as a result of the upsurge in the rate of break for the pipe while still 'young'.

4.3.3.3 PVC Pipe

Figure 4-11 presents the failure rate for the PVC pipe for the city of Calgary. As can be seen in the figure, the failure rate for the pipe material has been constant over time. It should be noted that the number of breaks for the pipe has been normalized using the overall length of these pipes.

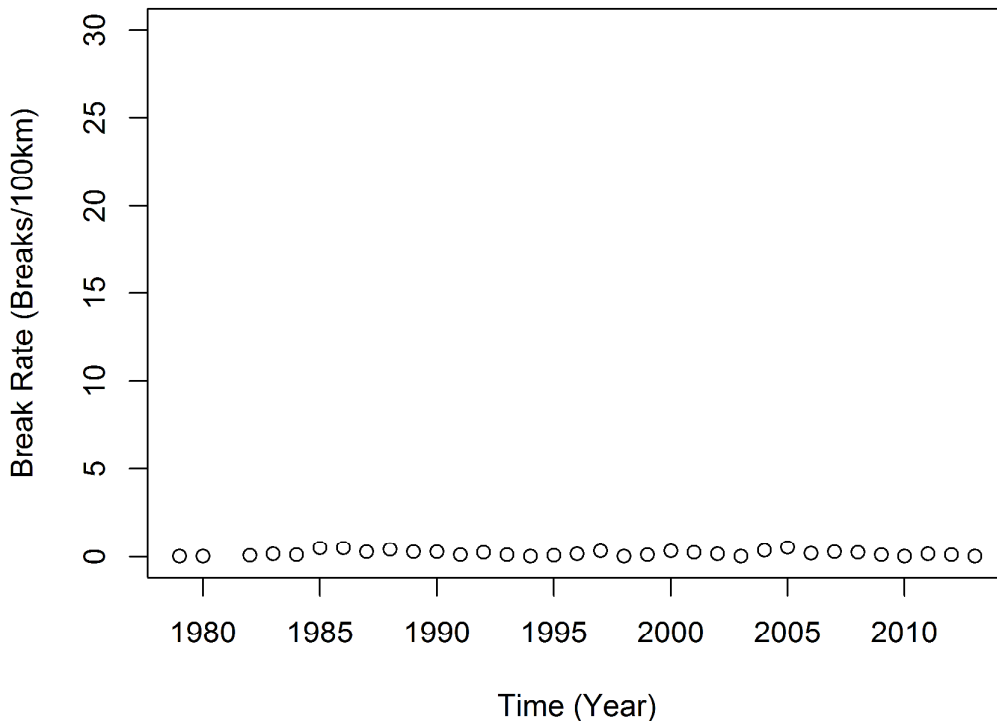


Figure 4-11: Annual break rate for PVC pipe material

As can be seen from the above figure, the rate of break for this pipe material has remain less than 1break/100km. the overall break rate for the material is 0.4breaks/100km. this result means that over the length of the entire water network, this pipe material has experienced less than 200breaks. Brander (2001) observed the same breaking trend and reported a break rate of 0.2breaks/100km.

The success of this pipe material as a suitable material is evident in the number of breaks observed for the pipe material and the increase in the percentage of coverage by this pipe for the city. The CCWN has since then adopted the use of this material as the main replacement. It is further thought that the resistance of the material to abrasion and many strong acids and caustics (Knight 2002) has contributed significantly to the material being used as replacement material for the city. However, Brander (2001) noted that though the breaks are rare, when they occur, they can be costly and at the same time damaging.

4.3.3.4 Overall Break Rate for CCWN

Figure 4-12 presents the yearly break pattern for CCWN. The utility has experienced a steady increase in the number of breaks in the first 30yrs of data recording. The break rate for the city was normalized over the entire length of the water network system.

As can be seen in the figure, the break rate for the city has risen rapidly for the first 30years and peaked in the year 1985 at a break rate of about 15breaks/100km. the breaks rate in 1970's increased considerably and was at an average of between 5breaks/100km and 10breaks/100km. the break rate for the 1960's and below was at below 5breaks/100km. After the year 1985, there was a significant reduction of breaks for the city up to early 1990's. This reduction can be attributed to removal metallic pipes around the city that were prone to rapid failure.

According to Pelletier et al. 2003, a utility can be classified as being in a good, acceptable and poor or degraded state if it has break rates/100km of below 20, between 20 and 39, and 40 and above respectively. From Figure 4-12, it may be concluded that the general condition of the system is in a good condition. From the analysis of the metallic pipes in the previous sections, it can be assumed that the break rate of the CCWN is mainly contributed by the metallic pipes

due to their rapid breaking patterns. It is nonetheless safe to assume that the good condition observed in the network is as a result of the PVC pipes which have the lowest breaking rate.

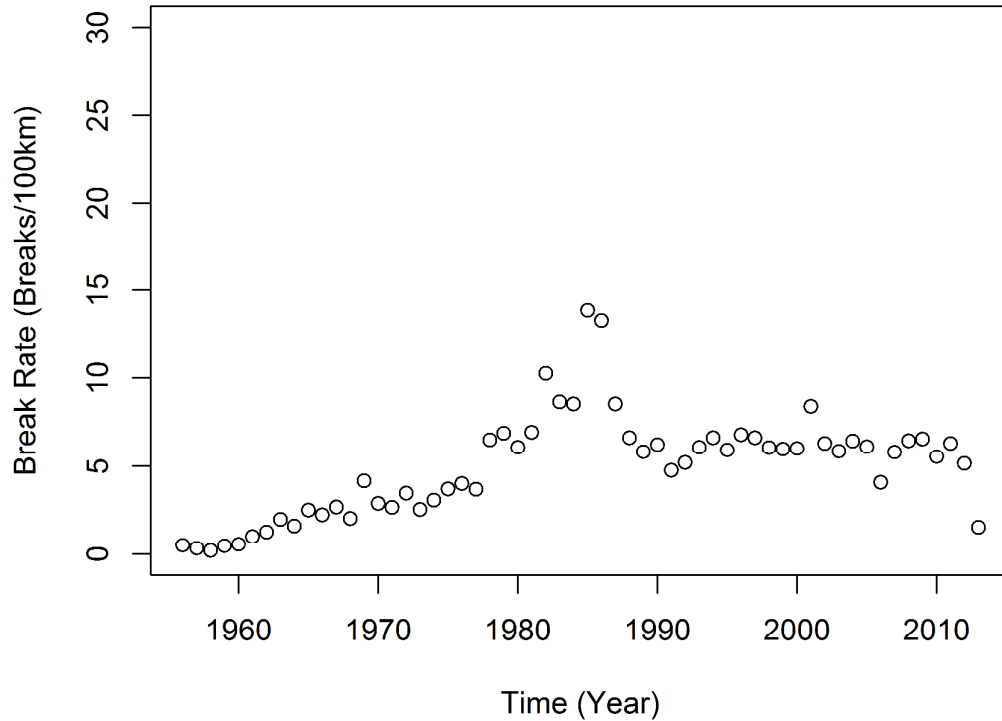


Figure 4-12: Annual Break Rate for CCWN

For the last 25 years the water utility has experienced a near constant break and for the last 3yrs a decreasing trend with an average of 7breaks/100km. This trend may be attributed to the use of PVC pipes (\cong 54% of water pipes) which has experienced less than 1 break per km/year since its introduction 35yrs ago; and a replacement program that has targeted pipes showing signs of high break patterns (Brander and Ng 2000). Also, contributing to the downward trend is the retrofitting program (cathodic protection) undertaken on DI pipes in the system.

4.3.4 Effects of weather parameters on breaks

4.3.4.1 Monthly Break Rate for CCWN

Figure 4-13 shows the breaking trend for the CCWN when the number of breaks are clustered according to the month when the break occurred. The figure reveals an interesting pattern of how the breaks occur.

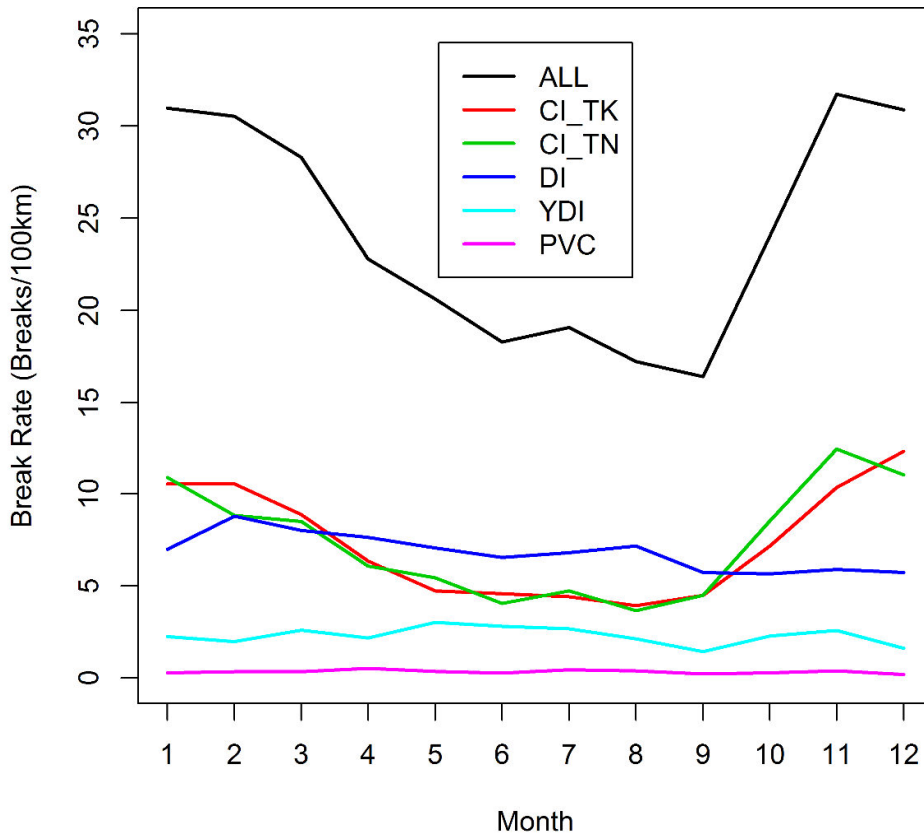


Figure 4-13: Monthly break-rate for CCWN utility

From the figure above, it is clear that different pipe materials respond to climatic changes differently. The pipe material that shows major variability with respect to its breaking pattern is the CI pipe. Both of the subsets of the pipe materials have a similar pattern with most of the breaks occurring during winter period and least breaks occurring in the summer. The breaks for both CI_TK and CI_TN materials begin to drop considerably as the spring sets in and reaches the lowest breaking rate in the month of October. This lowest point coincides with the end of summer when the soil has the lowest moisture content. In the same note, the breaks for

these materials begin to increase from the beginning of autumn and peaks in November before levelling off up until March when it starts to drop. Therefore, this finding seem to confirm that the ci pipe fails through circumferential breaks due to soil movement that result from the changing season patterns.

On the contrary, the DI and the YDI pipes does not seem to be mostly affected by the season pattern. As can be seen, there breaking trends seem to be uniform throughout the season. This result seem to support previous findings that the DI pipe fails mostly as a result of corrosion. This is so because it was made to withstand the external loads of a pipe. Similar to the breaking trends by the DI and YDI pipes, the PVC pipes have a uniform breaking pattern throughout the season. This result therefore seem to conclude that the pipe is not affected by the weather pattern.

The ALL line in the figure represents the breaking pattern of the CCWN for the entire season. It can be seen that the breaking pattern for the city is influenced by the CI pipe. This result is shown by the similarities of the breaking pattern for the whole City and the CI pipes. From the figure above, it is clear that from the month of March, the breaking rate for the city begin to drop considerably and reaches the lowest breaking rate in the month of October. This lowest point coincides with the end of summer when the soil has the lowest moisture content. In the same note, the period when the break rate begin to drop, it coincides with the end of winter. However, the number of breaks increases considerably from the end of summer and immediately peaks in the month of November before levelling off in December through February. The city records a lowest break rate of about 18breaks/100km while the highest break rate being 31breaks/100km.

The findings from this study reveal that the weather parameters have a significant impact on the break rate of a water network. It has been observed by many others that during winter, the colder temperatures causes more frost loads to be transmitted to the pipes thereby causing more breaks. The frost loads have been noted to be transmitted to as far as 2.4m below the ground during winter (Brander 2004). Therefore during the severe winter months, these loads targets the weaker pipes thereby causing high and repetitive break in these pipes (Brander and Ng 2000; Habibian 1994).

However during summer, there is considerable loss of moisture in the soil which makes the soil shrink. This movement of soil causes uneven loading in the pipe (Chan et al. 2005; Friedl et al. 2012). The study undertaken by Rajani et al. (1996) revealed that loading of the pipe increases when there is movement in the soil.

Findings from this study is in agreement with studies undertaken by other authors e.g. (Friedl et al. 2012; Harvey et al. 2013; Hu and Hubble 2007) who found out that more breaks occur during winter than in the summer.

4.3.4.2 Effects of Freezing Index

Figure 4-14 shows the evolution of the monthly breaks for various pipe materials for the CCWN in relation to Freezing Index (FI). The basis for this analysis is that a pipe material responds differently to weather patterns. FI has been included in this analysis as described in Section 3.2.1.

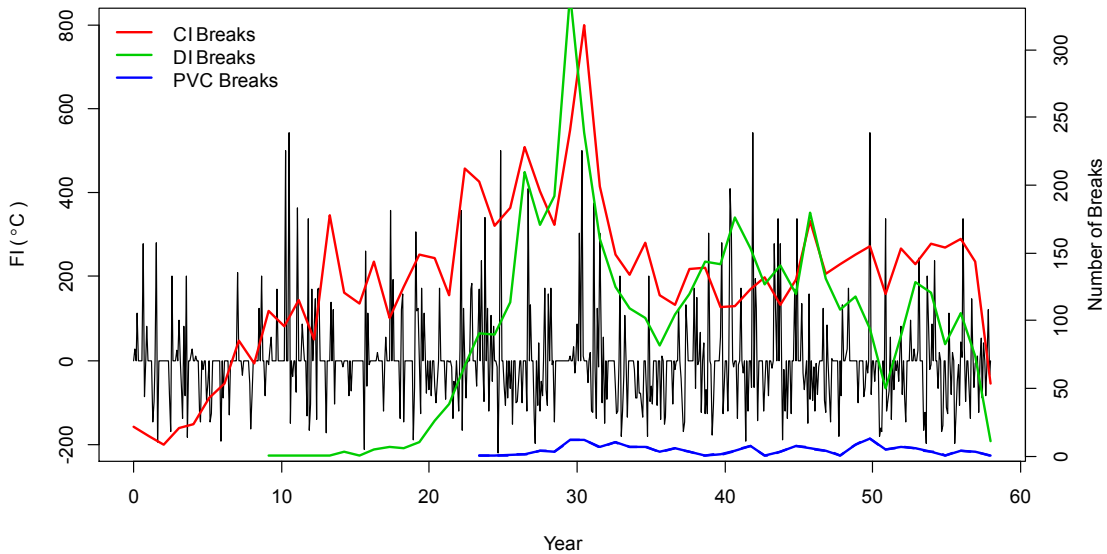


Figure 4-14: The effects of temperature on pipe break rate of CCWN utility

As can be seen in the above figure, the breaks of the city follows a cyclic pattern whereby high number of breaks occur when the temperature of the area is very low. The highest break for both CI and DI_A pipes was recorded after 30 years since recording began. Conversely, less breaks are being observed when the temperatures are high. Further, the figure indicates that the number of breaks experienced varies from season to season. From the trend of the breaks, it

seems that after a period of time, mostly in the range of 2-4years, the city experiences high number of breaks and this mostly affects the metallic pipes. Habibian (1994) explained that every season, the system is purged of the weaker pipes that resisted the occurrence of the breaks in the past season. Therefore, the spikes observed in this network is as a result of effects of the frost loads on the weaker pipes in the system.

The most distinctive result in this figure is the nature of the breaking trend for the PVC pipe material. It can be seen that the trend is near a constant which tend to indicate that it cannot be affected by different freezing patterns. The main reason for this trend is with the flexibility of the pipe. Though the non-metallic pipes have high coefficient of expansion, they cannot be much affected by the thermal changes due to their flexibility (American Water Works Association 2007).

4.3.5 Pipe Diameter in Relation with Material

Figure 4-15 presents the relationship between pipe diameters and different pipe materials for the CCWN. The number of pipes was normalized by the overall network pipe length, but not the individual material length.

As can be seen in the figure below, the proportions of metallic pipes decreases with the increase in diameter with more pipes having a diameter of 150mm. This high proportion of these pipes explains the high breaks experienced by the metallic pipes given in Section 4.3.3. Research (e.g. Kettler and Goulter, 1985, Mordak and Wheeler, 1988) has shown that small diameter pipes exhibit a high tendency of break rates than large diameter pipes because of thin pipe walls and small moment of inertia.

Further from the figure, the proportion of pipes by material reveals that CI pipes (CI_TK and CI_TN) are in the category of 150mm with proportions decreasing with increase in diameter. However, a closer look of the proportions for both CI_TK and CI_TN reveals that these pipes have almost the same proportions for the 150mm diameter with CI_TK having higher proportions for the larger diameter categories.

For the DI pipes represented by DI (wrapped and unwrapped) and YDI pipe subgroups, most of the pipes are in the larger diameter category of between 200mm and 300mm. However, analysis of the subgroups for these pipes reveals that both DI and YDI have almost the same number of pipes for the 150mm diameter pipes with YDI have more proportions for the larger diameter categories. The high proportion of the YDI pipes explains the reason for the drop in the number of breaks that the city has observed especially for all the DI (both YDI and DI) pipes.

In all diameter categories, PVC has the highest number of pipes. The most proportion of these pipes are found in 200mm diameter category with the least proportion found in the 100mm diameter. This finding seem to further confirm the choice of this pipe material as the replacement for the City of Calgary.

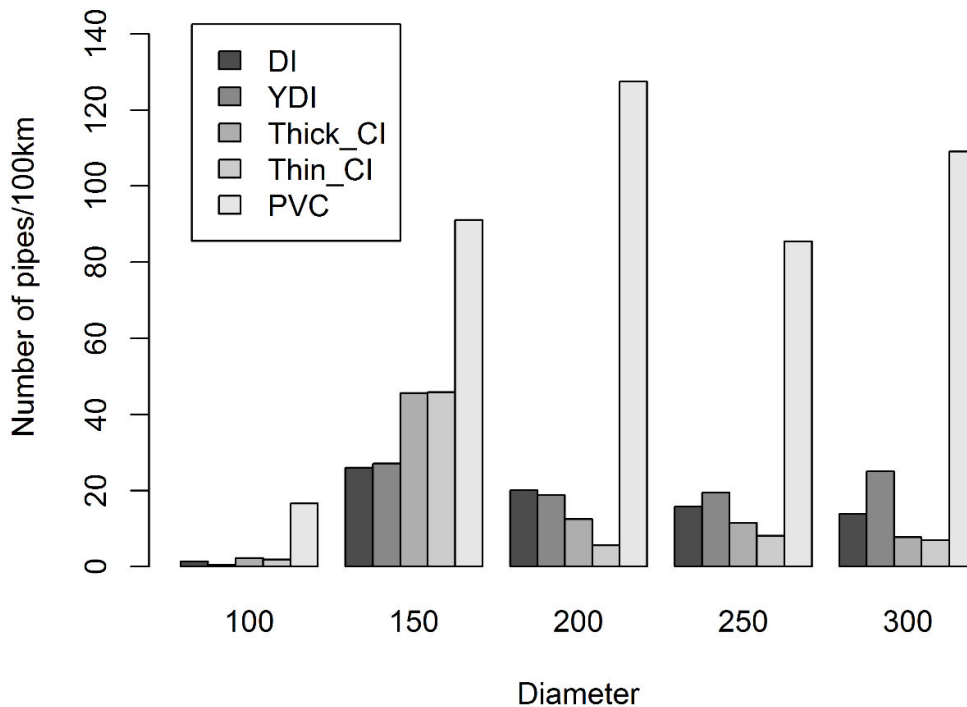


Figure 4-15: Pipe material in relation to diameters in the utility

Figure 4-16 shows the relationship between break type and pipe materials. As can be seen from the figure, each pipe material has a different breaking pattern. It indicates that both CI_TK and CI_TN pipes mainly failed as a result of circular break. This finding seem to confirm the result by Makar (1999); Rajani et al. (1996) and Habibian (1994) who found that mostly the CI pipes fail mainly due to circular breaks imposed by soil movement. Further, it can be seen that the

proportion of circular breaks for this material is higher in CI_TN than in the CI_TK pipes. This is attributed to CI_TN having thinner walls than its counterpart; in addition, weak points at both end sections of this pipe (Plate 4-2 and Plate 4-3) were created during manufacturing process making them disposed to circular breaks.

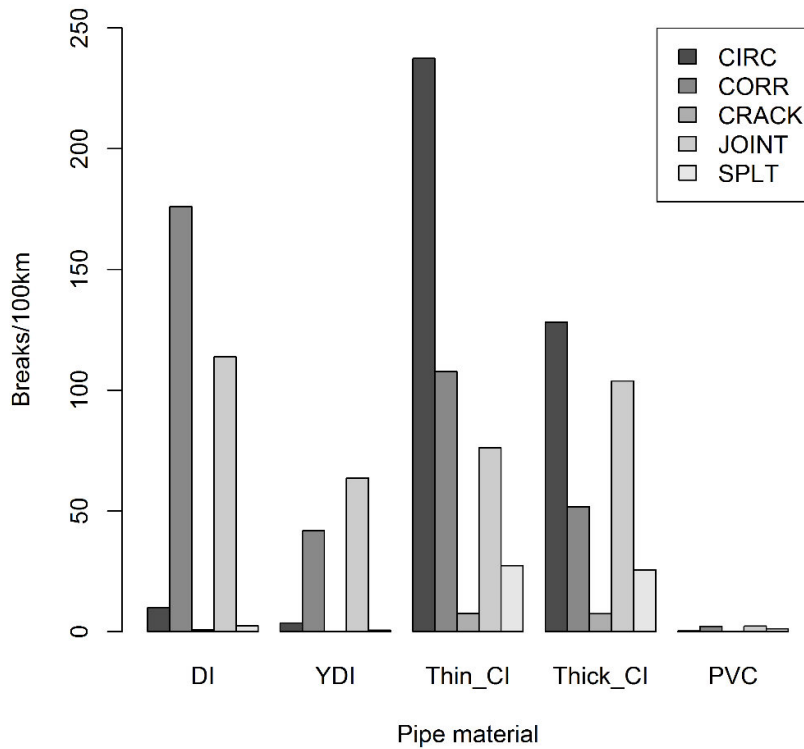


Figure 4-16: Relationship between pipe material and break-type

However, the DI pipes failed more as a result of corrosion while the YDI pipes failed as a result of leaking of joints and fittings. Studies have indicated that the DI pipes are more prone to failure by corrosion and can corrode at rates comparable to those of CI pipes. Therefore, the finding of this study seem to confirm this finding. Further, it is thought that the thinner walls compared to its counterpart pipe of similar diameter makes it prone to corrosion. It has been observed that the susceptibility of the wrapped and unwrapped DI pipes differs with 30% less breaks observed in wrapped than in unwrapped pipe (Brander and Ng 2000; Brander 2004).

Furthermore, it is noticeable that the PVC pipes seem not to be affected by any of the break-types affecting metallic pipes. Over a database of more than 18,000 pipes, only 0.7breaks/100km/yr have been observed. Research has revealed that this pipe material is resistant to abrasion, strong acids and caustics (Knight 2002) making it adaptable to different

micro-environments in the soil. Results from this pipe material seem to confirm findings by other authors e.g. (Davis et al. 2004) and interestingly there are cases where it has been reported that the rate of break for this pipe material decreases with increase in time (Davis et al. 2004). However, though with low breaking rates, concerns have been raised on the costly and damaging nature of the breaks observed for these pipes as compared to metallic pipes (Brander 2004).



Plate 4-2: Corrosion band located from 3'5" from end 9A



Plate 4-3: Corrosion band located from 3'4" from end 41A

Figure 4-17 depicts relationship between break type (i.e., circular (CIRC), corrosion (CORR), crack, joint, and split) and pipe diameter. From the figure, it can be seen that different pipe diameters are affected by the break-types differently. The figure reveals that the smaller diameter pipes are the most vulnerable pipes with the large diameter pipes effectively resisting most of these break-types.

From the analysis, it can be seen that majority of breaks occurred in 150 mm diameter pipes with the most significant break-type being circular or circumferential break, followed by corrosion, leaking from the joints, splitting and finally cracking. For the larger diameters, the most significant break-type seems to be leaking from the joints and corrosion.

The high proportion of the circular break-types is thought to be as a result of differential movement of expansive soils in the study area that results in volume change when subjected to wet and dry conditions (Rajani et al. 1996). The bedding material and the backfill used to fill the trench where the pipe was placed play an important role in propagating this break-type (Makar 2000). In the past studies, it has been revealed that this break-type mostly affect the smaller diameter pipes and contributes to about 70% of all the breaks occurring in a water network (Rajani et al. 1996). The result from this thesis therefore seem to be in agreement with the past studies.

For the corrosion break-type, though it seems to significantly affect the larger pipe diameters, it also contributes to the high number of breaks recorded in the smaller diameter pipes. As can be seen from the figure, this is the second most cause of breaks for the 150mm diameter pipe. This is as a result of smaller wall thickness and low moments of inertia that small diameter pipes have. However, the literature state that all the break-types are somehow associated with the corrosion through the formation of graphitization and pits (Makar and Kleiner 2000; Makar 2000; Rajani et al. 1996). It has been noted that the effects of this break-type increases with increase in the age of the pipe.

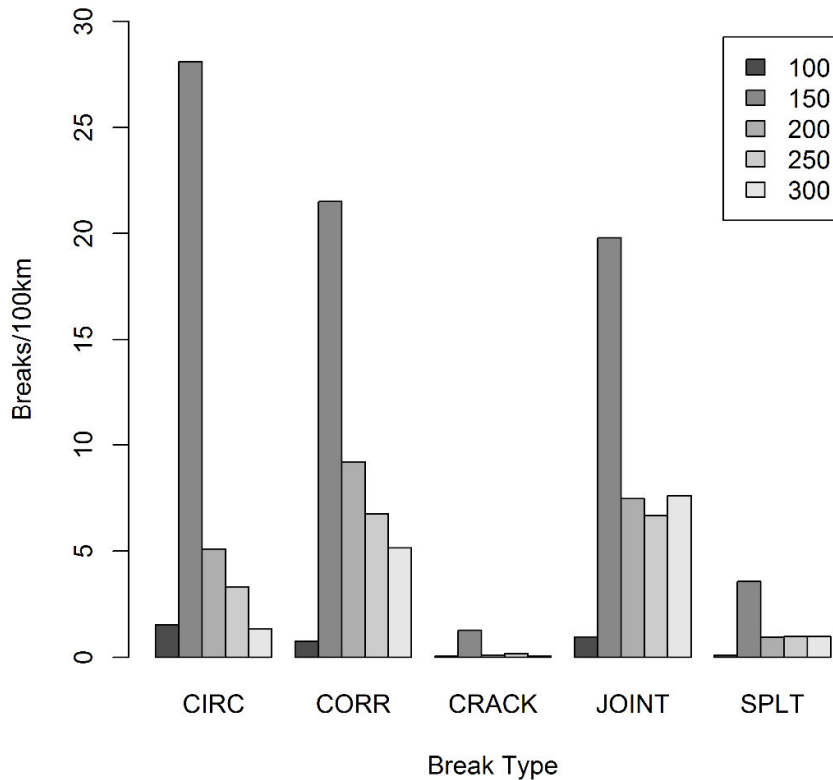


Figure 4-17: relationship between pipe diameter and break-type

Leaking of pipe joints, though mostly affecting the 150mm diameter pipes, have a significant effect on the break of the 200-300mm diameter pipes. Expansion and contraction of the pipe leads to the separation of the pipe thereby resulting in leaking pipes. This break-type when it occurs, causes erosion of bedding material and therefore may result in differential soil movement leading to circular and other break-types to occur. When the pipe has higher structural integrity, high moments of inertia and thicker pipe walls (large diameter pipes), the most vulnerable point is the pipe joints.

4.3.6 Spatial Distance and Break Occurrence

Figure 4-18 shows the relationship between the spatial occurrences of a break in relation to the previous break occurrences. This figure shows that the likelihood of a break occurring in the same vicinity as the last break decreases as the distance from the previous break increases. But within the first 10m, the number of breaks increases with the distance, then it subsequently decreases.

As can be seen in the figure, CI_TK, CI_TN and DI pipes almost have similar spatial breaking trend with YDI having a very low breaking trend. The DI pipe seem to have a higher re-breaks of over 200breaks when the distance from the last break is closer to zero. However, as the distance from the last break incidence increases, the number of breaks experienced by both the sub-groups of CI pipe overtakes the breaks from the DI pipe. These two pipes, CI_TK and CI_TN, experience breaks of over 280 and 240 breaks within the distance of about 10m with DI experiencing about 220breaks within the same distance.

Though the YDI pipe material show a similar trend as the other metallic mains, its influence by the occurrences of other breaks within the vicinity of the last break seems to have a little impact on the pipe. The breaking trend for the first 15m shows a constant breaking rate of about 49breaks which then decreases to about 25breaks after 25m.

Approximately, 33% of the breaks occurring after incidence of a break are within 10m of the first break. Hu and Hubble (2007) and Goulter and Kazemi (1987) reported that soil disturbance during pipe repair practice contributes significantly to the occurrence of another break within the same vicinity of the previous break. However, as the distance from the last break increases, the spatial distances ceases to be of influence to the occurrence of breaks and therefore, it is thought that other factors play more roles in the occurrence of breaks in the pipe. The findings from this thesis seem to conform to the finding found in the literature.

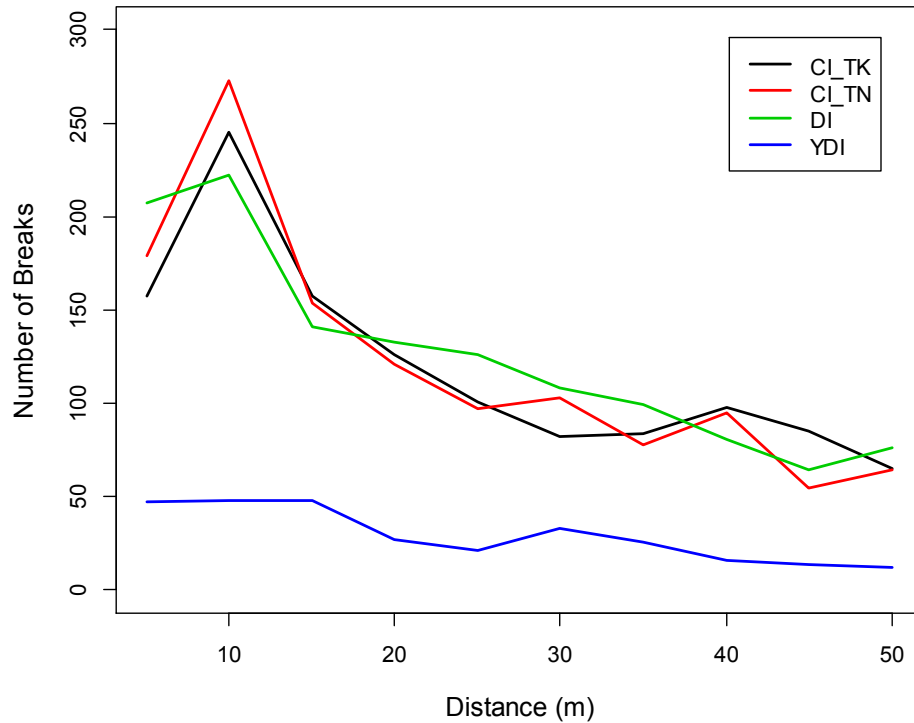


Figure 4-18: Spatial distance in relation to pipe breaks

4.3.7 Influence of Pipe Previous Failure History

The influence of previous pipe history was analyzed to establish its importance in the occurrences of breaks in the same pipe material. This was done by analyzing the number of breaks that the pipe has occurred after the occurrence of a break. The results of this analysis are presented in Table 4-1. This table shows the percentages of the number of breaks experienced by a pipe with respect to material after the occurrence of the first break. It should however be noted that the 7th re-break has been calculated by summing together all the breaks greater or equal to seven. It was expected that the percentage of the 7th and subsequent breaks should be much lower compared to the preceding breaks percentage.

Table 4-1: Number of breaks experienced by a pipe with respect to material type

Material	Percentage of Number of breaks					
	2 nd breaks	3 rd breaks	4 th breaks	5 th breaks	6 th breaks	>7 th breaks
CI_TN	32.3	18.6	10	5.4	2.9	3.8
CI_TK	24.0	13.8	7.9	4.3	2.47	4.8
DI	24.0	15.3	9.7	5.8	3.4	4.6
YDI	6.2	2.9	1.2	0.7	0.5	1.5

From the above table, it can be deduced that after the incidence of the first break in a pipe, there is higher chances of the same pipe failing again. This is highlighted by the higher percentages of the second break for all the pipe materials. However, it can be seen that the chances of pipe experiencing more than one break decreases as the number of breaks increases. It is assumed that as the pipe experience more breaks, there is likelihood that the pipe is replaced, thus eliminating this pipe from the system.

From Table 4-1, it can be highlighted that after the first break, the CI_TN pipe has the highest number of pipes with 2nd, 3rd and 4th breaks of 32.3%, 18.6% and 10%, respectively. This result highlights the vulnerability of the thin cast iron pipe to the occurrence of breaks in the water mains. CI_TK and DI also have a higher chance of re-breaking once the failure has occurred with both pipe materials having a percentage of 24%. However, the failure rate of the CI_TK pipe for the 3rd and subsequent breaks decreases as compared to the DI pipe. This result seem to confirm the vulnerability of the DI pipe which was initially thought to be stronger than the CI pipe. Past studies have confirmed that the DI pipe is more vulnerable much like the CI pipe due to thinner walls. The reason for the lower re-break percentages for the CI_TK pipe compared to the di pipe is due to the CI_TK having thicker walls.

Considering the YDI material, it is apparent that the pipe has a lower breaking trend of 2nd and subsequent breaks compared to the other pipe materials. From the table, it can be seen that the percentage of second break occurring is 6.2% which is 5 times lower compared to CI_TN pipe. It is thought that the main reason for the low re-break rate of the pipe is due to the cathodic protection and coating of the pipe which in the end prevents the effects of the break-types on the pipe.

Additionally from the table, it becomes apparent that the percentage of DI pipes with 5th and more breaks surpasses all other breaks from other pipe materials. This insinuates that the more the pipe stays in the ground, it becomes somehow uneconomical to sustain it therefore it should be replaced.

4.3.8 Influence of Type of Service Area on Number of Breaks

Two type of services are considered in this thesis for the analysis of breaks. These types are commercial and residential units. These types are thought to have influence on the number of breaks in the City of Calgary.

4.3.8.1 Commercial Units

Figure 4-19 shows the relationship between number of breaks and number of commercial units served by pipes in the network. The number of breaks considered for this analysis in this thesis have been normalized by the length of the pipe network. From the figure, it can be seen that the number of breaks observed decreases as the number of commercial units' increases. Therefore, there is an inverse relationship in the number of breaks observed and the number of commercial units served. The figure shows that most breaks occur to pipes serving less than five commercial units with the highest breaks being observed in the pipes serving only one commercial unit. However, it was expected that most of the breaks would occur to pipes serving high number of commercial units.

As can be seen from the figure, different pipe materials have different breaking trend for this factor. As expected, the most affected pipe is the CI_TN followed by CI_TK, DI and YDI, respectively. The CI_TN, CI_TK, DI, YDI have break rate of about 100breaks/km, 83breaks/100km, 50breaks/100km and 31breaks/100km, respectively, when the number of commercial units served is averagely two. The high number of breaks observed for low commercial units served is thought to result from small diameter pipes being used to supply water. As the number of commercial units increases, the diameter of the pipe supplying the water increases. Literature indicates that there is an inverse relationship between the diameter and number of breaks with small diameter pipes experiencing high number of breaks.

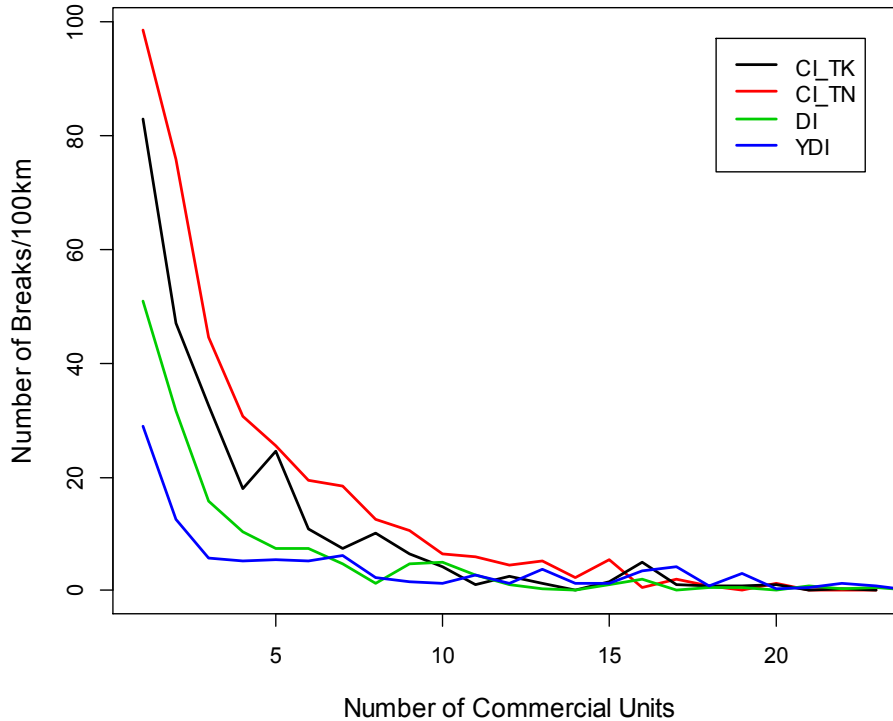


Figure 4-19: Pipe breaks observed in commercial units

4.3.8.2 Residential Units

Figure 4-20 shows relationship between number of breaks and number of residential units served by CCWN. The figure shows that the number of breaks increases as the number of residential units served increases and begins to decrease when the number of residential units served is more than 10 units. Most of the breaks observed are in pipes serving two to thirteen residential units with the least breaks being observed in pipes serving more than thirteen residential units. This high number of breaks observed in this category may be attributed to surge in water pressures in the affected pipes. Since most of the residents of the City of Calgary work during the day, when people return from work, it is believed that the pressures build up during night due to usage of water almost at the same time.

Further analysis of the response of different pipe materials to the breaking trend of this service type revealed that the most affected material is the CI_TN when the number of units served is less than five units. This is then followed by CI_TK, DI, and YDI pipes, respectively. The highest number of breaks for the pipe materials CI_TN, CI_TK, DI, and YDI are approximately 54 breaks/100km, 45 breaks/100km, 42 breaks/100km, and, 18 breaks/100km,

respectively. The number of breaks observed on these respective pipe materials peak when the number of residential units are 4, 7, 4 and 8 respectively.

However the breaking trend for these pipes changes as the CI_TK experiences more breaks compared to all other pipes when more than five residential units are served. The trend for the CI_TN reduces to almost below the breaks from the DI pipe. Further, the number of breaks from YDI pipe increases when the number of residential units surpasses 13 and rises past the breaks from all the pipe materials. It is thought the sudden increase in the number of breaks for this pipe materials is as a result of dominant use of the material in supply of water to residential units serving many units.

Generally, it seems that number of residential units served by water mains is also a factor on the number of breaks observed by a pipe. This is so because, it determines the size of the pipe to be used in supplying the residents clean and safe water. Therefore, when there number of residential units to be served are few, this will mean that the pipe size to be chosen will be a small diameter pipe. From the analysis carried out in the preceding sections, it is apparent that there is an inverse relationship between pipe size and number of breaks observed. Therefore, it is thought that the number of breaks observed in the water mains for 15 and more residential units is essentially due to large pipe size being used.

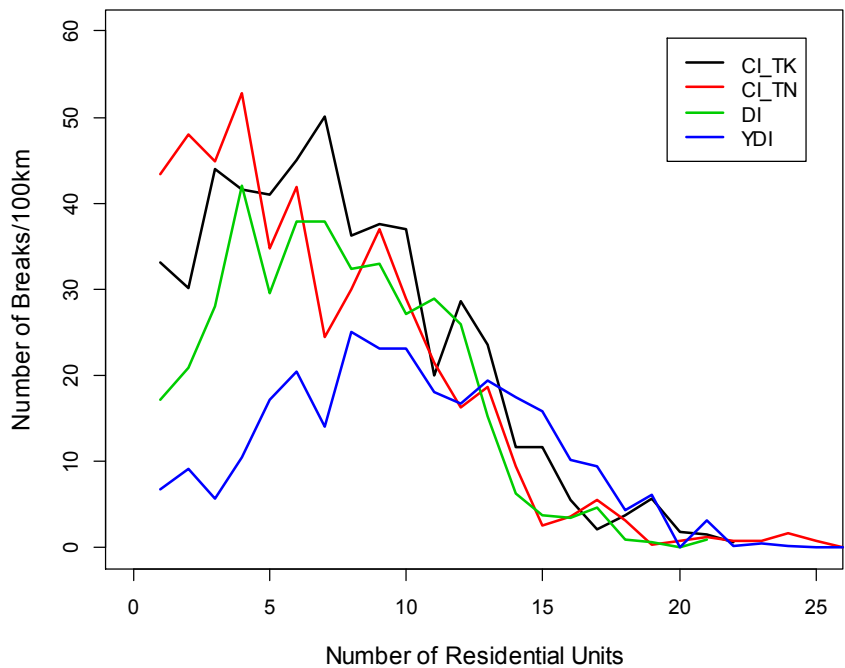


Figure 4-20: Pipe Breaks Observed in Residential Units

4.3.9 Cathodic Protection

Cathodic protection has been applied to some metallic pipes within the network. This was done when the number of breaks for the metallic pipes were observed to sky-rocket, and was done in areas where the soils were mostly aggressive. The aggressive soils are represented by the low resistivity of below 2000ohm-cm.

Figure 4-21 shows the ratio of retrofitted pipes to those non-retrofitted in relation to material type. As can be seen in the figure, most of the pipes retrofitted are DI and YDI with while the least retrofitted pipes were the subgroups of the CI pipes. It is apparent that the most pipe that is fitted with cathodic protection as expected is the YDI pipe which has a ratio of 1. This is so because, in its manufacture, the pipe was fitted with anode nodes to prevent the effects of corrosion on the pipe material.

The second most retrofitted pipe is the DI (wrapped and unwrapped) pipe which has a ratio of retrofitted to non-retrofitted pipes of 0.3. Since the pipes under this sub-group of the overall DI contain the pipes bare pipes, coated pipes, and furthermore, retrofitting was done on this pipe. Moreover, when the CCWN realized that the DI pipe was most affected by corrosion, cathodic protection was applied to these pipes in order to counter the rate of breaks encountered in these pipes (Brander and Ng 2000).

On the other hand, very small proportion of CI_TK and CI_TN pipes were protected against corrosion by retrofitting. In the case of the CI_TN pipe, a ratio of retrofitted to non-retrofitted pipes is at 0.1. This imply that the number of pipes that are not fitted with cathodic protection are more than those not fitted with it. Brander and Ng (2000) stated that they only embarked on retrofitting this pipe material in 1990's when it was observed that the number of breaks observed for these pipes was rising. This led to halting of all retrofitting process for all other pipes in order to concentrate on these pipes. Considering the CI_TK pipe, the ratio of retrofitted pipes to non-retrofitted pipes stands at less than 0.1. This is so because, most of these pipes were majorly not affected by breaks as compared to the CI_TN pipes. Generally, these pipes has a lower break rate as compared to their counterpart pipes of the same diameter.

The low levels of retrofitting observed for DI and both CI_TK and CI_TN points to the fact that most of these pipes were pulled out of the system when it was realized that these pipe were experiencing more breaks than expected. The number of breaks observed for these pipes peaked in the year 1980's. The success of retrofitting process is seen in the number of breaks

observed for the YDI pipe compared to breaks for the other metallic pipes. In all the cases of breaks for all the metallic pipes, the YDI pipe has a lower break rate of about 5times compared to the pipe with the most breaks.

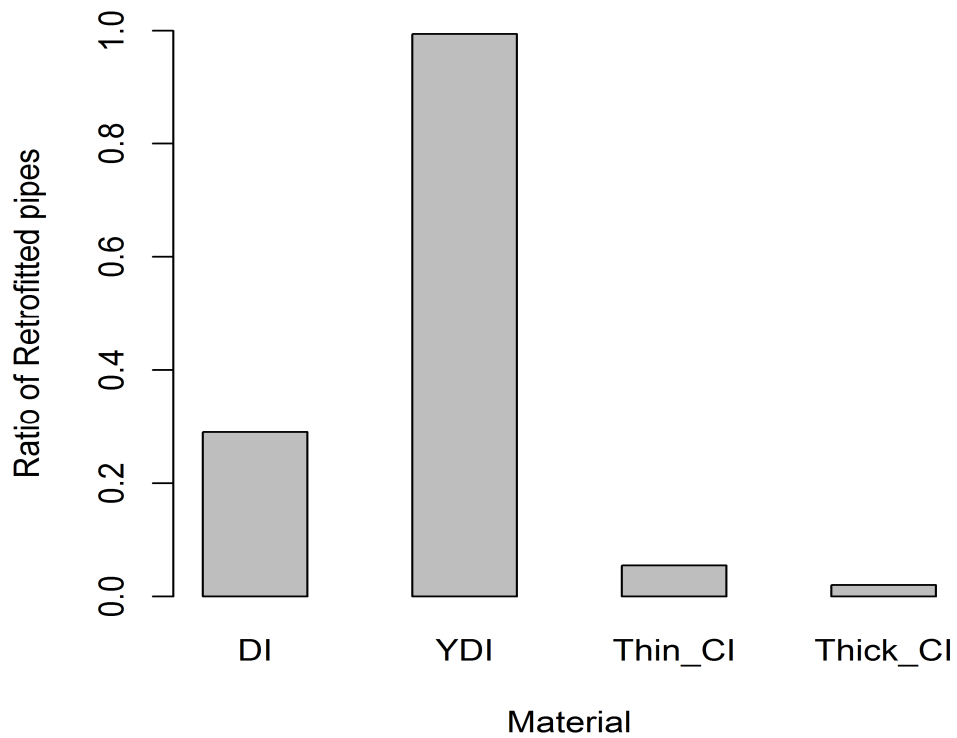


Figure 4-21: Relationship between retrofitted pipes and material type

Chapter 5

5 Discussion

This section presents results from the application of the three models to the CCWN utility. Results from model variable selection are then presented and finally, model training and validation is presented.

5.1 Model Variable Selection

Using random sampling method (Richardson et al. 1995), the data are split into training (70%) and test (30%) sets for use in model training and validation, respectively. The number of pipes considered in each data set is shown in Table 5-1.

Table 5-1: Number of pipes for training and testing of the models in each data set

Data set	CI		DI		PVC
	PRB=0	PRB>1	PRB=0	PRB>1	PRB>0
Training set	4412	1890	5018	2150	12,784
Test set	2921	1251	1802	771	5,478
Total	7333	3141	6820	2921	18,262

From each of the training sets, a total of seven covariates thought to have a significant contribution on pipe break rate were included in the variable selection. Results from selection process indicated that covariates with the most significant impact in influencing pipe breaks are length (z_0), diameter (z_1), MAT (z_2), soil resistivity, REST (z_3), FI (z_4), and RD (z_5). The MAT covariate represents different material types such as PVC pipes, thick-wall and thin-wall for the CI pipe, and unwrapped and wrapped DI and YDI pipes. The selection process is done using the methodology presented in Section 3.2.2.

5.2 Model Training

5.2.1 Effects of Covariates on Pipe Failure

Using the training data set, Cox-PHM, WPHM and Poisson Model are fitted with the variables found to be significant in Section 5.1. Covariates significant at 95% level were included in the model.

Table 5-2, Table 5-3, and Table 5-4 show significant covariates and the corresponding coefficients for each model used in the study for DI_A, CI and PVC pipe strata, respectively. The contribution of covariate to the hazard of a pipe is portrayed by the positivity or negativity of the coefficients (Cox 1972) with exponential values closer to one indicating marginal contribution to the hazard. More positive coefficients indicate higher risk of pipes failing and vice versa.

Table 5-2: Selected covariates and corresponding coefficients Cox-PHM and WPHM¹ for DI strata

	Cox-PHM		WPHM		Poisson	
	PRB=0	PRB>1	PRB=0	PRB>1	PRB=0	PRB>1
Intercept			5.61	4.41	-0.8	-0.89
Scale			1.82	0.91		
Coefficients, β						
Length	0.0007	0.001	-0.0007	-0.001	0.00054	0.0007
Diameter	0.187		-0.194		0.326	
MAT	-0.593		0.522		-0.62	
REST		-0.096		0.102		-0.12
FI	0.0005		-0.0005			
RD	0.021	-0.0007	-0.023	0.00076	0.018	-0.012

¹Dividing the WPHM coefficients by the scale parameter give Cox-PHM parameters. Negative sign in WPHM coefficients indicate a higher risk

Table 5-3: Selected covariates and corresponding coefficients for Cox-PHM and WPHM for CI strata

	Cox-PHM		WPHM		Poisson	
	PRB>1	PRB=0	PRB>1	PRB=0	PRB>1	PRB=0
Intercept			4.83	5.93	-1.1	-0.93
Scale			1	1.85		
Coefficients, β						
Length	0.0014	0.0013	-0.0013	-0.0013	0.0009	0.00079
Diameter	-0.155		0.160		-0.050	
MAT		0.101		-0.12		0.024
REST		-0.119		0.111		-0.466
FI	0.0003	0.0006	-0.0003	-0.0006	0.0001	0.0004
RD	0.0013	0.0178	-0.0015	-0.0186		-0.013

Table 5-4: Selected covariates and corresponding coefficients for Cox-PHM and WPHM for PVC strata

	Cox-PHM	WPHM	Poisson
	PRB>0	PRB>0	PRB>0
Intercept		6.45	-3.55
Scale		4.74	
Coefficients, β			
FI	0.008	-0.008	0.0086
RD	0.029	-0.027	0.028

The above tables indicate that pipe diameter, only influences the occurrence of a pipe break in metallic pipes after an occurrence of a break. Results signify that small diameter pipes (<200mm) have a high risk of breaking compared to large diameter pipes. For example, for DI_A pipes, the hazard rate is scaled by a factor of 1.2 for pipes (results from Cox-PHM and WPHM) while PM predicted an increase by a factor of 1.3. For CI pipes, large diameter (>200mm) pipe reduces the hazard rate by factor of 0.86 (results from Cox-PHM) and 0.95 (results from Poisson Model). These findings confirms past studies on susceptibility of small diameter pipes to breaks (Kettler and Goulter 1985).

Pipe material influences the occurrence of pipe failure differently. A higher hazard rate, as expected, for CI_TN pipes compared to CI_TK pipes was observed. Results from all models

revealed that CI_TK pipe material reduces the hazard rate by about 0.85 compared to the CI_TN. It is thought that the smaller wall thickness of the CI_TN pipe (20% less) compared to CI_TK pipe (Brander 2001) may be contributing to more breaks observed for this material type. In a similar study done by Brander and Ng (2000) reveals that CI_TK halves breaks observed in CI_TN. Therefore, the results from this study seems to confirm the findings by Brander and Ng (2000) and the result seem to conclude that the CI_TK is more superior pipe compared to the CI_TN.

YDI pipe has a high significant reduction in the hazard rates compared to wrapped and unwrapped DI pipe. It was found that the pipe contributed significantly to the reduction of the hazard to about 0.5 compared to the DI pipe, signifying a reduction of breaks by almost half breaks for this pipe material. This reduction can be attributed to protection of the pipe against corrosion using coating and anode protection (Brander and Ng 2000). It has been reported that up to 90% reduction of breaks has been observed from this pipe material (Brander 2001).

From the discussion of the pipe materials considered, it has been noted that pipe material plays a very crucial role in the occurrence of the number of breaks for a pipe. This is due to the fact that different pipe materials responds to the effects of various factors affecting the pipe differently. Though the importance of material type has been demonstrated, care should be adopted while concluding these results as there may be other factors influencing the occurrence of a pipe break.

The positive coefficients of FI from above tables indicate that as the temperature decreases, more breaks are expected. However, the three models predicted a marginal increase of the hazard (increase by a factor of 1) for DI_A, CI and PCV pipes. It is thought the 3m installation depths used in the city (Brander 2001) has masked the influence of the temperature by limiting the penetration of frost loads. Similar results have been obtained by Kleiner and Rajani (2002) and Kleiner and Rajani (2010). Studies have shown that frost load penetration can only penetrate to waterlines at depths less than 1.52m and 2.23m in clay trenches with granular backfill and rocky trenches with clay or granular backfill respectively (Zhao et al. 2001).

Table 5-2 and Table 5-3 show that pipes installed in high soil resistivity, resistivity > 2000ohm-cm, are less likely to fail than those installed in soils with low resistivity. Studies have attributed the occurrence of corrosion to low resistivity soils (Makar 2000; Makar et al. 2001) by

increasing the pitting rate (Liu et al. 2010). Table 5-2 reveal that once the DI_A pipe has experienced a break; high corrosive soils begin to influence the failure of pipes more while pipes in low corrosive soils experiences less breaks. The results from Table 5-2 indicate both Cox-PHM and WPHM predicted a reduction of the hazard by factor of 0.9 while PM predicts a reduction by a factor 0.88. Prediction results from Cox-PHM, WPHM and Poisson (Table 5-3) for the CI pipes indicate the reduction by 0.88 for both Cox-PHM and WPHM with low rates of 0.62 being predicted by PM. Therefore, this result seem to confirm the observation by Brander (2001) that most breaks in CCWN occur in areas with corrosive soils than in non-corrosive soils .

In the past studies, some authors (e.g. Le Gat and Eisenbeis 2000; Røstum 2000) have found a strong linear correlation between break and length. However in this study, pipe length marginally (increases the hazard is by a factor of 1- results from the three models) increases the rate of breakage (Table 5-2, Table 5-3 , and Table 5-4) with longer pipe breaking more than shorter pipes. The effects of this covariate seem to impact only DI_A and CI pipes, either with one or more breaks.

From these tables, results indicate that high RD compared to low RD increases the probability of metallic and non-metallic pipes breaking through increased shrinkage of expansive soil (Fuchs-Hanusch et al. 2011; Gould et al. 2009). However, the hazard rate from all models increases marginally by a factor of 1.02. The heterogeneity of soils in CCWN (Brander and Ng 2000; Brander 2001) seems to lessen the impact of this covariate to areas with expansive soils (clay).

The covariate values obtained from the models in this study differed from those estimated in previous studies (e.g. Le Gat and Eisenbeis 2000; Røstum 2000) except for length and diameter as shown in Table 5-5 . These two covariates were comparable the those estimated by Fuchs-Hanusch et al. (2012).

Table 5-5: Parameter values for different covariates from past studies

Variables	Parameter values	Stratification	Reference
z_0	0.002	PRB=1	Fuchs-Hanusch et al. (2012)
	-0.51	PRB=0	Røstum (2000)
	-0.768	DI	Martins (2011)
	-0.29	PVC	Le Gat and Eisenbeis (2000)
	-0.37	CI	Le Gat and Eisenbeis (2000)
z_1	0.001	DI	Martins (2011)
	-1.62	PRB=1	Fuchs-Hanusch et al. (2012)
	0.004	PRB=0	Røstum (2000)
	0.55	DI	Le Gat and Eisenbeis (2000)
	0.52	CI	Le Gat and Eisenbeis (2000)
z_2	0.042	PRB=1	Fuchs-Hanusch et al. (2012)
	0.082	DI	Martins (2011)

As seen from table above, covariates affect pipes differently in the every water network; therefore it is doubtful to have a model that is suitable to all networks.

From the above discussion on the effects of covariates on the pipe failure, it is clear that the covariates play an important part. These factors are thought to combine together in affecting the pipe failure. It was however noted that the most significant factors influencing pipe failure in the CCWN are the physical factors compared to environmental factors.

5.2.2 Model Break Prediction

To test the prediction accuracy of the built models, observed breaks from the training sets for the DI_A, CI and PVC pipe strata were predicted using Cox-PHM, WPHM and PM. Figure 5-1 shows prediction results from the training set for the CI pipe stratum. As can be seen from the figure all the models produced close results especially the WPHM and the PM, though the results from WPHM seems closer to the observed breaks than the results from PM and Cox-PHM.

This figure shows that WPHM best predicted the observed increasing break rate and decreasing break rate for CI pipes, but could not fully capture the rapid break rate for the CI pipes. The increasing break rate of CI is best predicted by the PM, but could not capture the slow breaking phase of pipe breaks. The result from Cox-PHM indicate the case of underestimation for all the

phases of pipe breaks, but somehow the model seemed to capture the constant break rate for the pipes. The failure of the Cox-PHM model to capture the failure trend of CI pipe is mainly attributed to the undefined baseline hazard. Overall, WPHM captured the breaking trend of CCWN utility better than both Poisson Model and Cox-PHM.

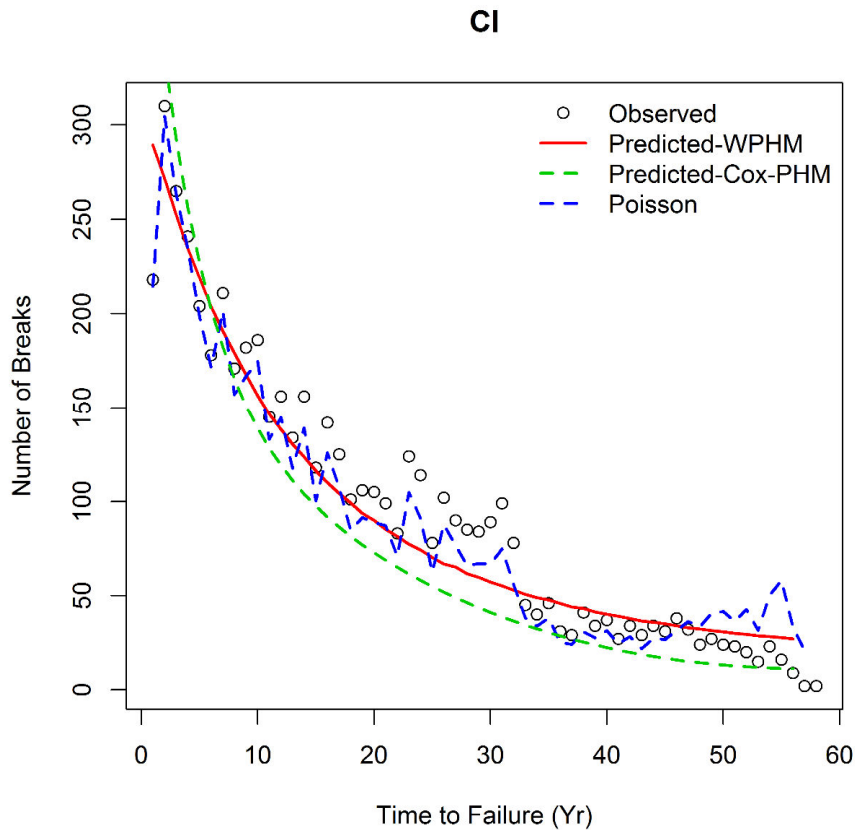


Figure 5-1: Prediction of Breaks for CI Strata from CI Data Training Set

Figure 5-2 presents the modelling results for the DI_A pipe from the WPHM, Cox-PHM and PM. As can be seen in this figure, varied results were predicted by the models with closer prediction results given by WPHM.

The WPHM captured the increasing breaking rate, the decreasing rate and somehow captured the rapid breaking phase of the pipes. On the other hand, the PM seem to underestimate all the braking phases of the pipe break but notably, it overestimated the slow breaking phases of the pipe. In this phase, it significantly predicted very high number of breaks for the DI_A pipe. Notable from Cox-PHM, the model somehow predicted well the increasing trend and the

decreasing trend of DI_A breaks while significantly underestimating the rapid breaking phase of the pipe.

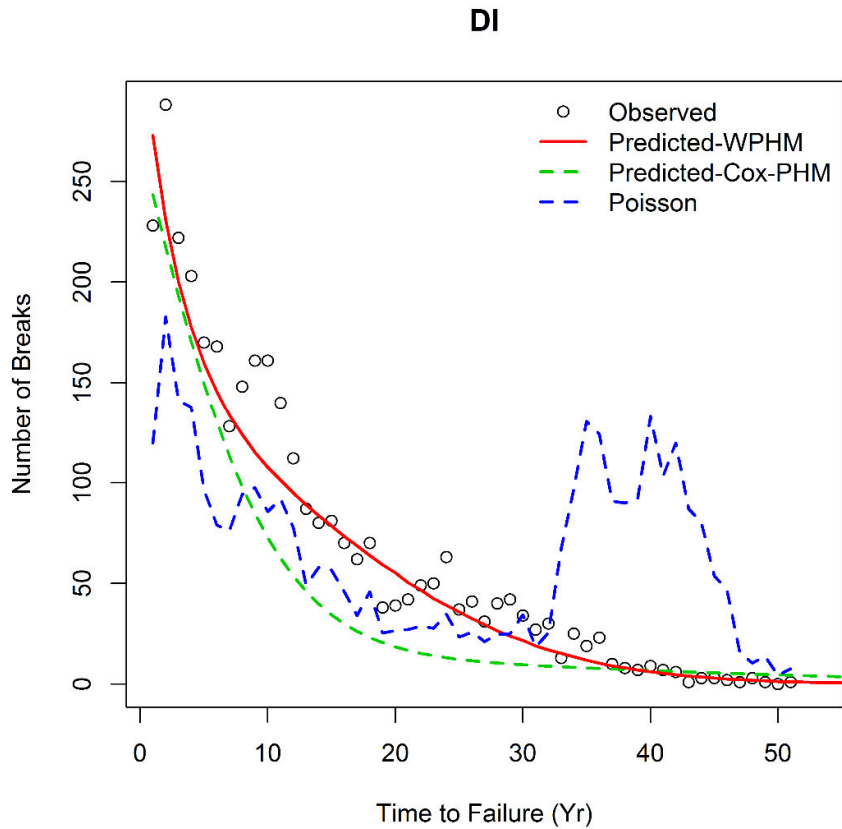


Figure 5-2: Prediction of Breaks for DI Strata from DI Data Training Set

Similarities in the breaking trends of the models applied in the prediction of the metallic pipes were observed. It can be seen that for both prediction of the breaks from both pipes, WPHM results indicated it is a better model in the prediction of breaks from these pipes.

Literature has indicated that the use of WPHM for prediction of failures elicits varied prediction results. Some authors reporting a case of underestimation (e.g. Le Gat and Eisenbeis 2000), others reporting overestimation (e.g. Røstum 2000) while others have reported accurate prediction results (e.g. Martins et al. 2013). These varied results are attributed to model dependence on approximate procedures like Monte Carlo simulation in calculation of the number of failures (Le Gat and Eisenbeis 2000). However, Røstum (2000) resolved that the nature of prediction results obtained from this model is unique to a particular water utility.

In both CI and DI_A pipe break predictions, Cox-PHM underestimated the fast breaking phases but somehow well predicted the slow breaking phase. These results seem to question the validity of the Cox-PHM model in the prediction of the fast breaking phase. Similar findings were reported by Andreou et al. (1987) and Karaa and Marks (1990). These authors recommended the use of Cox-PHM in the prediction of the slow breaking phase and the results obtained in this study corroborates this recommendation. The weakness of the Cox-PHM in not being able to fully capture the fast breaking phases of the pipes is mainly attributed to the undefined baseline hazard function.

Similarly, the PM overestimated the prediction slow breaking phases of CI and DI_A strata. It is assumed that the reason for the failure of Poisson model to capture this phase is due to the assumption of constant failure rate for all the pipes in strata. Therefore, it has no way of differentiating between slow breaking pipes and those that have a high likelihood of failure. Similar case was reported by C. Martins (2011).

Figure 5-3 presents the prediction results of pipe breaks from the PVC pipe strata using the training data set. As can be seen in the figure, varied prediction results are given by the three models applied to this data set. It is clear from the figure that the PM produced results that are closer to the observed pipe breaks data from the strata. However, the number of breaks predicted from both Cox-PHM and WPHM departed considerably from the observed data. The WPHM model mimicked the prediction results for the first 8years then started departing from the observed data, however, the Cox-PHM overestimated the results from the onset. However, from the 10year, the WPHM overtook the Cox-PHM in overestimating prediction of pipe breaks.

The result from the survival models seem to point to the failure of these models in capturing the failure trend of the pipes. However, the output from the PM seem to point that count models and constant hazard models seem to be more suitable to the breaking pattern for these pipes.

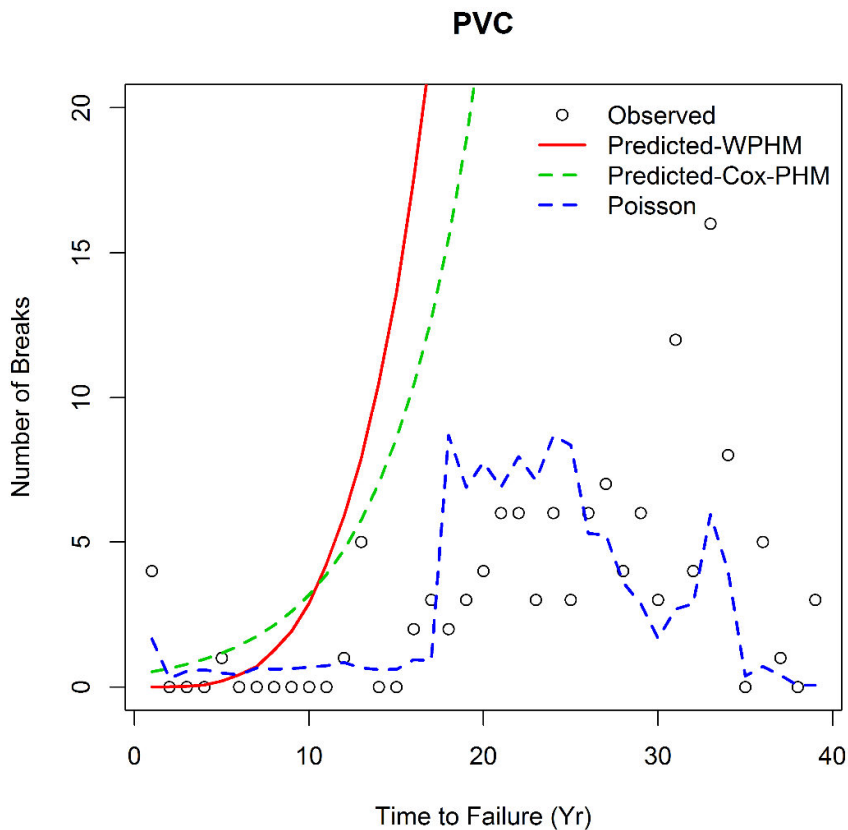


Figure 5-3: Prediction of Breaks for PVC Strata from PVC Data Training Set

The incapability of Cox-PHM and WPHM to capture the failing trend of pipes in this stratum is partly attributed to less number of observed failures (only 168 recorded breaks in over 18,000 individual pipes). However, the author reasons that the major reason is because these pipes have a constant failure rate while Cox-PHM and WPHM are time-dependent models. Literature have reported similar findings in UK and Australia and in one incident, reduction in number of failures with increase in time was observed for PVC pipes (Davis et al. 2004).

From the prediction results, it is therefore fairly reasonable to assume that the break trend of the PVC pipe completely differs from that of metallic pipes as the number of breaks does not increase as the pipe ages. Furthermore, from the analysis given in Section 4.3.3 for the break rate of the PVC pipe, the risk factors do not seem to accelerate its failure rate.

5.2.3 Checking for Model Appropriateness for Cox-PHM and WPHM

It is important that proportional hazard models are tested before their application in the break prediction in order to determine their suitability.

5.2.3.1 Cox-PHM

The methodology presented in Section 3.1.1.3.2 was employed in the determination of the validity of the Cox-PHM model in break prediction. Figure 5-4, Figure 5-5, and Appendix A-2 presents the Schoenfeld's residuals for the CI pipe strata with no previous breaks. The results indicate that the proportionality assumption of the model is not violated and therefore the models are fit for breaks prediction.

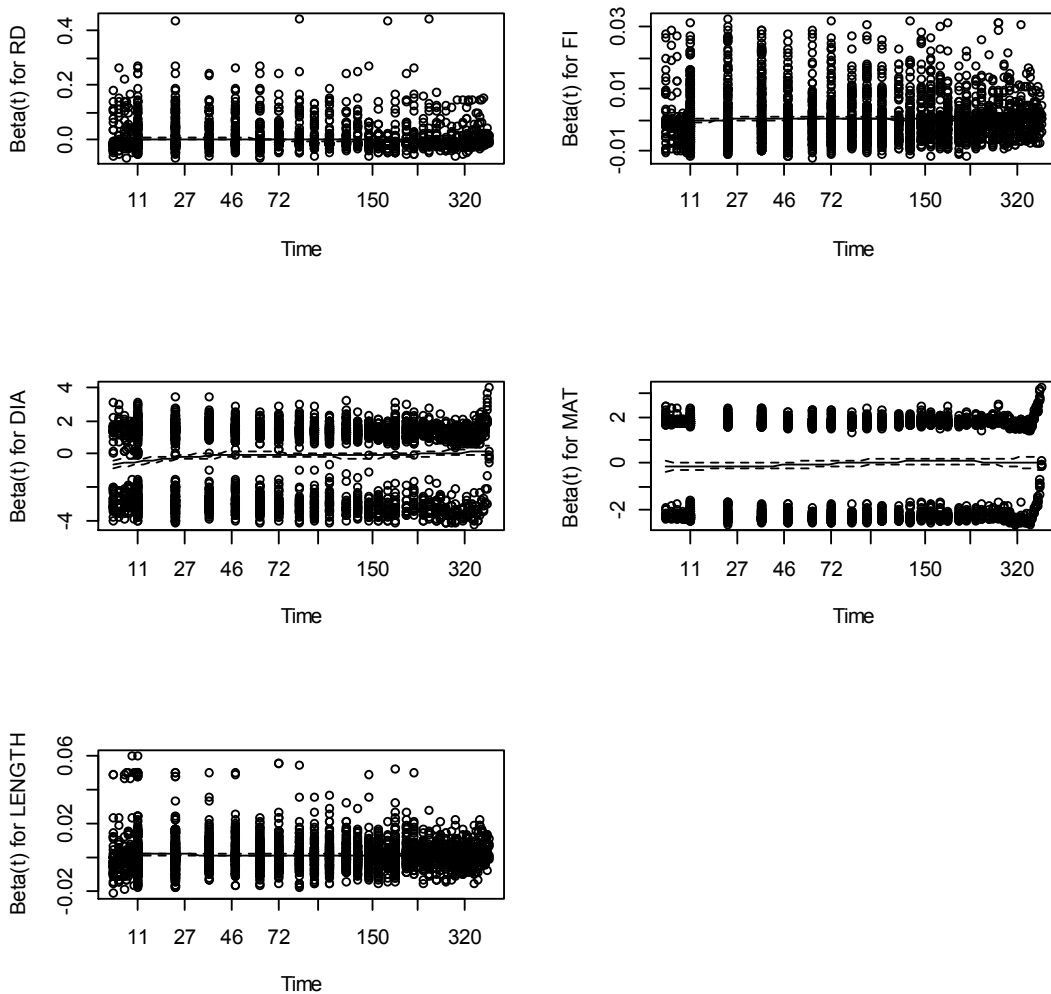


Figure 5-4: Schoenfield's residuals for the CI pipe strata with no previous breaks

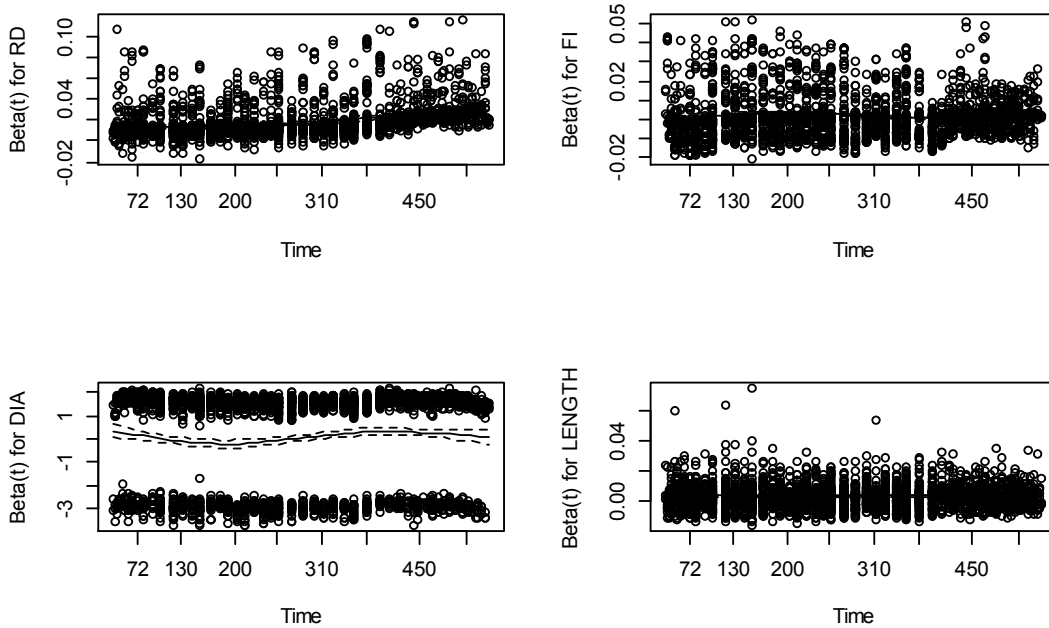


Figure 5-5: Schoenfield's residuals for the CI pipe strata with more than one break

5.2.3.2 WPHM

To test for the proportionality of the WPHM model, a plot of the model covariates is done. The plot of twice the minus logarithm of the survival function for each individual covariate in the model is done and the result should be approximately parallel if the model is appropriate. The result from Figure 5-6 and Figure 5-7 shows that WPHM model is appropriate for modelling the pipe breaks for CWW.

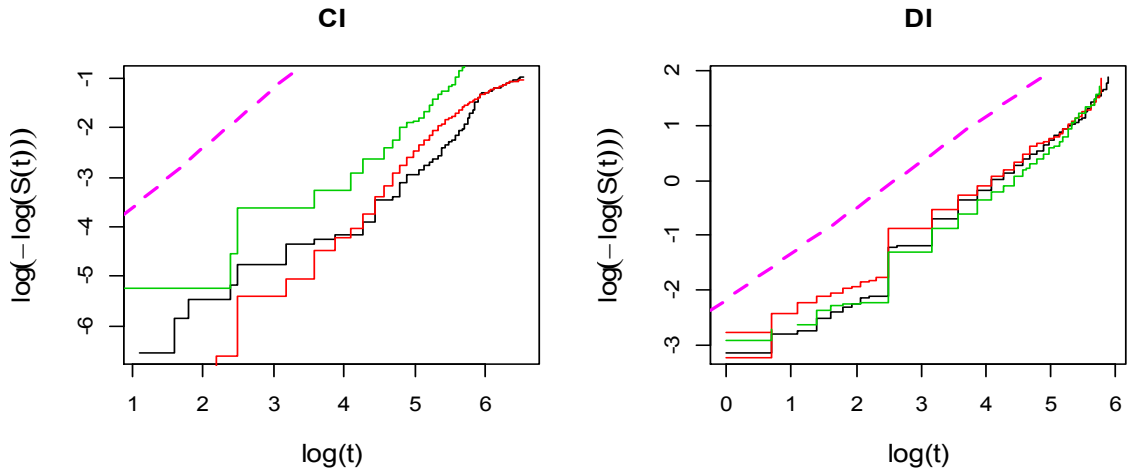


Figure 5-6: WPHM appropriateness using graphical method for pipes with more than one break

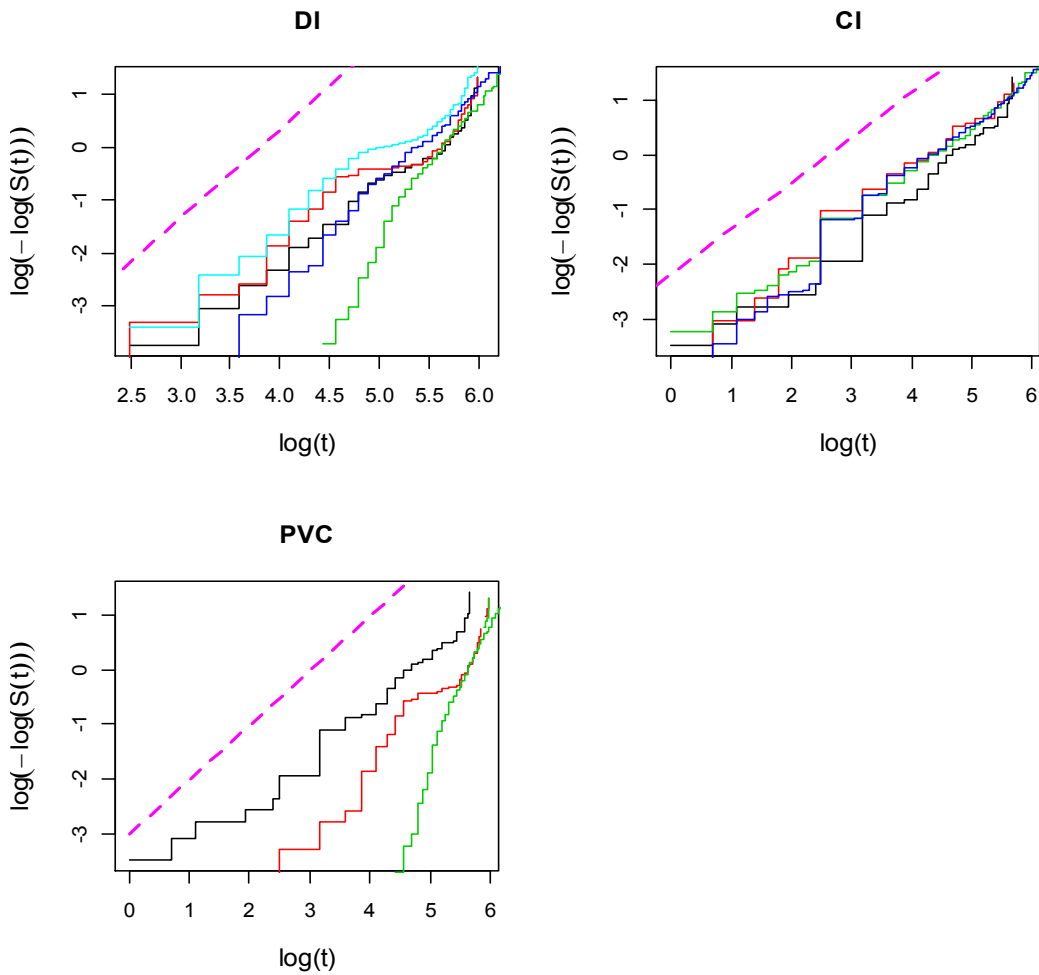


Figure 5-7: WPHM appropriateness using graphical method for pipes with no previous breaks

5.3 Baseline Hazard Estimation

By applying the methodology described in Section 3.1.1.2, baseline hazards for Cox-PHM are approximated by applying curve smoothing techniques such as linear and non-linear least squares (*nls*) relations (Rossiter 2009).

The results of curve fitting techniques described above for ductile iron (DI_A) and cast iron (CI) and PVC pipe strata are presented in Figure 5-8 and Figure 5-9. Figure 5-8 shows the fitted curve for the pipe strata with no previous break while the Figure 5-9 presents the fitted curves for pipe strata with more than one previous break.

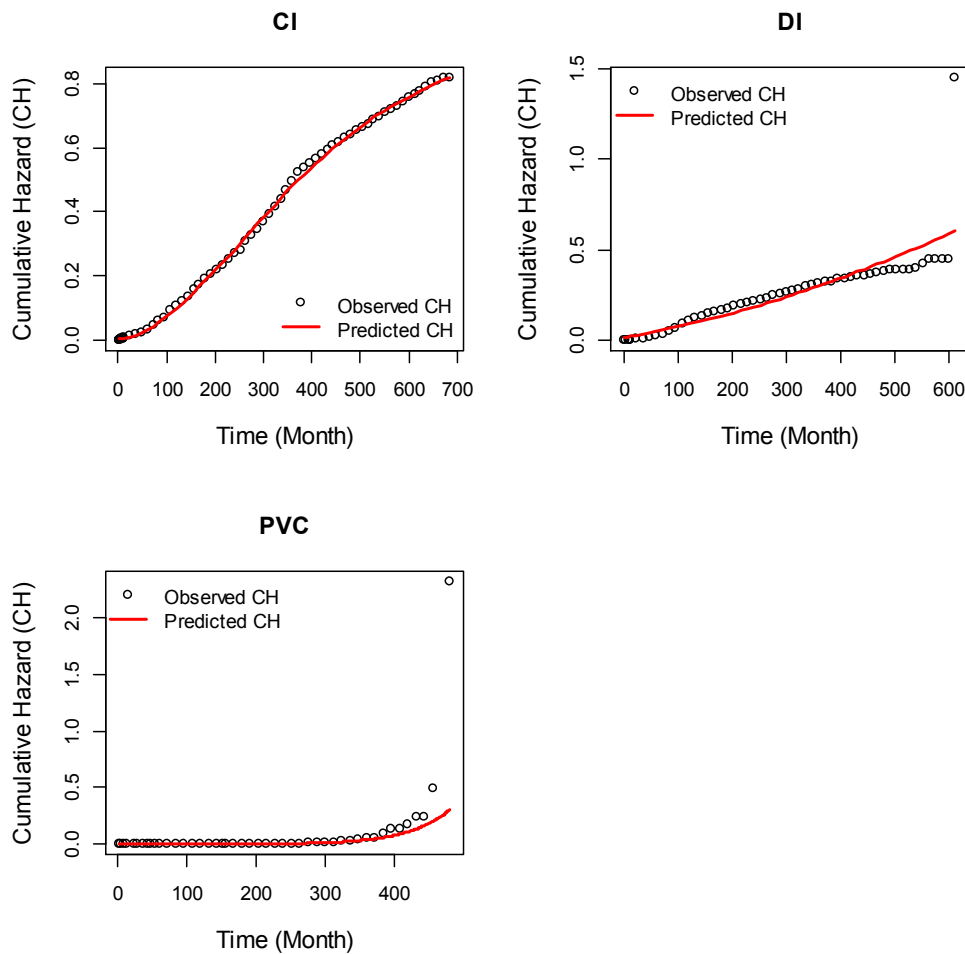


Figure 5-8: Observed and predicted CH for pipes with PRB =0 for CI, DI_A and PVC pipe strata

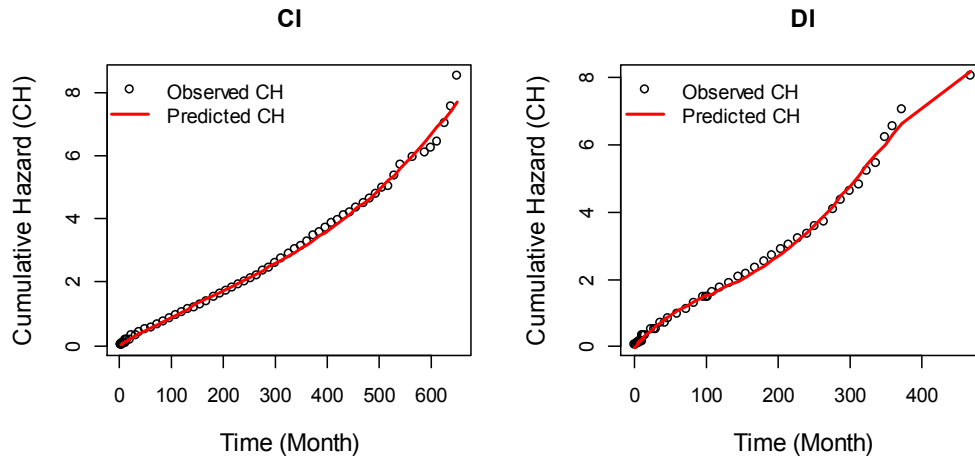


Figure 5-9: Observed and predicted CH for pipes with PRB >0 for CI and DI_A pipe strata

The hazards for the CI and DI_A strata were estimated using the linear relations where the hazards were fitted polynomial equations. Analysis of variance (ANOVA) was performed on any competing equations and it emerged that second, third and fourth order polynomial equations best described the hazards of these strata respectively. The cumulative hazard equation for the PVC strata was estimated using the *nls* method. The resulting cumulative hazard equations for each of the predicted hazards are given in Table 5-5

Table 5-6: Estimated cumulative hazards equations for DI, CI and PVC pipe strata from Cox-PHM

Strata by material	Strata by PRB	Estimated Equation
DI_A	PRB=0	$10^{-4} (-560 + 262t - 1.64t^2 + 0.0066t^3 - 0.000074t^4)$
	PRB>1	$-0.017 + 0.00053t - 0.00000071t^2$
CI	PRB=0	$0.0026 + 0.00975t - 0.0000095t^2 + 0.0000002t^2$
	PRB>1	$10^{-5} (456 + 13t + 0.68t^2 - 0.0012t^3 + 0.00000065t^4)$
PVC	PRB>0	$0.0001 \exp(0.00172t)$

5.4 Estimation of Time to Failure of a Pipe

Mean Time to Failure (MTF) represent the average time it take a pipe to experience a break. In order to establish time of replacement of a pipe, MTF of each pipe depending on the ordered occurrences of breaks in a pipe. As a result, seven ordered pipe breaks for both DI and CI pipes

were generated from the failure data, and Table 5-7 presents the number of pipes in each stratum.

As can be seen in the table, the number of pipes reduces as the number of pipes increases. Most of the breaks observed for a CI and DI_A pipes with one break is about 3 and 6 times, respectively, more than the breaks experienced by pipe with 2 breaks.

Table 5-7: Number of pipes in each of the defined strata in pipe data set

Type of pipe	Strata						
	PRB1	PRB2	PRB3	PRB4	PRB5	PRB6	PRB7
CI	5312	1768	1034	572	312	174	290
DI	7168	1042	620	368	220	135	188

For each stratum, a model was developed by considering all the covariates in the pipe failure data. Significant covariates that influence the occurrence of a pipe break reduces as pipe changes from slow to fast breaking phases. Table 5-8 and Table 5-9 present significant covariates as well as the corresponding coefficients that depict the influence of these covariates on each pipe stratum failure.

Table 5-8: Significant covariates and corresponding coefficients for CI pipe strata

Covariates	PRB 1	PRB 2	PRB 3	PRB 4	PRB 5	PRB 6	PRB 7
Diameter	1.26	0.90	0.84	0.85		0.61	0.77
Length	1.00	1.00	1.00	1.00	1.00		1.00
MAT	1.53						
RD	1.01	1.00					
FI	1.00						

Table 5-9: Significant covariates and corresponding coefficients for DI pipe strata

Covariates	PRB 1	PRB 2	PRB 3	PRB 4	PRB 5	PRB 6	PRB 7
Diameter	1.71			0.98			
Length	1.00	1.00	1.00	1.00		1.00	
MAT	1.75				0.68		1.44
RD	1.00						
FI	1.00	1.00	1.00				

The influence of these covariates in the failure of the pipe is similar to the explanation given in Section 5.2.1.

5.4.1 Survival Curves

Survival curves indicate the probability of pipes that have not broken at time t and the higher the curve, the longer it takes for the first break to occur. It should be noted that the curves were developed from only Cox-PHM.

First, survival curves for DI_A and CI pipes were developed to give an indication of the failure trend of the two pipe materials under study. Figure 5-11 presents the survival curves for CI, both thick and thin, and DI_A (wrapped and unwrapped DI and YDI) pipes. From Figure 5-11, the survival curves of both pipes begin to fall immediately after installation signifying susceptibility of these pipes to deterioration mechanisms. However from this figure, there exist significant differences in failure rates of the DI_A and CI pipes.

Like in the case of CI pipes, there seem to be no significant difference in the failure rate for both pipes for approximately 90months after installation. Thereafter, difference between the failure rates of the two pipes becomes apparent. As expected, CI_TK pipe seem to break slowly than its counterpart. In addition to differences in wall thickness, past studies have revealed that difference in metallic composition of these pipes contributes to these differences (Brander and Ng 2000; Brander 2001).

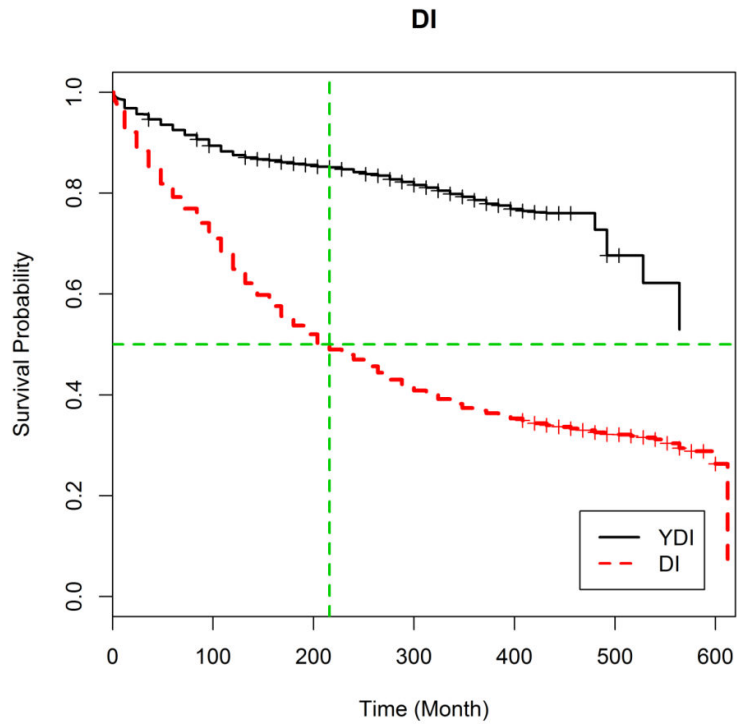


Figure 5-10: Survival curves for YDI and DI pipe

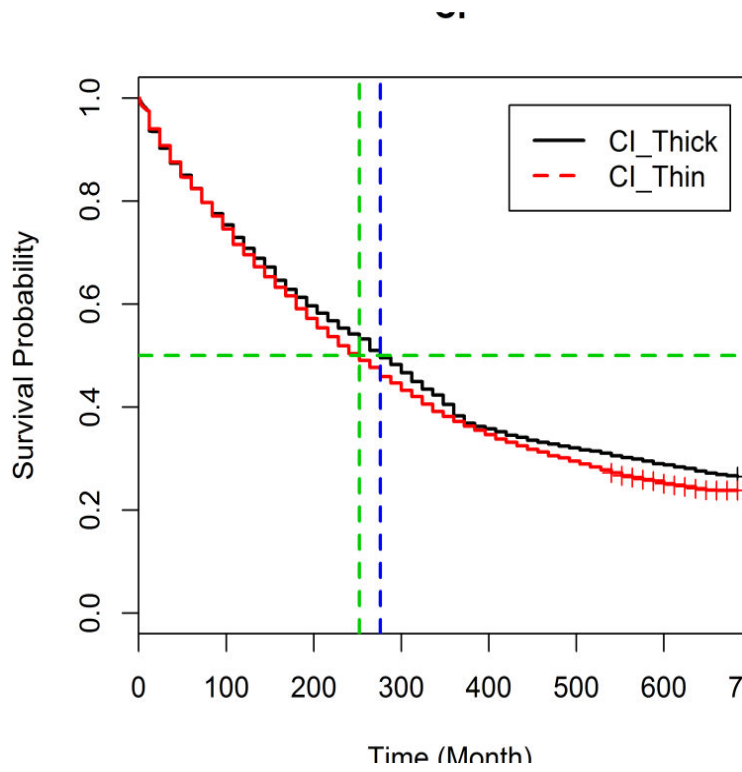


Figure 5-11: Survival curves for thick and thin CI pipe

On the other hand, survival curves for DI_A pipe depicts a different pattern for the DI (wrapped and unwrapped) and YDI pipes. The figure shows a steeper curve for DI with YDI depicting a gentle failing trend. This result indicates a lower break rate for the YDI pipes compared to DI pipe. A study undertaken by Brander (2001) revealed that the YDI pipes is protected against corrosion both with coatings and anode protection. Further, the study revealed that this pipe has led to up to 90% reduction of pipe breaks. When pipe breaks were ordered according to the number of break occurrences and their survival curves drawn, varied results were observed. As shown in Figure 5-12, the steepness of the survival curves for a pipe with two and more breaks increases significantly indicating that the likelihood of a pipe breaking after experiencing a break rises. This finding collaborates the outcome in past studies e.g. Fuchs-Hanusch et al. (2012) and Pelletier et al. (2003). The later concluded that once repair and maintenance has occurred, chances of pipe break occurring increases.

Further, as can be seen in the figure below, there exist significant difference in steepness of survival curves for a pipe with two to four breaks (CI) and two to three (DI) but after this, survival curves seem to almost form one curve. This result seem to signify that once the pipe has experienced more than three (DI) or four (CI) breaks, the hazard remains constant (Fuchs-Hanusch et al. 2012; Park et al. 2008). Moreover, the pipe seem to have moved from the slow breaking phase to the fast breaking phase where the arrival of pipe breaks follows Poisson distribution (Andreou et al. 1987b).

Additionally, owing to the steepness of the curve for DI compared to CI pipe, it is clear that after the first failure, DI pipe becomes more disposed to failure than CI pipe. Studies have revealed that DI pipe can be more prone to failure as it can be more prone to corrosion attack (Makar et al. 2001; Rajani and Kleiner 2003).

CI

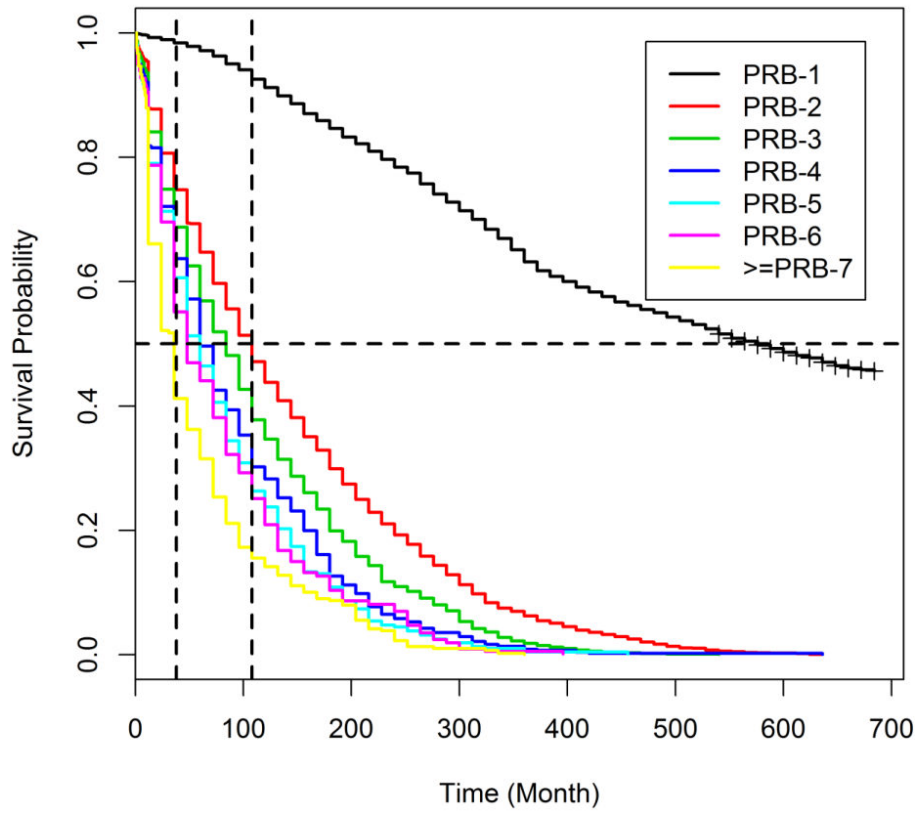


Figure 5-12: Survival curves for CI pipe with different ordered breaks

DI

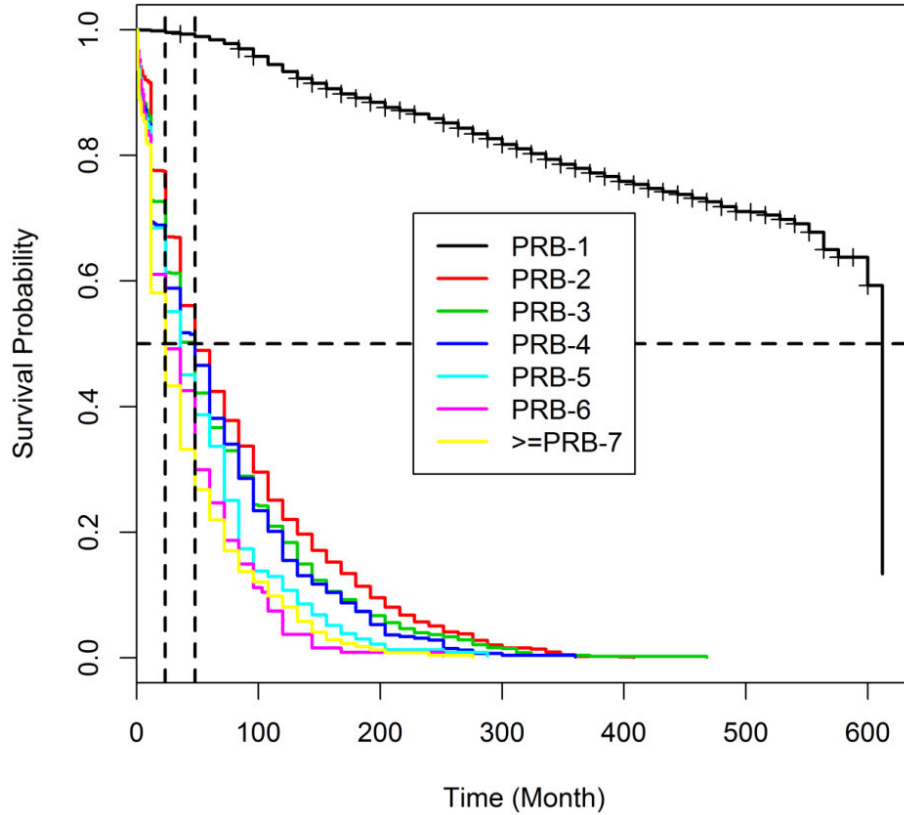


Figure 5-13: Survival curves for DI pipe with different ordered breaks

From these results, it can be concluded that after the occurrence of three or four breaks, the pipe becomes more prone to failure, therefore, its replacement may be more reasonable to undertake than a repair. This is so because after the third or fourth break, the pipe moves from the slow breaking phase and therefore the cumulative cost of repair might become higher in the long run.

5.4.2 Mean Time to Failure (MTF)

In order to estimate the average time it takes for a pipe to fail, failure data were stratified into seven ordered groups. For each stratum a model was developed and its MTF calculated. The results from these models are presented in Table 5-10. The seventh group in each pipe category represents pipes with seven and more breaks lumped together.

Table 5-10: Mean Time to Failure for DI and CI pipes ranked on previous failures

Strata	DI	CI
PRB=1	612	576
PBR=2	48	108
PRB=3	48	84
PRB=4	48	60
PRB=5	36	60
PRB=6	24	48
PRB=7	24	36

It is clear from Table 5-10 that the MTTF for pipes with no previous breaks is greater than for the pipes with more than one break pointing that after the occurrence of the first break, the likelihood of recurrence is high. As can be seen from the table above, the time to second break for DI and CI pipe reduces significantly to 48months and 108months, respectively. Parallel results have been found by Park et al. (2008).

Remarkably, the MTTF of DI pipe is higher than the MTTF of the CI pipe up to the occurrence of the first break. However, after the occurrence of the first failure, time to failure of CI supersedes the MTTF of DI pipe. Interestingly, past studies reveals that CI pipe is a brittle material, and therefore, the adoption of DI pipe was thought to reduce the number of break observed in water mains (Makar and Kleiner 2000; Makar et al. 2001; Rajani and Makar 2001). However, Rajani and Kleiner (2003) found that DI pipe could equal the number of breaks observed from the CI pipe. Therefore, the findings from this study confirm the findings from this study by Rajani and Kleiner (2003).

5.5 Sensitivity Analysis for different Break Types

To study the impact of covariates on the break-type, pipe data from CI and DI_A strata were stratified according to four break-types; ALL (represent corrosion, circular, leak, crack, and split break-types), circular (CIRC), corrosion (CORR) and leak (LEAK) break-types. Analysis of CIRC break-type for DI_A was not performed owing to less number of pipes that broke due to this break-type. Analysis for PVC stratum was not done owing to less number of breaks. Selection of variables were done using Akaike (1973) methodology and models for each break type developed and parameterized using Equations 3-4, 3-11, and 3-14, for Cox-PHM, WPHM

and Poisson Model respectively. The Significant variables and corresponding coefficients for CI and DI pipes are shown in Table 5-11, Table 5-12, and Table 5-13.

Table 5-11: Sensitivity analysis for ALL break-types and corrosion break-type on CI pipe

	ALL breaks			Corrosion		
	WPHM	Cox-PHM	Poisson	WPHM	Cox-PHM	Poisson
Intercept	7.36		-0.903	8.74		-1.85
Scale	0.89			0.83		
Coefficients, β						
LENGTH, β_0	-0.004	0.0041	0.0017	-0.006	0.0056	0.0035
DIA, β_1	-0.107	0.101	0.122			
PMP, β_2	-0.48	0.46	0.204	-0.43	0.41	0.3
FI, β_3	-0.0021	0.002	0.0009	-0.004	0.0041	0.0026
RD, β_4			0.0041	-0.014	0.0134	0.095

Table 5-12: Sensitivity analysis for Circular and Leak break-type on CI pipe

	Circular			Leak		
	WPHM	Cox-PHM	Poisson	WPHM	Cox-PHM	Poisson
Intercept	8.87		-1.98	8.09		-1.25
Scale	0.87			0.79		
Coefficients, β						
LENGTH, β_0	-0.0048	0.0047	0.0032	-0.0041	0.0041	0.0022
DIA, β_1	-0.516	0.507	0.36	0.37	-0.36	-0.184
PMP, β_2	-1.02	0.985	0.68	-0.114	0.113	-0.044*
FI, β_3	-0.0028	0.0025	0.0013	-0.0034	0.0032	0.0021
RD, β_4	-0.0132	0.0127	0.0062	-0.0166	0.0156	0.0104

Table 5-13: Sensitivity analysis for ALL and Corrosion break-type for DI pipe

	All break types			Leak		Corrosion			
	WPHM	Cox- PHM	Poisso n	Cox- PHM	Poisso n	WPHM	WPHM	Cox- PHM	Poisso n
Intercept	7.96		-1.6		-1.9	8.48	8.48		-2.05
Scale	0.78					0.68	0.76		
Coefficients, β									
LENGTH, β_0	-0.0017	0.00066	0.0006						
DIA, β_1	-0.46	0.43	0.32	0.43	0.39	-0.43	-0.41	0.41	0.33
PMP, β_2	-0.8	0.8	0.64	0.69	0.57	-0.63	-1.09	1.12	1.0
FI, β_3	-0.0041	0.004	0.0023	0.0055	0.0038	-0.0057	-0.005	0.0048	0.003
RD, β_4				0.019	0.015	-0.02	-0.015	0.015	0.01

From above tables, significant covariates do not vary greatly depending on pipe material and break-type. Results indicate that both physical and environmental covariates are important in the occurrence of a particular break-type in a pipe. The covariates found to significantly increase the hazards in all break-types (ALL, CORR, LEAK CIRC) in both material types were diameter and PMP (the covariate Pipe Manufacturing Period (PMP) represents the age of the pipe).

As expected, old pipe in the system have a high risk of failing compared new pipes. Table 5-11 and Table 5-12, reveals that age of the pipe contributes to increase in the occurrence of circular break more than occurrence of any other break for CI pipes. Cox-PHM and WPHM predicted that old pipes increase this occurrence by a factor of 2.7 while Poisson predicts a hazard increase of a factor of 2 as compared to new pipes. Therefore, it is apparent that as CI pipe ages; it becomes more susceptible to circular breaks. This finding further confirms the finding by Makar et al. (2001) and in Figure 4-17 that CI pipe fails more as a result of circular breaks. For DI_A pipes (see Table 5-13), this covariate increases the hazard for the old pipes for the occurrence of corrosion break by a factor of 2.7 and 3 from Poisson Model and both WPHM and Cox-PHM models, respectively, more than occurrence of any other break.

Except for leak break-type (Table 5-13), smaller diameter pipes increases the hazard of occurrence of a particular break-type as compared to large diameter pipes. For CI pipe, the

most impacted break-type is the circular break-type. Poisson Model and both Cox-PHM and WPHM predicts increase by a factor of 1.43 and 1.66, respectively, as compared to larger diameters. Further,

Table 5-12 indicates that most of the pipe failing as a result of leak from joints and fittings are the large diameter pipes. It seems that as these pipes age, there joints and fittings loosen thereby increasing the chances of leak occurrence. From Table 5-13, smaller diameter pipes increases the occurrence of all break-types (ALL, CORR, LEAK) for DI pipes almost equally. Both Cox-PHM and WPHM predicts increase of the hazards by a factor of 1.5, while Poisson predicts an increase by a factor of 1.4. From Table 5-11, analysis of the CORR break-type for CI pipe indicate the occurrence of corrosion break-type is not influenced by the diameter as it was found not to be significant.

FI was found to be very significant in both pipe materials but increased the rate of failure marginally for all the pipe materials. All the models predicted an increase of the hazard by a factor of 1. For both pipe materials, RD covariate was significant when CORR, LEAK and CIRC break-types were considered separately but was not significant when all break-types were pooled. Just as for the FI covariate, the influence of this covariate on the hazard rate was found to be low with all the models predicting an increase the hazards by a factor of 1. The occurrence of either leak or corrosion break-type for DI_A pipes (Table 5-13) is not influenced by the length of the pipe. This covariate was found to contribute marginally to the occurrence of CI break-types and ALL break-types for DI_A pipes.

From this analyses, both environmental and physical pipe attributes influences the occurrence of a particular break-type occurring in a pipe. However, the contribution of these variables differs with physical pipe attributes influencing the occurrence of particular break-type more than environmental variables.

5.6 Model Verification

To test the accuracy of the models, the prediction of the number of breaks in the test data set was carried out using 30% of the data in each stratum. Model parameters used in verification were estimated from the training data set and are contained in Table 5-2, Table 5-3, and Table 5-4. The results from each pipe stratum are discussed in the following paragraphs.

The observed and predicted breaks from Cox-PHM, WPHM and Poisson Model for DI pipes are presented in Figure 5-14. The diagonal line in this figure indicates the ideal line (where the predicted breaks equals the observed breaks) which signifies the best fit of data between the observed and the predicted breaks.

From this figure, it is clear that most of the predicted and the observed breaks from the WPHM are concentrated on the ideal line. This result signifies that the model seem to predict well the observed breaks from the DI_A pipe strata. For the PM, in the early stages of the pipe life, most of the predicted breaks are below the ideal line indicating that the model underestimated the prediction of breaks for the DI_A pipe. However, in the fast breaking phase of the pipe, most of the breaks are concentrated on the ideal line indicating that the model predicted well the breaks in this phase. Therefore, prediction accuracies of the PM varies as the pipe moves from one phase to another with high accuracies obtained as the pipe moves from the slow to fast breaking phase.

Further from Figure 5-14, for Cox-PHM it can be seen that most of the predicted breaks in the slow and fast breaking phases are above the ideal line. This result indicate that the model significantly underestimated the breaks of these phases. However, for the last periods of the pipe life, the model seemed to accurately predict the breaks as can be seen by the concentration of breaks on the ideal line. Similar to results from training section, the accuracy of prediction from Cox-PHM seem to decrease as the pipe moves from slow to fast breaking phases of pipe life.

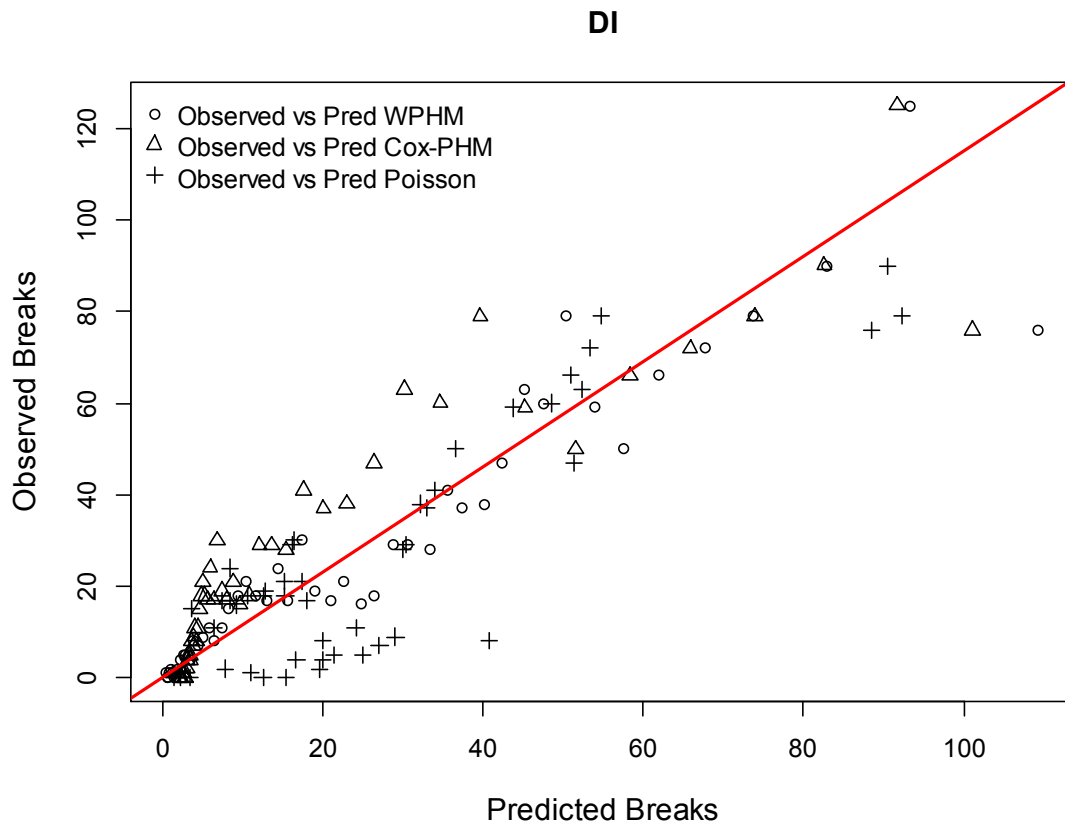


Figure 5-14: Observed and Predicted number of breaks for DI pipe strata

The observed and predicted breaks from Cox-PHM, WPHM and Poisson Model for CI pipes are presented in Figure 5-15. Similar to the results from the DI_A pipe, the diagonal line in this figure indicates the ideal line (where the predicted breaks equals the observed breaks) which signifies the best fit of data.

As can be seen in the figure, most of the breaks from the WPHM are concentrated on the ideal line. This result indicates that the model predicted the breaks for this pipe stratum well. However, as indicated by breaks below the ideal line, the model somehow overestimated some of the breaks in the fast breaking phase.

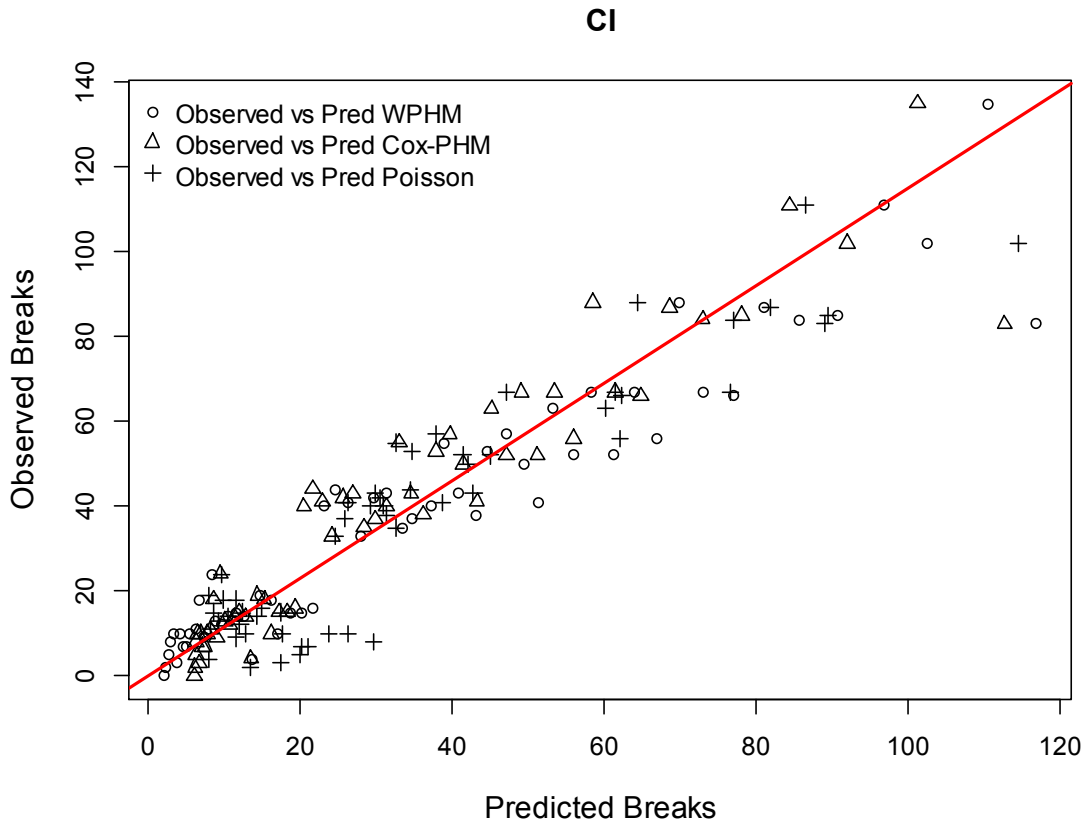


Figure 5-15: Observed and Predicted number of breaks for CI pipe strata

For the PM, most of the breaks in the early phase of pipe life are below the ideal line indicating the levels of overestimation of the breaks for this pipe (Figure 5-15). However as the pipe moves from slow to fast breaking phase, the prediction accuracy of the model increases. This result is shown by breaks closer to the ideal line in the figure.

Modelling results from Cox-PHM from the above figure indicates that during the slow breaking phases of the pipe, most of the breaks are closer to the ideal line. This result indicated that the model somehow predicted well the breaks in this phase. However, as the pipe shifts from slow to fast breaking phase, most of the breaks are observed to lie above the ideal line indicating the high levels of underestimation of predicted breaks.

Correspondingly, the observed and predicted breaks from Cox-PHM, WPHM and Poisson Model for PVC pipes are presented in Figure 5-16. Similar to the results from the PVC pipe,

the diagonal line in this figure indicates the ideal line (where the predicted breaks equals the observed breaks) which signifies the best fit of data.

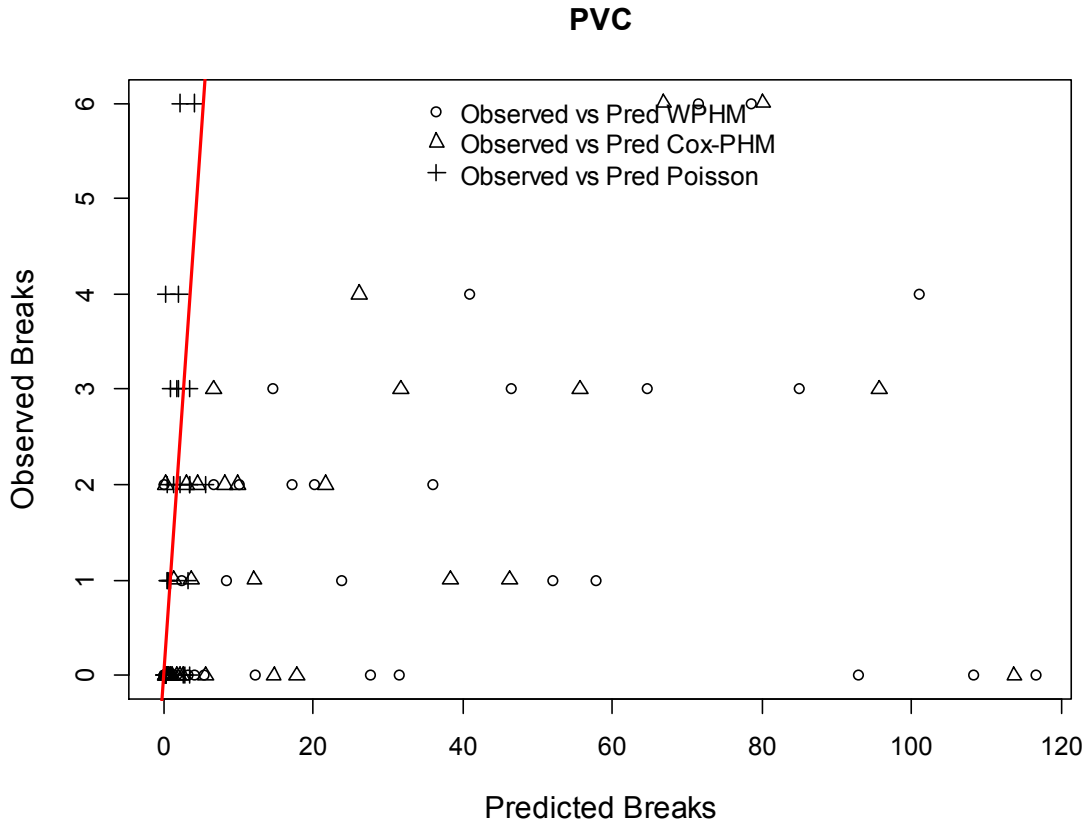


Figure 5-16: Observed and Predicted number of breaks for PVC pipe strata

As can be seen from the figure below, most of the breaks from the PM lie close to the ideal line. The result signifies that the model predicted best the breaks for pipe stratum. However for Cox-PHM and WPHM, most of the breaks are under and more so, further away from the ideal line. Therefore, this signifies that these two models overestimated the breaks for the PVC pipes.

From the prediction results, it is therefore fairly reasonable to assume that the break trend of the PVC pipe completely differs from that of metallic pipes as the number of breaks does not increase as the pipe ages. Furthermore, from the analysis given in Section 4.3.3 for the break rate of the PVC pipe, the risk factors do not seem to accelerate its failure rate.

In summary, the prediction results from the three models elicited varied results in the prediction of breaks for the metallic and PVC pipes. Cox-PHM seems to somehow predict well the slow breaking phase of metallic pipes but as the number of breaks increases (rapid breaking phase), its prediction accuracy decreases. For the PM, overestimation of the breaks is observed in the slow breaking phases of metallic pipes, but the accuracy of the prediction increases as the pipe enters fast breaking phase. In all cases, WPHM better explains both the breaking phases of the metallic pipes. In the case of the PVC break prediction, both Cox-PHM and WPHM poorly predicted the number of breaks, but PM predicted accurately the breaks.

5.7 Model Evaluation

The performance of the Cox-PHM, WPHM and PM in the prediction of pipe breaks is presented in Table 5-14. As previously discussed in Section 3.2.4, performances of the models were measured using Root Relative Square Error (RRSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Relative Absolute Error (RAE).

Table 5-14: Performance of Cox-PHM, WPHM, and PM on Break Prediction for DI, CI and PVC strata

Models	RRSE			MAE			RMSE			RAE		
	DI	CI	PVC	DI	CI	PVC	DI	CI	PVC	DI	CI	PVC
Cox-PHM	0.47	0.5	1.56	10.4	10.6	28.9	14	16	45.0	0.5	0.41	1
WPHM	0.33	0.31	1.9	5.8	7.3	29.1	9.37	9.7	55.1	0.26	0.29	1
PM	0.43	0.35	0.94	10.1	8.9	1.08	12.2	11	1.6	0.44	0.36	0.79

As can be seen from the table, when all the evaluation measures were performed for all the models considered, it is apparent that the WPHM was ranked the highest. In the case of the RRSE measure for the DI pipe, WPHM had a performance value of 0.33 against 0.47 and 0.43 from Cox-PHM and PM, respectively. The same case under the same performance measure applies for the CI pipe. Similarly for the MAE measure, WPHM had the highest performance value of 5.8 compared to 10.1 and 10.4 from PM and Cox-PHM, respectively. Moreover, it can be noted from the table that the performance of the PM and the Cox-PHM were close in some cases of evaluation. For example, the performance measures MAE and RAE indicated very close performance values for both the CI and DI_A pipes. These results shows that the models almost produced similar results though it is clear that PM is a far superior model.

Generally for all the performance measures for metallic pipes, results from Table 5-14 indicate that the best performing model in the prediction of these pipes is the WPHM. This is then followed by the PM and Cox-PHM, respectively.

For the prediction of the PVC pipe failure, all the four performance measures indicate that the PM performed best in the prediction of pipe breaks for the PVC pipe. In the case of the RRSE performance measure, PM had the highest performance value of 0.94 compared to 1.56 and 1.9 from the Cox-PHM and WPHM, respectively. Similarly, RAE performance indicator had a value of 1.08 for PM compared to 28.9 and 29.1 from the Cox-PHM and WPHM, respectively. For the other indicators, they follow similar trend in the order of PM, Cox-PHM and WPHM. However, it can be noted from the table that the performance of the Cox-PHM and WPHM were very close especially for MAE and RAE indicating that these models almost produce the same prediction results, though the Cox-PHM was more superior than the WPHM in all cases.

In conclusion, the model performance results indicate that WPHM is well suitable for modelling the metallic pipes breaks while PM is well suited for the PVC breaks pipes. In all the categories analyzed, Cox-PHM gave the lowest performance indicating that the model gives low prediction results for the metallic and PCV pipes.

Chapter 6

6 Conclusion and Recommendations

This section presents the summary of findings of this study. Also, recommendations on the suitability of the models for predictions of pipe breaks are outlined.

6.1 Conclusion

This thesis presented the use of Proportional Hazard Models (Cox-PHM and WPHM) as well as PM in the prediction of number of pipe breaks for the City of Calgary Water Distribution Network (CCWN). The use of Cox-PHM model in prediction of number of pipe breaks had been restricted by the undefined baseline hazard function thereby limiting it to application in evaluation of the effects of covariates on the hazard rate of pipes. However, the methodology proposed and applied in this thesis involves the application of simple curve fitting techniques to estimate the baseline hazard equation. The resulting equation was used in prediction of the number of pipe breaks for CI, DI_A and PVC pipes. The results from this model were then compared to results from WPHM and Poisson Model.

The major findings from the analysis of pipe data were:

- Most of the failures occurring on the CCWN occur on the metallic pipes with very small number of breaks occurring on PVC pipes.
- The overall shape of breaking pattern for the CCWN is influenced by the CI pipe material. DI, YDI and PVC pipe material does not seem to be affected by seasonal variation.
- The type of failure occurring on a pipe is dependent on the pipe material type with corrosion failure type significantly affecting DI pipe while circumferential break mostly affecting CI pipe. However, PVC pipe seem not to be prone to any break-type.
- Pipe failure follows the weather patterns with most failures occurring during winter period with less breaks occurring in summer.
- An inverse relationship between pipe diameter and pipe failure exists. As the diameter increases, the number of breaks observed by a pipe decreases.

- A higher likelihood of re-break exist for a pipe that has a failure history, with a higher chance of failure occurring within a short distance from the location of the last failure
- Cathodic protection significantly reduces the chances of pipe failing
- Though age of the pipe does contribute linearly to the increase in the pipe breaks, however, younger pipes seem to be more prone to pipe breaks than older pipes

6.1.1 The Most Important Covariates

- Pipe physical covariates, environmental and operational factors were found to be critical in influencing the occurrence of a pipe failure.
- However, the most critical covariates were found to be physical and operational covariates
- Impacts of covariates differs according to pipe material with metallic pipes more prone than PVC pipes
- DI pipe more prone to failure than the CI pipe.

6.1.2 Covariates Impacting a Particular Break-Type

- The most important covariates are physical and environmental covariates
- However, the most critical are age of pipe and the pipe diameter

6.1.3 The Robust Model for the City of Calgary Water Network

- The accuracy of prediction by the three models differed depending on the pipe material.
- The WPHM captured the failing trends of the metallic pipes better than the Cox-PHM and PM models
- Accuracy of the PM model increased as the number of breaks increases, while the accuracy of the Cox-PHM decreased with increase in breaks
- PM explained the failing trend of the PVC pipes better

6.1.4 Suitability of the Cox-PHM for break prediction

- The model produced good results for the CCWN
- Captured the slow breaking phases of the pipe better than the fast breaking phases, thus suitable for young systems
- Appealing method because it does not impose distribution on data

6.2 Recommendation

From the study conducted in this thesis, the following recommendations are made:

- Priority replacement should be given to DI more than CI pipes
- PVC pipe should be used in replacement and rehabilitation of the city's system, however final decisions should be made in accordance with life cost analysis of the pipe.
- The metallic pipes should be replaced once the fourth break has occurred
- Combination of models should be applied in the prediction of breaks for the city of Calgary, with WPHM model being applied in metallic pipes while pm be used for PVC pipes
- Kenyan water companies should be encouraged to keep pipe infrastructure data in order to facilitate a similar study to be done, and ensure that informed decision on pipe replacement and rehabilitation is done.

Chapter 7

7 References

- Abeyssekera, W., and Sooriyarachchi, M. (2009). “use of Schoenfeld’s global test to test the proportional hazards assumption in the cox proportional hazards model: an application to a clinical study.” *Journal of the National Science Foundation*, 37(1), 41–51.
- Akaike, H. (1973). “Information theory and an extension of the maximum likelihood principle.” *Breakthroughs in Statistics*, 1, 610–624.
- American Water Works Association. (2007). *Distribution system inventory , integrity and water quality*.
- American Water Works Service Co., I. 2002. (2002). *Deteriorating buried infrastructure management challenges and strategies*.
- Andreou, S. (1986). “Predictive models for pipe break failures and their implications on maintenance planning strategies for deteriorating water distribution systems.” *PhD Thesis*, Massachusetts Institute of Technology.
- Andreou, S., Marks, D., and Clark, R. (1987a). “A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory.” *Advances in Water Resources*, 10(1), 2–10.
- Andreou, S., Marks, D., and Clark, R. (1987b). “A new methodology for modelling break failure patterns in deteriorating water distribution systems: Applications.” *Advances in Water Resources*, 10(1), 11–20.
- Asnaashari, A., McBean, E. a., Shahrour, I., and Gharabaghi, B. (2009). “Prediction of watermain failure frequencies using multiple and Poisson regression.” *Water Science & Technology: Water Supply*, 9(1), 9.
- Berardi, L., Kapelan, Z., Giustolisi, O., and Savic, D. A. (2008). “Pipe deterioration models for water distribution systems.” *Journal of hydroinformatics*, 10(2), 113–126.
- Betrie, G., Tesfamariam, S., Morin, K., and Sadiq, R. (2013). “Predicting copper concentrations in acid mine drainage: a comparative analysis of five machine learning techniques.” *Environmental monitoring and assessment*, 185(5), 4171–82.
- Boxall, J., O’Hagan, A., Pooladsaz, S., and Saul, A, J. (2007). “Estimation of burst rates in water distribution mains.” *Water management*, 160(2), 73–82.
- Brander, R. (2001). “Water pipe materials in Calgary, 1970-2000.” *AWWA Infrastructure Conference Proceedings*, Orlando, Florida, 1–10.

- Brander, R. (2004). "Minimizing break rates on PVC water mains." *Proceedings of Plastic Pipe XII Conference*, Milan, Italy, 1–10.
- Brander, R., and Ng, B. (2000). "Developing a condition assessment technique for water mains." City of Calgary Waterworks Division.
- Census, C. R. (2013). *Civic census results for the city of Calgary*. Election and Information Services, Calgary, Canada, 1–123.
- Chan, D., Gould, S., and Davis, P. (2005). "Data analysis and laboratory investigation of the behaviour of pipes buried in reactive clay." 206–211.
- Christodoulou, S. (2010). "Water network assessment and reliability analysis by use of survival analysis." *Water Resources Management*, 25(4), 1229–1238.
- Clair, A., and Sinha, S. (2012). "State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models!" *Urban Water Journal*, 9(2), 85–112.
- Cox, D. (1972). "Regression models and life-tables." *Journal of the Royal Statistical Society*, 34(2), 187–220.
- Crowther, M. J., and Lambert, P. C. (2013). "stgenreg : A stata package for general parametric survival analysis." *Journal of statistical software*, 53(12), 1–17.
- Davis, P., Moglia, M., Gould, S., and Burn, S. (2004). "Physical probabilistic models to estimate failure rates in PVC pipe networks." *World Water Congress 2004*, 1–10.
- Davis, P., Silva, D., Marlow, D., Moglia, M., Gould, S., and Burn, S. (2008). "Failure prediction and optimal scheduling of replacements in asbestos cement water pipes." *Journal of Water Supply: Research and Technology—AQUA*, 57(4), 239.
- Demyanov, S., Bailey, J., Ramamohanarao, K., and Leckie, C. (2012). "AIC and BIC based approaches for SVM parameter value estimation with RBF kernels." *Journal of Machine learning*, 97–112.
- Duchesne, S., Beardsell, G., and Bouchard, K. (2012). "A survival analysis model for sewer pipe structural deterioration." *Computer-Aided Civil and Infrastructure Engineering*, 00(2012), 1–15.
- EnvironmentCanada. (2013). "Climate data online." <http://weather.gc.ca/city/pages/ab-52_metric_e.html>.
- Friedl, F., Schrotter, S., Kogseder, B., and Fuchs-hanusch, D. (2012). "Early failure detection model for water mains due to seasonal climatic impacts." *World Environmental and Water Resources Congress 2012*, 3021–3032.
- Fuchs-Hanusch, D., Friedl, F., and Kogseder, B. (2011). "Effect of seasonal climatic variance on water main failures in moderate climate regions." *Conference Proceedings of LESAM*.

- Fuchs-Hanusch, D., Kornberger, B., Friedl, F., and Scheucher, R. (2012). "Whole of life cost calculations for water supply pipes." *Water Asset Management International*, 8, 19–24.
- Le Gat, Y., and Eisenbeis, P. (2000). "Using maintenance records to forecast failures in water networks." *Urban Water*, 2(3), 173–181.
- Gibbons, M., Griffin, D., Macey, C., and Homeniuk, R. (2008). "AC Enigma: Using opportunistic sampling to assess asbestos-cement water main performance in Winnipeg." (204).
- Gould, S., Boulaire, F., Marlow, D., and Kodikara, J. (2009). "Understanding how the Australian climate can affect pipe failure." *OzWater 09, Melbourne, Australia, 16-18 March 2009*, 1–8.
- Goulter, I. C., and Kazemi, A. (1987). "Spatial and temporal groupings of water main pipe breakage in Winnipeg."
- Goulter, I., Davidson, J., and Jacobs, P. (1993). "Predicting water-main breakage rates." *Journal of Water Resources ...*, 119(4), 419–436.
- Gowlter, I. C., and Kazemi, A. (1989). "Analysis of water distribution pipe failure types in Winnipeg, Canada." *Journal of Transportation Engineering*, 115(2), 95–111.
- Greene, W. (2003). *Econometric analysis*. Prentice Hall, 509–602.
- Gupta, R. K. (2013). "Regression models use in data splitting techniques." *Journal of Mathematics*, 5(5), 53–56.
- Habibian, A. (1994). "Effect of temperature changes on water-main breaks." *Journal of transportation engineering*, 120(2), 312–321.
- Harris, N., and Radlinski, M. (2013). "Asbestos-cement pipe condition assessment and development of optimum replacement methodology." *Forensic Engineering 2012*, 268–279.
- Harvey, R., McBean, E. a., and Gharabaghi, B. (2013). "Predicting the timing of watermain failure using artificial neural networks." *Journal of Water Resources Planning and Management*.
- Hu, Y., and Hubble, D. (2007). "Factors contributing to the failure of asbestos cement water mains." *Canadian Journal of Civil Engineering*, 34, 608–621.
- Hu, Y., and Vu, H. (2006). "Field performance of water mains buried in expansive soil dans : 1 st International Structural Specialty." 1–13.
- Hu, Y., and Vu, H. (2011). "Analysis of soil conditions and pipe behaviour at a field site." *Canadian Geotechnical Journal*, 847–867.
- Hu, Y., Wang, D., and Chowdhury, R. (2010). "condition assessment methods for ac pipe and current practices." *nrcc*, (September), 1–54.

- Hu, Y., Wang, D., and Cossitt, K. (2008). "Asbestos cement water mains: history, current state, and future planning." *American Water Works Association Infrastructure conference proceedings...*, 1–13.
- Hu, Y., Wang, D., Cossitt, K., and Chowdhury, R. (2010). "AC Pipe in North America : inventory , breakage , and working environments." *Journal of Pipeline Systems Engineering and Practice*, (November), 156–172.
- Jenkins, S. P. (2005). "Survival analysis." University of Essex, unpublished Lecture Notes manuscript, Institute for Social and Economic Research,.
- Jones, D. (1983). "Statistical analysis of empirical models fitted by optimization." *Biometrika*, 70(1), 67–88.
- Karaa, F. A. (1984). "a decision support model for the investment planning of the construction and rehabilitation of mature water distribution systems."
- Karaa, F., and Marks, D. (1990). "Performance of water distribution networks: Integrated approach." *Journal of Performance of Constructed Facilities*, 4(1), 51–67.
- Kettler, A. J., and Goulter, C. (1985). "An analysis of pipe breakage in urban water distribution networks." *Canadian Geotechnical Journal*, 12(1982), 286–293.
- Kleiner, Y., and Rajani, B. (2000). "Considering time-dependent factors in the statistical prediction of water main breaks." *American Water Works Association Infrastructure conference proceedings...*, 1–12.
- Kleiner, Y., and Rajani, B. (2001). "Comprehensive review of structural deterioration of water mains: statistical models." *Urban Water*, 3(3), 131–150.
- Kleiner, Y., and Rajani, B. (2002). "Forecasting variations and trends in water-main breaks." *Journal of Infrastructure Systems*, 8(4), 122–131.
- Kleiner, Y., and Rajani, B. (2010). "I-WARP: Individual Water mAin Renewal Planner." *Drinking Water Engineering and Science*, 3(1), 25–41.
- Kleiner, Y., and Rajani, B. (2010). "I-WARP: Individual Water mAin Renewal Planner." *Drinking Water Engineering and Science*, 3(1), 71–77.
- Knight, M. L. (2002). "Failure analysis of PVC and CPVC piping materials."
- Kumar, D., and Klefsjö, B. (1994). "Proportional hazards model: A review." *Reliability Engineering & System Safety*, 44(2), 177–188.
- Kumar, D., Klefsjö, B., and Kumar, U. (1992). "Reliability analysis of power transmission cables of electric mine loaders using the proportional hazards model." *Reliability Engineering & System Safety*, 37, 217–222.

- Kuraoka, S., and Rainer, J. H. (1996). "Damage to water distribution system caused by the 1995 Hyogo-ken Nanbu earthquake." *Canadian Journal of Civil Engineering*, 23(3), 665–677.
- Liu, Z., Sadiq, R., Rajani, B., and Najjaran, H. (2010). "Exploring the Relationship between Soil Properties and Deterioration of Metallic Pipes Using Predictive Data Mining Methods." *Journal of Computing in Civil Engineering*, 24(3), 289–301.
- Loganathan, G. V, Park, S., and Sherali, H. D. (2002). "Threshold break rate for pipeline replacement in water distribution systems." *Journal of Water Resources Planning and Management*, 128(4), 271–279.
- Mailhot, A., Pelletier, G., and Noel, J.-P. V. (2000). "Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application." *Water Resources Research*, 36(10), 3053–3062.
- Makar, J. (1999). "Failure analysis for grey cast iron water pipes." *AWWA Distribution System Symposium*, 1–8.
- Makar, J. (2000). "A preliminary analysis of failures in grey cast iron water pipes." *Engineering Failure Analysis*, 7(1), 43–53.
- Makar, J., Desnoyers, R., and McDonald, S. (2001). "Failure modes and mechanisms in gray cast iron pipe." *Underground Infrastructure Research*, 1–10.
- Makar, J., and Kleiner, Y. (2000). "Maintaining water pipeline integrity." *AWWA Infrastructure Conference and Exhibition*.
- Markovic, S., Cerekovic, N., Kljajic, N., and Rudan, N. (2012). "Rainfall analyses and water deficit during growing season in Banja Luka Region." *1st International Congress of ...*
- Marlow, D., Davis, P., Beale, D., Burn, S., and Urquhart, A. (2010). "Remaining asset Life: a state of the art review." *WERF, Alexandria, Virginia*.
- Martins, A. (2011). "Stochastic models for prediction of pipe failures in water supply systems." Universidade Technica de Lisboa, Instituto Superior Tecnico, PhD Thesis.
- Martins, A., Leitão, J. P., and Amado, C. (2013). "A comparative study of three stochastic models for prediction of pipe failures in water supply systems." *Journal of Infrastructure Systems*, 1–37.
- Mordak, J., and Wheeler, J. (1988). "Deterioration of asbestos cement water mains." *Final Report to the Department of the Environment*.
- O'Day, D. (1982). "Organizing and analyzing leak and break data for making main replacement decisions." *Journal of the American Water Works Association, Asce*, 66(3), 911–917.

- Park, S., Kim, J., Newland, A., Kim, B. J., and Jun, H. (2008). "Survival analysis of water distribution pipe failure data using the proportional hazards model." *World Environmental and Water Resources Congress 2008 Ahupua'a*.
- Peel, M., Finlayson, B. L., and McMahon, T. A. (2007). "Updated world map of the Köppen-Geiger climate classification." *Hydrology and Earth Systems*, 4, 439–473.
- Pelletier, G., Mailhot, A., and Villeneuve, J. (2003). "Modeling water pipe breaks — Three case studies." *Journal of Water Resources Planning and Management*, 129(2), 115–123.
- Qi, J. (2009). "Comparison of proportional hazards and accelerated failure time models."
- Rajani, B., and Kleiner, Y. (2001). "Comprehensive review of structural deterioration of water mains: physically based models." *Urban Water*, 3(3), 151–164.
- Rajani, B., and Kleiner, Y. (2003). "Protection of ductile iron water mains against external corrosion: Review of methods and case histories." *Journal American Water Works Association*, 95(11), 110–125.
- Rajani, B., Kleiner, Y., and Sink, J. (2012). "Exploration of the relationship between water main breaks and temperature covariates." *Urban Water Journal*, 9(2), 67–84.
- Rajani, B., and Makar, J. (2001). "A methodology to estimate remaining service life of grey cast iron water mains." *Canadian Journal of Civil Engineering*, 1272(2000), 1259–1272.
- Rajani, B., and Tesfamariam, S. (2004). "Uncoupled axial, flexural, and circumferential pipe – soil interaction analyses of partially supported jointed water mains." *Canadian Geotechnical Journal*, 41, 997–1010.
- Rajani, B., and Tesfamariam, S. (2005). "Estimating time to failure of ageing cast iron water mains under uncertainties." *Water management for the 21st century*, 1–7.
- Rajani, B., and Tesfamariam, S. (2007). "Estimating time to failure of cast-iron water mains." *Proceedings of the ICE - Water Management*, 160(2), 83–88.
- Rajani, B., Zhan, C., and Kuraoka, S. (1996). "Pipe –soil interaction analysis of jointed water mains." *Canadian Geotechnical Journal*, 33(3), 393–404.
- Richardson, A., Ampt, E., and Meyburg, A. (1995). *Survey methods for transport planning*. Eucalyptus Press.
- Rogers, P. (2011). "Prioritizing water main renewals : Case study of the Denver water system." (August), 73–81.
- Rogers, P., and Grigg, N. (2009). "Failure assessment modeling to prioritize water pipe renewal: two case studies." *Journal of Infrastructure Systems*, (September), 162–171.
- Rossiter, D. (2009). "Technical note: Curve fitting with the R environment for statistical computing." *International Institute for Geoinformation Science & Earth Observation*, 0–17.

- Røstum, J. (2000). “Statistical modelling of pipe failures in water networks.” Norwegian University of Science and Technology, Hydraulics and Environmental Engineering Department.
- Royston, P. (2011). *Estimating a smooth baseline hazard function for the Cox*. University College London, Department of Statistical Science, 1–16.
- Royston, P., and Altman, D. G. (2013). “External validation of a Cox prognostic model: principles and methods.” *BMC Medical Research Methodology*, 13, 33.
- Shamir, U., and Howard, C. (1978). “An analytic approach to scheduling pipe replacement.” *Journal of the American Water Works Association*.
- Smith, T., Smith, B., and Ryan, M. (2003). “Survival analysis using Cox proportional hazards modeling for single and multiple event time data.” *statistic and data analysis*.
- Team, R. (2013). “R: A language and environment for statistical computing.” *Vienna, Austria: R Foundation for Statistical Computing*, (Version 3.0.0 (2013-04-03)).
- Tesfamariam, S., and Rajani, B. (2004). “Impact of uncertainties on the translation of the remaining pipe wall thickness to structural capacity.” *Infra 2004 Urban Infrastructure: Managing the assets, Mastering the Technology*, 1–16.
- Thornthwaite, C. (1948). “An approach toward a rational classification of climate.” *Geographical review*, 38(1), 55–94.
- Toumbou, B., Villeneuve, J.-P., Beardsell, G., and Duchesne, S. (2012). “A general model for water distribution pipe breaks: development, methodology and application to a small city in Quebec, Canada.” *Journal of Pipeline Systems Engineering and Practice*, 1–40.
- Vanrenterghem-Raven, A., Eisenbeis, P., Juran, I., and Christodoulou, S. (2004). “Statistical modeling of the structural degradation of an urban water distribution system: case study of New York City.” *World Water & Environmental Congress 2003*, 1–10.
- Walski, T., and Pelliccia, A. (1982). “Economic analysis of water main breaks.” *Journal of the American Water Works Association*, 74(3), 140–147.
- White, P. E. (1985). *interim report: usage and performance of asbestos cement pressure pipe*.
- Wood, A., and Lence, B. J. (2009). “Using water main break data to improve asset management for small and medium utilities : District of Maple Ridge , B . C .” *Journal of Infrastructure Systems*, 15(2), 111–119.
- Zhang, D. (2005). “Analysis of survival data-Lecture notes.” North Carolina State University.
- Zhao, J., Rajani, B., and Daigle, L. (2001). “Thermal performance of trench backfills used for frost protection of water service lines.” *Canadian Geotechnical Journal*, 38(1), 161–174.

8 Annexes

8.1 A-1: Climatic Data

Date/Time	Year	Month	Mean Max Temp (°C)	Mean Temp (°C)	Mean Min Temp (°C)	Extr Min Temp Flag	Total Rain (mm)	Total Snow (cm)	Total Precip (mm)
1956-01	1956	1	-9.6	-14.7	-19.7	S	0	34.5	34.5
1956-02	1956	2	-6.3	-11.8	-17.3		0	11.2	11.2
1956-03	1956	3	1	-4.5	-9.9		0.5	24.4	24.9
1956-04	1956	4	7.6	1.7	-4.3		0.8	28.2	29
1956-05	1956	5	17.4	10.4	3.3		21.6	8.6	30.2
1956-06	1956	6	19.8	13.6	7.3	S	130.6	0	130.6
1956-07	1956	7	22.9	16.6	10.2		38.6	0	38.6
1956-08	1956	8	21.8	15.3	8.8	S	79.2	0	79.2
1956-09	1956	9	17.3	10.5	3.6	S	21.1	0	21.1
1956-10	1956	10	9.7	3.1	-3.6	S	0.3	20.8	21.1
1956-11	1956	11	7.7	1.6	-4.6	S	0	12.2	12.2
1956-12	1956	12	-0.3	-6.8	-13.3		0	21.3	21.3
1957-01	1957	1	-9.3	-14.6	-19.9		0	25.1	25.1
1957-02	1957	2	-3.6	-9.9	-16.1		0	13.7	13.7
1957-03	1957	3	2.8	-1.9	-6.5		0.3	16.5	16.8
1957-04	1957	4	9.3	3.4	-2.5		8.4	17.8	26.2
1957-05	1957	5	17.7	10.7	3.6		10.7	10.4	21.1
1957-06	1957	6	19	12.9	6.7		64	0	64
1957-07	1957	7	24.2	16.7	9.2		41.7	0	41.7
1957-08	1957	8	19	13.1	7.1	S	78.7	0	78.7
1957-09	1957	9	18.8	12.5	6.1		26.7	0.3	26.9
1957-10	1957	10	6.5	1	-4.6		3	47.2	50.3
1957-11	1957	11	4.2	-1.1	-6.4		0.3	23.9	24.1
1957-12	1957	12	3.8	-2.4	-8.6		0	5.3	5.3
1958-01	1958	1	3.7	-2.2	-8.1		0	7.9	7.9
1958-02	1958	2	-6.3	-10.7	-15.1		0	19.6	19.6
1958-03	1958	3	-3.6	-7.8	-12		0	25.1	25.1
1958-04	1958	4	8.6	3.7	-1.3	S	11.9	30.7	42.7
1958-05	1958	5	20.9	13.1	5.2		15.5	0	15.5
1958-06	1958	6	19.1	13.5	7.8		96	1	97
1958-07	1958	7	22.3	15.9	9.4		60.7	0	60.7
1958-08	1958	8	24.7	17.1	9.5	S	17.3	0	17.3
1958-09	1958	9	17.6	10.7	3.8		56.9	0.3	57.2
1958-10	1958	10	15.3	7.5	-0.4		2	0	2
1958-11	1958	11	1.9	-3.2	-8.3		0	17.3	17.3
1958-12	1958	12	0.7	-4.4	-9.5		0	6.4	6.4
1959-01	1959	1	-8	-13	-17.9		0	11.2	11.2

1959-02	1959	2	-4.4	-10.3	-16.2	S	0.3	17.3	17.5
1959-03	1959	3	6.1	0.9	-4.3		0	8.1	8.1
1959-04	1959	4	10.4	3.7	-3.1		1	24.9	25.9
1959-05	1959	5	13.5	7.2	0.8		10.4	35.8	46.2
1959-06	1959	6	19.4	13.4	7.3		116.6	0	116.6
1959-07	1959	7	24.9	17.4	9.8		58.4	0	58.4
1959-08	1959	8	19.1	13.3	7.4	S	65.8	0	65.8
1959-09	1959	9	14.5	8.8	3		17	1	18
1959-10	1959	10	9.3	4.1	-1.2		5.3	4.3	9.7
1959-11	1959	11	1.2	-4.3	-9.8		0.8	36.6	37.3
1959-12	1959	12	3.9	-1.5	-6.9		0	14	14
1960-01	1960	1	-4	-9.4	-14.7		0	19.8	19.8
1960-02	1960	2	-3.2	-8.1	-13		0	31.2	31.2
1960-03	1960	3	0.6	-5	-10.5		0.3	5.6	5.8
1960-04	1960	4	11	4.6	-1.9		0	30.2	30.2
1960-05	1960	5	15	8.9	2.7		51.8	1	52.8
1960-06	1960	6	19.5	13.3	7	S	86.1	0	86.1
1960-07	1960	7	26.3	18.6	10.9		42.7	0	42.7
1960-08	1960	8	21.9	15.2	8.5	S	41.7	0	41.7
1960-09	1960	9	19.2	11.4	3.6		11.7	0	11.7
1960-10	1960	10	13.5	7.2	0.8		5.8	14	19.8
1960-11	1960	11	2.2	-3.3	-8.8		0	15.5	12.7
1960-12	1960	12	1.5	-4.7	-10.8		0	26.2	17.5
1961-01	1961	1	2.6	-3	-8.6		0	9.1	5.6
1961-02	1961	2	1.2	-4.2	-9.5	S	0	44.2	30
1961-03	1961	3	4.4	-0.8	-6		1	5.6	4.8
1961-04	1961	4	8.4	2.6	-3.2		6.6	37.1	37.8
1961-05	1961	5	16.4	10.5	4.6	S	34	10.9	42.7
1961-06	1961	6	25.9	17.7	9.4		9.9	0	9.9
1961-07	1961	7	23.7	17.1	10.5		153.7	0	153.7
1961-08	1961	8	25.8	18	10.1		26.7	0	26.7
1961-09	1961	9	14.9	8	1		20.8	2	22.1
1961-10	1961	10	10.3	4.5	-1.4		3	47.5	41.4
1961-11	1961	11	2.8	-3.5	-9.8		0	5.6	3.8
1961-12	1961	12	-5.1	-11.1	-17	S	0	17.8	12.4
1962-01	1962	1	-1.6	-8	-14.3		0	16	11.7
1962-02	1962	2	-3.9	-9.6	-15.3		0	17.3	11.2
1962-03	1962	3	-0.2	-6.1	-11.9		0	14.2	10.4
1962-04	1962	4	13.9	6.4	-1.2		0	20.1	15
1962-05	1962	5	14.9	8.8	2.6		51.6	10.9	56.9
1962-06	1962	6	21.3	14.5	7.7		45.2	0	45.2
1962-07	1962	7	22.8	15.5	8.2		33.3	0	33.3
1962-08	1962	8	22.6	15.5	8.3		51.3	0	51.3
1962-09	1962	9	19.3	11.6	3.8		21.6	11.4	26.9

1962-10	1962	10	14.7	7.7	0.6	S	8.1	2	9.7
1962-11	1962	11	7.3	1.4	-4.5		1.3	3.6	4.3
1962-12	1962	12	2.1	-4.7	-11.4		0	10.2	7.6
1963-01	1963	1	-4.6	-11.1	-17.6		0.3	36.1	25.4
1963-02	1963	2	2.2	-2.9	-7.9		0.5	13	7.9
1963-03	1963	3	6.2	0.2	-5.9	S	0	18.8	14.2
1963-04	1963	4	10.8	4.2	-2.4		6.9	14	19.1
1963-05	1963	5	15.9	9.2	2.4		2.8	19.6	19.6
1963-06	1963	6	20.6	14.2	7.8		146.3	0	146.3
1963-07	1963	7	23.4	16.9	10.3	S	90.4	0	90.4
1963-08	1963	8	23.8	16.8	9.8		16.8	0	16.8
1963-09	1963	9	22.2	14.8	7.3	S	50.5	0	50.5
1963-10	1963	10	16.4	8.9	1.4		0	0	0
1963-11	1963	11	2.7	-3.8	-10.2		0	19.6	13.2
1963-12	1963	12	-1.6	-8.3	-14.9		0	23.4	21.1
1964-01	1964	1	-0.8	-6.4	-11.9		0	5.8	1.8
1964-02	1964	2	5.8	-0.1	-5.9		0	2	2
1964-03	1964	3	0.4	-6.3	-12.9	S	0	17.5	8.4
1964-04	1964	4	10.2	3.7	-2.9	S	0.3	18.3	13.2
1964-05	1964	5	15.6	9.9	4.1		58.7	5.1	63.8
1964-06	1964	6	20.1	13.9	7.7		101.6	0	101.6
1964-07	1964	7	24.9	17.4	9.9		72.1	0	72.1
1964-08	1964	8	22.4	15	7.6	S	8.1	0	8.1
1964-09	1964	9	14.4	8.4	2.4		51.3	6.1	57.4
1964-10	1964	10	15.2	8	0.8		7.1	15.2	18
1964-11	1964	11	1.1	-3.8	-8.7		0	29	20.8
1964-12	1964	12	-11.7	-17.2	-22.6	S	0	33.5	24.1
1965-01	1965	1	-5.6	-11.8	-17.9		0	17.3	11.7
1965-02	1965	2	-1.5	-8.1	-14.6	S	5.1	16.8	16
1965-03	1965	3	-2.8	-8.2	-13.6		0.3	24.1	16
1965-04	1965	4	9.3	3.5	-2.4	S	2.3	11.7	12.4
1965-05	1965	5	15.2	8.4	1.6		36.1	8.6	44.2
1965-06	1965	6	18.6	12.5	6.4		169.9	0	169.9
1965-07	1965	7	23.3	16.9	10.4		117.6	0	117.6
1965-08	1965	8	23	16.6	10.2	S	69.9	0	69.9
1965-09	1965	9	10.4	5.1	-0.2		52.6	33	81.5
1965-10	1965	10	15.9	8.4	0.8	S	11.2	0.3	11.4
1965-11	1965	11	0.2	-5.7	-11.6		0.3	43.7	31
1965-12	1965	12	-3.9	-9.6	-15.3	S	0	12.4	7.6
1966-01	1966	1	-12.6	-18	-23.3	S	0	11.7	10.2
1966-02	1966	2	-0.9	-7.2	-13.4		0	7.4	5.3
1966-03	1966	3	5.7	-0.6	-6.8		0	5.8	4.8
1966-04	1966	4	7.2	0.8	-5.7		0.8	59.7	42.7
1966-05	1966	5	17.3	10.2	3	S	57.7	1	58.4

1966-06	1966	6	18	11.9	5.7	S	79.2	0	79.2
1966-07	1966	7	21.7	15.6	9.5		113.3	0	113.3
1966-08	1966	8	20.6	13.9	7.1	S	30.5	0	30.5
1966-09	1966	9	20.8	13	5.1	S	4.6	0	4.6
1966-10	1966	10	11.8	5.2	-1.4		6.9	14.5	14.5
1966-11	1966	11	-1.5	-7.7	-13.9		0.5	53.8	33.8
1966-12	1966	12	-2.4	-7.7	-12.9	S	0	8.9	5.6
1967-01	1967	1	-5.6	-11.7	-17.7		0	32	20.6
1967-02	1967	2	1.3	-5.3	-11.8		0	13.7	9.7
1967-03	1967	3	-2.3	-7.9	-13.4		0	26.4	19.6

8.2 A-2: Cox Proportional Hazard Assumption

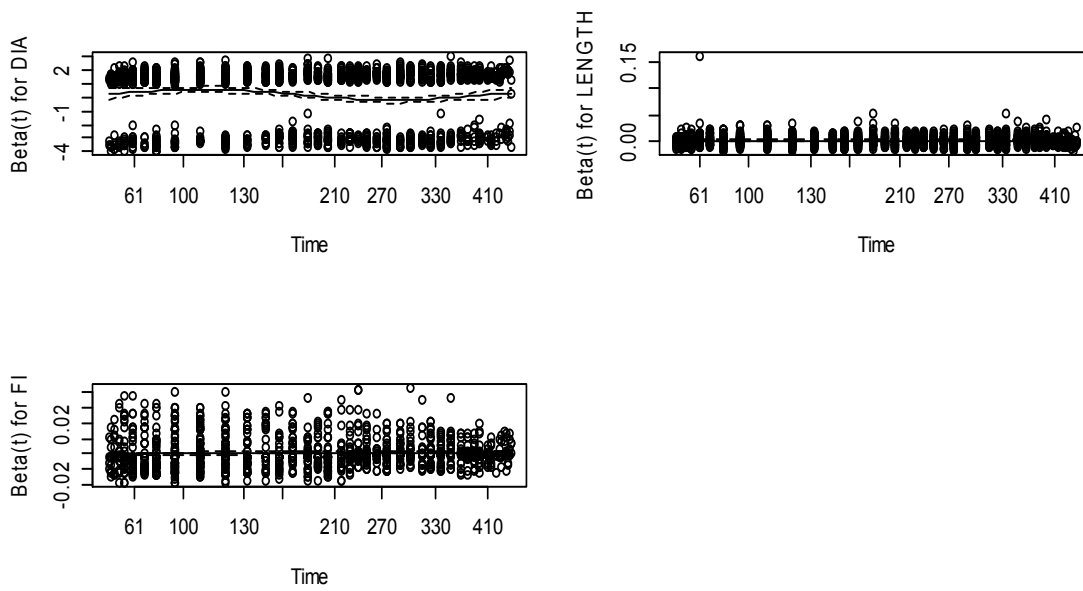


Figure 8-1: Schoenfeld's residuals for the DI pipe strata with less than one break

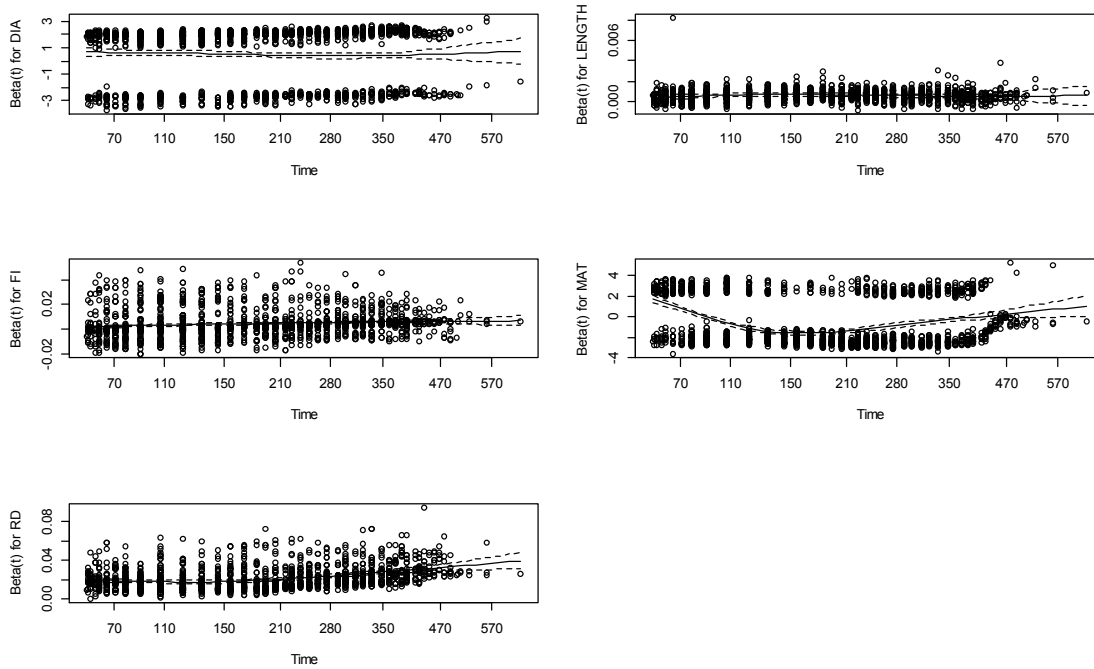


Figure 8-2: Schoenfeld's residuals for the DI pipe strata with more than one break

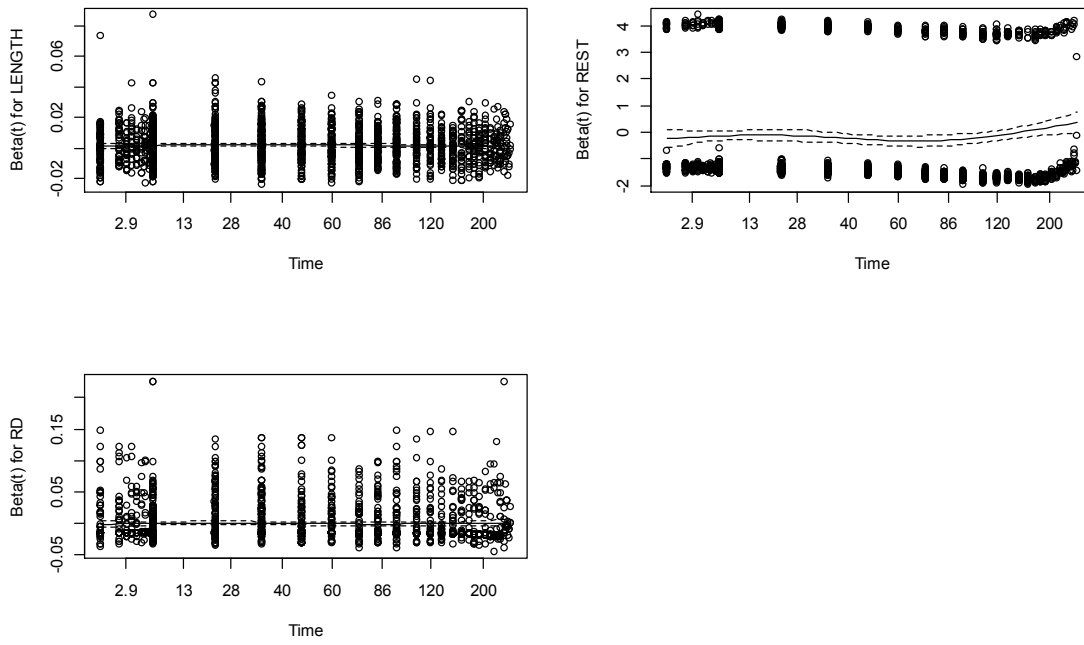


Figure 8-3: Schoenfield's residuals for the DI pipe strata with more than one break

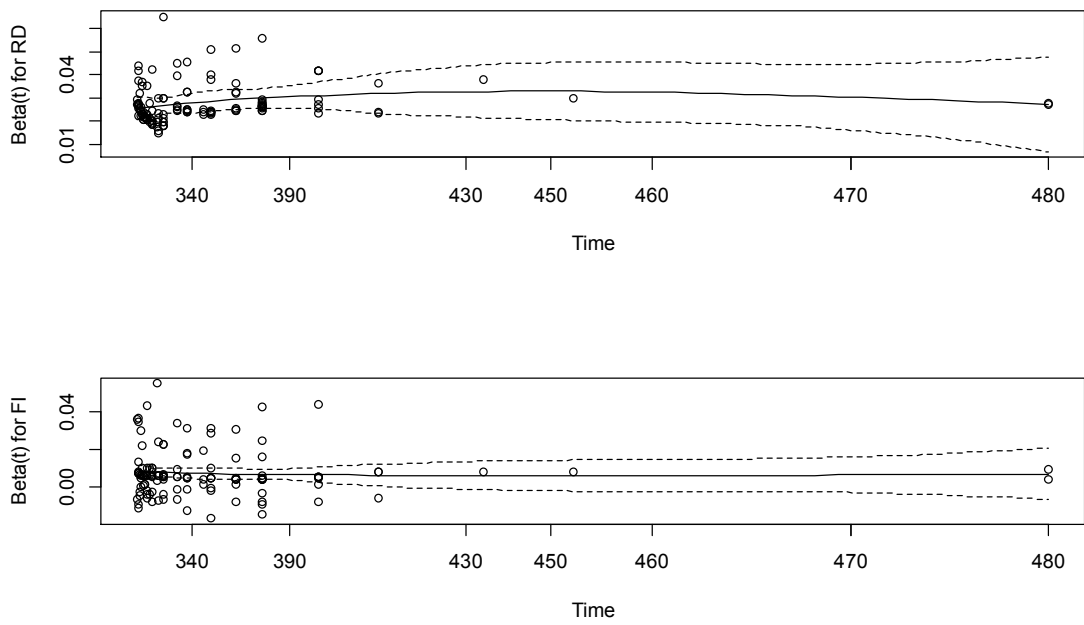


Figure 8-4: Schoenfield's residuals for the PVC pipe strata