12-2017

Systematic Review for Water Network Failure Models and Cases

Yufei Gao
University of Arkansas, Fayetteville

Systematic Review for Water Network Failure Models and Cases

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Industrial Engineering

by

Yufei Gao
Wuhan University of Technology
Bachelor of Science in Material, 2014

December 2017
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Shengfan Zhang, Ph.D.
Thesis Director

Haitao Liao, Ph.D.
Committee Member

Manuel D. Rossetti, Ph.D.
Committee Member

Wen Zhang, Ph.D.
Committee Member

Abstract

As estimated in the American Society of Civil Engineers 2017 report, in the United States, there are approximately 240,000 water main pipe breaks each year. To help estimate pipe breaks and maintenance frequency, a number of physically-based and statistically-based water main failure prediction models have been developed in the last 30 years. Precious review papers focused more on the evolution of failure models rather than modeling results. However the modeling results of different models applied in case studies are worth reviewing as well.

In this review, we focus on research papers after Year 2008 and collect latest cases without repetition. A total of 64 papers are qualified following the selection criteria. Detailed information on models and cases are summarized and compared. Chapter 2 provides a summary and review of failure models and discusses the limitation of current models. Chapter 3 provides a comprehensive review of collected cases, which include network characteristics and factors. Chapter 4 focuses on the main findings from collected papers. We conclude with insights and suggestions for future model selection for pipe failure analysis.

Contents

Chapter 1 Introduction

Drinking water is delivered via one million miles of pipes across the U.S. Aging pipe has been one of the major challenges facing the water industry due to the limitation of funding availability. Many of those pipes were laid in the early to mid 20th century with a lifespan of 75 to 100 years (American Water Works Association (AWWA), 2017). Using these average life estimates and counting the years since the original installations shows that these water utilities will face significant needs for pipe replacement over the next few decades. Some components in water and sanitation conveyance systems in the United States and Europe are more than 100 years old (AWWA 2017). Aging pipes present many technical limitations for effective water provisioning. Firstly, degradation of infrastructure system integrity leads to system losses and water leaks. The water lost in the conveyance process is often referred to as "nonrevenue water" because it leaves the system prior to the water meter, which is generally used to define cost paid by the user. Secondly, supplied water by pipes with breaks generally carries a higher risk of contamination, which could lead to various potential health impacts for users. As estimated in the American Society of Civil Engineers (ASCE) 2017 Infrastructure Report Card, in the United States, there are approximately 240,000 water main pipe breaks each year (ASCE 2017). As a result, 10% to 30% of total water is non-revenue water, while in England this value has recently been estimated to be 25% (ASCE 2015). It is projected that above 1 million miles of water mains need replacement, as estimate by AWWA (2017). The replacement cost is estimated to be approximately $1 trillion to maintain and expand service to meet demand over the next 25 years (ASCE 2017). However, constrained by the limited resources available, efficient maintenance and management of water infrastructure, particularly pipe maintenance and repair in the distribution system, is challenging but imperative.

To deal with this problem, a number of physically-based and statistically-based water main failure prediction models have been developed in the last 30 years. Physical models predict breaks by simulating the mechanics of pipe failure and the capacity of a pipe to resist failure. Statistical models are developed with historical data on pipe breaks to identify failure patterns, and they extrapolate these patterns to predict future pipe breaks (MJ Nishiyama, 2013).

1.1 Motivation

Most papers about water network failure focused on failure model development and validation, with case studies using one or more real database of networks. Previous review papers on failure models summarized the evolution of models, compared the differences between various models, and defined a variety of classification of models (Clair & Sinha, 2012; Nishiyama & Filion, 2013). However, all these discussion and comparisons did not mention much about the applied cases. The application of each single case and the specific conclusion for real data are seldom reviewed in the past 20 years. The characteristics of cases covers a lot of information such as region (water and air temperature, the depth of pipes), pipeline scale (the number of pipes ranging from tens to thousands), date of construction (which is highly associated with pipe material used), the state of maintenance and record (the frequency of maintenance and the integrity of maintenance record).

A better understanding of the relationship between failure prediction and network characteristics would be useful for failure model selection when an analyst works on another similar real case. In addition, previous conclusions from case studies could be used as validation for prediction and direction for analysis in the future. Therefore, a comprehensive review for water network cases and results is necessary and worthwhile.

1.2 Objectives and Design of Systematic Review

The overall goal of this paper is to provide a comprehensive review of recent water network failure models and cases. In this review, we focus on research papers after Year 2008 and collect latest cases without repetition. Detailed information on models and cases such as attributes of networks considered in the models are summarized and compared. Papers selected in this research are searched by key words: pipe failure, water distribution, failure prediction, pipe break, pipe deterioration. A few papers were collected from the citation of pervious review papers (Genevieve Pelletier 2003; Berardi et al. 2008). After paper collection, case screening was processed by several principles: remove papers before 2008 and keep papers that have the case study part. According to the flow diagram in **Figure 1**, 64 cases were collected in total.

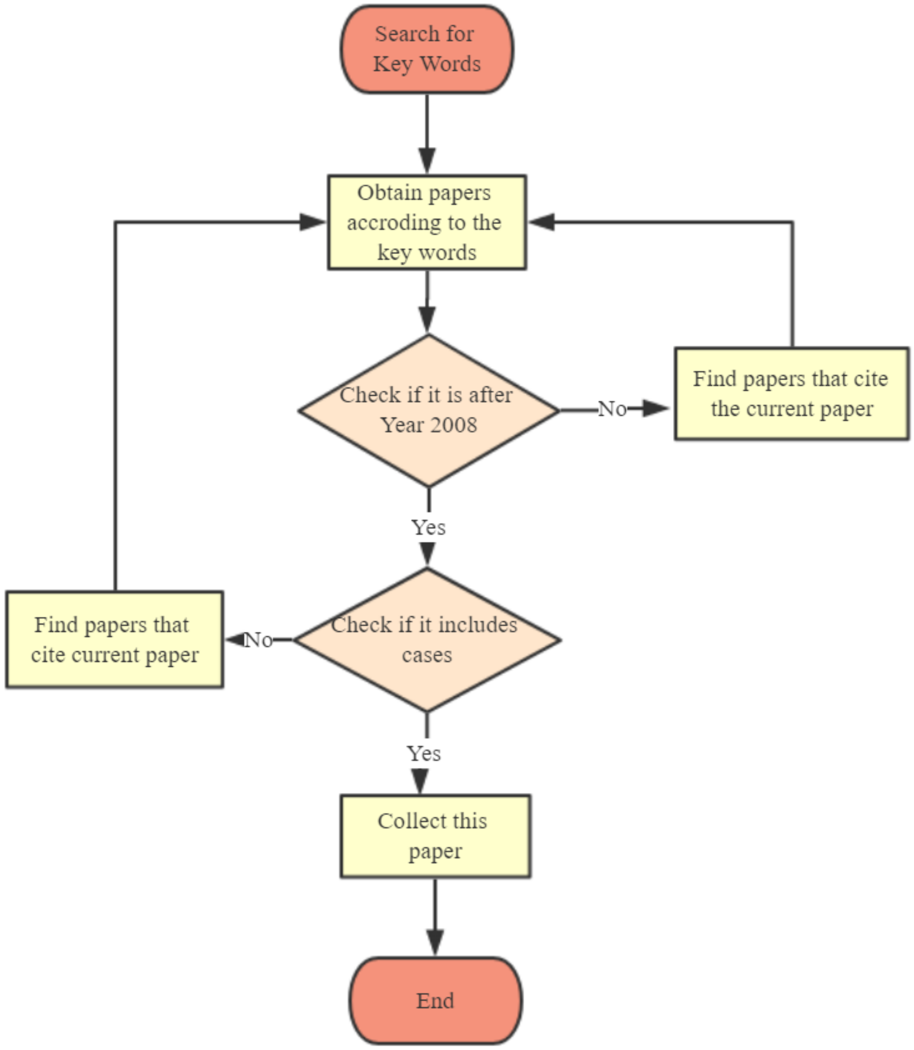

Figure 1. Selection Criteria Flow Diagram

1.3 Organization of the Thesis

The remainder of the thesis is organized as follows. Chapter 2 provides a brief review of failure models in previous review papers and discusses the limitation of current models. Chapter 3 provides a comprehensive review of collected cases which include network characteristics and factors. Chapter 4 focuses on the main findings from collected papers. Common points are extracted as insights and suggestion for future model selection in pipe failure.

Chapter 2 Review of Water Network Failure Models

During the last three decades, researchers developed different models to predict the failure of water pipes for a reliable infrastructure management. These failure prediction models can be classified into four categories: deterministic, statistical, stochastic, artificial intelligence models. In the next few sections, we first review each category in detail with a focus on the studies in the last decade and then summarize in Section 2.5.

2.1 Deterministic Models

Deterministic models usually are used in cases where the relationship between inputs and output is clear. In two approaches the deterministic models can be applied: empirical and mechanistic. Empirical approach tries to find the relations between failure rates as the output and the features and attributes of a group of pipes as the inputs, while the mechanistic approach can forecast the remaining useful life of an individual asset (just one pipe). Many papers (Kwietnieswki et al. 1993; Kowalski 2013; Kutylowska 2014) used a similar definition of failure rate. The value of λ is determined from operational data using number of pipe failures in unit time interval divide average pipeline length in a time period and the observation time. The problem of these models is that a deterministic model can be applied just in a specific location (Clair and Sinha 2012).

2.2 Probabilistic Models

Probabilistic models analyze the probability of an event occurring (Creighton 1994). The probability of occurrence is one and the probability of the event that cannot happen is zero. The other probability of occurrence should be between 0 and 1 (Mitrani 1998). Information about asset conditions and attributes are required to develop a probabilistic model. The output or dependent variable would be a range of values instead of the specific number. These models need extensive data and typically used in infrastructure assets (Clair and Sinha 2012). It should be noted that the probabilistic approach commonly increases the computational complexity of the models (Moglia 2007).

The Evolutionary Polynomial Regression (EPR) technique was first presented by Giustolisi and Savic (2006). The technique utilizes the huge potential of conventional numerical regression techniques and the strength of Genetic Algorithm in solving optimization problems (Xu et al. 2011). Later, this approach was used by other researchers in several engineering fields. Savic et al. (2006) and Ugarelli et al. (2008) used EPR to model the sewer pipe failures. Berardi et al. (2008) and Xu et al. (2011) applied the EPR to develop deterioration models for water distribution networks. Rezania et al. (2008) utilized the EPR methodology to evaluate the uplift capacity of suction caissons and shear strength of reinforced concrete deep beams. Elshorbagy and El-Baroudy (2009) compared the EPR and Genetic Programming to develop the prediction model of soil moisture response.

Guistolisi and Savic (2009) tested the EPR-MOGA (an improved EPR) to develop a model to forecast the groundwater level based on the amount of rainfall each month. El-Baroudy et al. (2010) utilized the EPR to develop the evapotranspiration process then compared the efficiency of

Evolutionary Polynomial Regression to Artificial Neural Networks (ANNs) and Genetic Programming (GP). Markus et al. (2010) applied EPR, ANNs and the naive Bayes model to forecast weekly nitrate-N concentrations at a gauging station. Ahangar-Asr et al. (2011) applied EPR to predict mechanical properties of rubber concrete. Fiore et al. (2012) used EPR to provide the predicting torsional strength model of reinforced concrete beams.

Moglia et al. (2007) developed a physical probabilistic failure prediction model based on the fracture mechanics of cast 30 iron water pipes. The random independent variables were added to the inputs, and then Monte-Carlo simulation technique was applied to deal with the computational complexity of the model. The developed model without failure data, degradation and load data, was not capable of estimating failure rates of water pipes. Whereas, with these data, it can predict failure rates more accurately.

Li et al. (2009) used the mechanically-based probabilistic model to predict remaining useful life and failure probability of buried pipes. They considered the effect of random inputs and used Monte-Carlo simulation framework to calculate cumulative distribution function (CDF) of remaining useful life of pipelines. But, they did not consider the correlation of defects for a pipeline having more than one corrosion defects. Also, they found CDF more suitable than probability density function (PDF) and reliability index in describing the probability of failure.

It should be mentioned that this technique requires a large historical dataset that contains a number of data points collected over a period to develop a promising statistical model (Clair and Sinha

2012). There has been an extensive effort during the past decades to develop the failure rate prediction model by using statistical approach.

Berardi et al. (2008) developed a water pipe deterioration model using Evolutionary Polynomial Regression. As it is mentioned before, they used a dataset that was classified into homogeneous groups based on the age and diameter of the pipe. The developed model can predict the number of breaks in each group. Then, for predicting the failure rate for each pipe, a general structural deterioration model based on EPR aggregated model was developed.

Wang et al. (2009) utilized five multiple regression models for a range of pipe materials (gray cast iron, ductile iron without lining, ductile iron with lining, PVC, and hypericin) to forecast the annual failing rate of individual water pipe rather than a homogeneous group. The overall model robustness was measured by F-test and the significant of each independent variable was measured by t-test. The model was validated using 20% of their collected dataset that was randomly selected.

Wang et al. (2010) employed the Bayesian inference to assess the condition of water pipes. Ten factors from three pipe materials (cast iron, ductile cast iron, and steel) were used to generate factor weight. According to the results of these experiments, the age of pipe is the most critical variable 28 while, the model was not sensitive to some factors like trench depth, electrical recharge, and some road lanes.

Xu et al. (2011) developed two prediction models for failure rate using Evolutionary Polynomial Regression and Genetic Programming, and then they compared the results of these two models.

Results were measured based on; 1) error between predicted and actual data, 2) parsimony of generated equation, and 3) ability to justify the generated equations based on the engineering knowledge. The results showed that EPR has some advantages over GP in equation uniformity and parameters estimation, while GP was better to find the complex relations.

Osman and Bainbridge (2011) employed two statistical deterioration models to predict future failures of water pipes: rate-of-failure models (ROF) and transition-state (TS) models. ROF model extrapolates the failure rate for a specific group of water pipes that were classified based on age and some environmental factors. This model does not differentiate the times between successive pipe breaks for an individual segment while, the transition-state model focuses on finding the time between successive failures for the water pipes. TS models are dependent on the availability of sufficient and accurate data, but ROF models can be applied to limited historical data. The stresses in the buried pipes, which increase the probability of pipe failure, might be caused by the ground movement.

Kabir et al. (2015) presented Bayesian Model Averaging (BMA) method to select the most critical explanatory variables. Then the Bayesian Weibull Proportional Hazard 29 Model (BWPHM) is applied to provide the survival curves and to forecast the failure rate of two pipe types: cast iron and ductile iron.

Kabir et al. (2014) assessed the risk of failure of metallic water pipes using a Bayesian Belief Network (BBN). Bayesian Belief Network can be interpreted as a probabilistic graphical model that can represent a collection of some covariates and their probabilistic relationships. This model recognizes the most vulnerable and sensitive pipe segments through the water pipe networks. The proposed model is good just for small to medium utilities with limited data.

Jenkins et al. (2014) tried to address the problem of limited, incomplete, or uncertain data in water distribution networks. Two main modification were added to Weibull hazard rate models (WPHM) to improve the prediction performance of the models: the expert opinion and the spatial analysis. But these two modifications were not tested in the other utilities.

Francis et al. (2014) analyzed the water distribution systems to develop a pipe breaks prediction model using Bayesian Belief Networks (BBNs). They illustrated that assessing water pipe network is not only important for the failure prediction model but also is crucial for avoiding water loss and water quality degradation.

Kabir et al. (2015) stated that uncertainty regarding quality and quantity of databases became a major concern for failure prediction model development of infrastructure assets. Thus, they tried to reduce these uncertainties by developing failure prediction model for water mains using a new Bayesian belief network based data fusion model. The proposed model can identify the most vulnerable and sensitive pipe in the entire network, as well as the total number of pipes that require the immediate and appropriate action like maintenance, rehabilitation, and replacement

Konstantinos Kakoudakis et al. (2017) presented a new approach for improving pipeline failure predictions by combining a data-driven statistical model, i.e. evolutionary polynomial regression

(EPR), with K-means clustering. The EPR is used for prediction of pipe failures based on length, diameter and age of pipes as explanatory factors. Individual pipes are aggregated using their attributes of age, diameter and soil type to create homogenous groups of pipes. The created groups were divided into training and test datasets using the cross-validation technique for calibration and validation purposes respectively. The K-means clustering is employed to partition the training data into a number of clusters for individual EPR models

2.3 Stochastic model

A stochastic model is a tool for estimating probability distributions of potential outcomes by allowing for random variation in one or more inputs over time. Poisson process, nonhomogeneous Poisson process, Yule process are classified in this type. To see occurrences of pipe breaks over a certain period as stochastic point processes is one of the common ways to model them. (Kleiner and Rajani, 2001; Gat and Eisenbeis, 2001). One of the point processes that is often used is the non-homogeneous Poisson process (NHPP). This is because its great flexibility allows it to capture the non-linear relationship of the break rate with time without giving up on the inclusion of suitable pipe factors (Loganathan et al., 2002). Li Chik et al. (2016) used the NHPP, hierarchical beta process (HBP), and a newly-developed Bayesian simple model (BSM) for short-term failure forecasting with a few water utility failure data sets. After close analysis of the prediction curves, they found that the performance of the three models are of great similarity in terms of pipe ranking. However, compared with the other models, the BSM is relatively simpler, which has given it more edges. The covariate, the number of known past breaks, can be very important when it comes to the relative ranking of the pipes in the network. The NHPP and HBP are recommended if the total number of failures in the network is required.

2.4 Artificial Intelligence and Machine Learning Methods

Artificial intelligence and machine learning models, which include Artificial Neural Networks (ANN), Least square support vector machine method (LS-SVM) and Fuzzy set theory models, become more and more popular in recent years due to its capability of dealing with complex data. ANN is a method that can predict pipe failure and deterioration of infrastructure specially buried pipes. The ANN follows the pattern of the human brain using its generalization capabilities. Thus, this technique is able to process information even under large, complex, and uncertain environment. The high-quality database is needed for supervised training and forecasting the future condition of the pipes. Moreover, ANN needs several controlling factors including: number of hidden layers, the number of neurons in each hidden layer, activation functions, the number of training epochs, learning rate, and momentum term. However, ANN is considered as a "Black- 32 Box" technique. Therefore, it is not able to provide insight into the relationship between dependent and independents variables (Clair and Sinha 2012; Moselhi and Hegazy 1993, Atef et al. 2015, Shirzad et al. 2014). Fuzzy Logic is a mathematical method in the field of artificial intelligence that widely used by researchers to assign a value to a certain degree of membership instead of crisp values such as zero and one. This method is known to deal with systems that are subject to uncertainties and ambiguities. Fuzzy Logic is applicable in infrastructure assets like oil and gas, water, bridges and highways (Siler and Buckley 2005, Clair and Sinha 2012).

Jafar et al. (2010) employed ANN to analyze the urban water mains. Six ANN models that predict the failure rate of water pipes of a city in France were developed then, they tried to estimate the optimal rehabilitation/replacement time for the same network. These prediction models were tested and validated using cross validation. In the first part of this article, data collection was explained then development and validation of ANN models were discussed. In the data collection part, correlation and chi2 method were applied to select the most critical inputs.

Asnaashari et al. (2013) studied two different methods to forecast the water pipe's failure rate. Multi Linear Regression (MLR) and ANN were utilized, and their results were compared. The value of R-Squared showed that the ANN model (R2=0.94) is more promising while the MLR technique (R2=0.75) is just good enough for preliminary assessment. Shirzad et al. (2014) compared the predictive performance of ANN and Support Vector Regression (SVR) in forecasting the water pipe's breakage rate. In addition, they investigated the effect of hydraulic pressure (average and maximum hydraulic pressure values) on precision of predicting the pipe's failure rate. The results showed that the ANN model is more accurate, but it is not suitable for generalization purposes. Thus, for management purposes, SVR might be more appropriate.

Kutyłowska (2014) predicted the failure rate of pipes in an urban water utility using ANN. They employed quasi-Newton approach to train the model. The house connections and distribution pipes are considered as two different sections in database, and the results for both were acceptable.

Aydogdu and Firat (2014) incorporated two methods: fuzzy clustering and Least Squares Support

Vector Machine (LS-SVM) in order to estimate the failure rate of water pipes. At first, they developed failure rate estimation model using LS-SVM, and then fuzzy clustering method is utilized to define nine sub-regions for predictive performance improvement of the model. For model evaluation they employed some measurement indexes such as Correlation Coefficient (R), Efficieny (E) and Root Mean Square Error (RMSE).

2.5 Summary and Limitation of Previous Studies

Table 1. Classification of Models and Corresponding Number of Cases in the Literature.

Classification	Models	Number of Cases
Deterministic Models	Failure Rate	8
Probabilistic Models	Linear Regression, Evolutionary Polynomial Regression (EPR), Weibull Proportional Hazard Model (WPHM), Bayesian Belief Network, Weibull/Exponential Distribution (WE)	31
Stochastic Models	Poisson process, NHPP, Yule process	12
Artificial Intelligence/Machine Learning	Artificial Neural Network(ANN), Fuzzy Clustering; Least square support vector machine method (LS-SVM)	15

As shown in **Table 1**, it is obvious that statistical models have been the most popular method for failure prediction compared to other model types. Statistical models are used most frequently in the latest 10 years although it mostly requires large number of available factors in dataset. Deterministic model mainly refers to failure rate model which is usually only need the failure number in a period of time, so it is easy to use than others. Stochastic models mostly used for data that only include process information even under large, complex, and uncertain environment. In most cases, datasets were clustered into different groups, based on the pipe material, and then one model was developed for each group. Thus, there are several models just for one network that might be tough to implement in the real world. Several techniques were utilized by the other authors. Particularly, ANNs are commonly used in many studies. ANN is able to develop accurate prediction models in complex and uncertain environments. However, EPR is selected because it does not require large datasets for training and unlike ANN, it enables the recognition of correlations among dependent and independent variables. Being as such, EPR is not a “Black-Box” technique, but it is classified as a “Grey-Box” technique that can provide insight into the relationship between inputs and the output. The process of development and selection of EPR contains the engineering 36 knowledge that allows the user to understand the generated equations and correlation between variables involved. In ANN, each attempt delivers particular output, which can be different in other attempts with the same inputs and features, while, in EPR or generally regressions, all similar attempts lead to the same equations as the output. Advantage summary form

Chapter 3 Review of Contributing Factors

In this section, we summarize factors contributing to water network failure with two parts. We first discuss classification of various factors in the literature, and then summarize the effects of commonly used factors on network failures.

3.1 Classification of Factors

InfraGuide (2003) classified the factors contributing to the deterioration of water pipes to three main categories: physical, environmental and operational. According to this classification, physical factors include pipe material, pipe wall thickness, pipe age, pipe vintage, pipe diameter, type of joints, thrust restraint, pipe lining and coating, dissimilar metals, pipe installation and pipe manufacture. Pipe bedding, trench backfill, soil type, groundwater, climate, pipe location, disturbances, stray electrical currents, and seismic activity are considered as the environmental factors, while other researchers included rainfall, traffic and loading, and trench backfill as the environmental factors as well (Kabir et al. 2015). The internal water pressure, transient pressure, leakage, water quality, flow velocity, backflow potential, and O&M practices are examples of operational factors. Others considered the nature and date of last failure (e.g., type, cause, severity), nature of maintenance operations (e.g., TV inspections, pipe cleaning, cathodic protection), nature and date of last repair (e.g., type, length), water quality and construction method as operational factors that affect the failure rate of water pipes (InfraGuide 2003). The specific explanation of each factor is shown in **Table 2.**

Table 2. Factors that contribute to water system deterioration (InfraGuide 2003)

Classification	Factor	Explanation
Physical	Pipe material	Pipes made from different materials fail in different ways.
	Pipe wall thickness	Corrosion will penetrate thinner walled pipe more quickly.
	Pipe age	Effects of pipe degradation become more apparent over time.
	Pipe vintage	Pipes made at a particular time and place may be more vulnerable to failure.
	Pipe diameter	Small diameter pipes are more susceptible to beam failure.
	Type of joints	Some types of joints have experienced premature failure (e.g., leadite)
	Thrust restraint	Inadequate restraint can increase longitudinal stresses.
	Pipe lining and coating	Lined and coated pipes are less susceptible to corrosion.
	Dissimilar metals	Dissimilar metals are susceptible to galvanic corrosion.
	Pipe installation	Poor installation practices can damage pipes, making them vulnerable to failure.
	Pipe manufacture	Defects in pipe walls produced by manufacturing errors can make pipes vulnerable to failure. This problem is most common in older pit cast pipes.
Environmental	Pipe bedding	Improper bedding may result in premature pipe failure.
	Trench backfill	Some backfill materials are corrosive or frost susceptible.
	Soil type	Some soils are corrosive; some soils experience significant volume changes in response to moisture changes, resulting in changes to pipe loading. Presence of hydrocarbons and solvents in soil may result in some pipe deterioration.
	Groundwater	Some groundwater is aggressive toward certain pipe materials.
	Climate	Climate influences frost penetration and soil moisture. Permafrost must be considered in the north.
	Pipe location	Migration of road salt into soil can increase the rate of corrosion.
	Disturbances	Underground disturbances in the immediate vicinity of an existing pipe can lead to actual damage or changes in the support and loading structure on the pipe.
	Stray electrical currents	Stray currents cause electrolytic corrosion.
	Seismic activity	Seismic activity can increase stresses on pipe and cause pressure surges.
Operational	Internal water pressure, transient pressure	Changes to internal water pressure will change stresses acting on the pipe.
	Leakage	Leakage erodes pipe bedding and increases soil moisture in the pipe zone.
	Water quality	Some water is aggressive, promoting corrosion
	Flow velocity	Rate of internal corrosion is greater in unlined dead-ended mains.
	Backflow potential	Cross connections with systems that do not contain potable water can contaminate water distribution system.
	O&M practices	Poor practices can compromise structural integrity and water quality.

Jon Røstum (2000) proposed another classification method which considered all the factors into 4 types: structural, external, internal, and maintenance. **Table 3** provides more details about it.

Table 3. Factors affecting structural deterioration of water distribution pipes (Jon Røstum, 2000)

Structural Variables	External/Environmental Variables	Internal Variables	Maintenance Variables
Location	Soil type	Water velocity	Date of failure
Diameter	Loading	Water pressure	Date of repair
Length	Groundwater	Water quality	Location of failure
Year of construction	Direct stray current	Water hammer	Type of failure
Pipe material	Bedding condition	Internal corrosion	Previous failure history
Joint method	Leakage rate		
Internal protection	Salt for de-icing of road		
External protection	Temperature		
Pressure class	External corrosion		
Wall thickness			
Laying depth			

3.2 Effect of Factors in Previous Papers

In this section, we list and describe factors that are commonly identified to have the greatest impact on pipe failure. Conclusions on these factors are also summarized.

Age and installation period

We can see the features of different failures in different phases of the installation process. After the installation has been done, compared with time, these features will become more reliant on the construction practice in each phase. The break rate in one construction phase might be higher than that in another phase (Mosevoll, 1994). Sometimes, compared with pipes that are relatively young, pipes that are older will be less prone to the effect of failures. For example, the walls of grey cast iron pipes are produced by newer casting methods, and for the same external loads, these thinner walls may cause more corrosion as well as more stress. It is only in the 1930s that we managed to use backfill to extend the lifetime of pipes. Time has witnessed the improvements of the jointing techniques, which make a higher degree of deflections at joints become possible. From 1950s to 1960s, when the number of houses just kept rising at a rapid rate, compared with the quality of the buildings, people often placed more emphasis on the quantity. During this time, houses of a rather bad quality as well as the poor skill of the construction workers could often be seen in the reports (Sundahl, 1997). According to the report written by Andreou et al. (1987), compared with pipes that failed at a later stage, pipes that failed in the initial stage usually have better performance. Besides, Wengström (1993) has discovered that we cannot rely on pipe records to find out the age dependency. This is also why he drew up the conclusion that it is possible for us to hide the age dependency via repairs. In other words, after being repaired for around four times, pipes will usually need to be taken out of the ground. sessing pipe

Corrosion

One of the causes of the need to replace a pipeline is corrosion as it can lead to degradation of pipes that are made of grey cast iron, ductile iron and steel (Mosevoll, 1994). The internal corrosion has great reliance on the features of the transported water (e.g. pH, alkalinity, bacteria and oxygen content) while the external corrosion is reliant on the surroundings of the pipe (e.g. soil characteristics, soil moisture, and aeration). However, Kumar Dey (2003) put forward the idea that when we are doing the prediction, we also need to take into consideration the external corrosion as its intensity will change according to the different conditions. In this regard, it is different from the internal corrosion.

Diameter

The idea that pipes with small diameters are most prone to failures can be found in a large number of literature works in the field. (Rajeev, 2003). Pipes with diameters that do not exceed or are equal to 200mm failure the most often. The strength of smaller pipes is usually are usually smaller, and their walls are also thinner. Also, they are usually constructed in a different way and their joints are usually not as reliable. These are the reasons why smaller pipe dimensions fail more frequently (Wengström, 1993). Another possible cause for this is the lower velocities in smaller pipes, which can cause the suspended materials in the water to settle, and this can make it easier for the bacteria to grow. (National Research Council. (2006)).

Pipe length

The length of pipes, regardless of which network they are in, varies from one to another. For long pipes (e.g. >1000m), external conditions including the condition of the soil as well as the traffic might be different depending on the pipe. Røstum et al. (1997) advised us to choose pipes that are 100m long so that the external conditions for the same pipe will be the same as well. Eisenbeis (1999) found out that the hazard function is of a similar proportion to the square root of length.

Pipe material

Cast iron pipes (i.e. grey cast iron and ductile iron pipes) are used in a great number of water works despite the fact that they have long been notorious for their high failure rates. This can also explain the increasing use of new materials such as PVC and PE in water networks. The material features of these pipes vary a great deal from each other, and analysis of different materials must be done separately. Recent studies have been focusing on analyzing pipes that are made of PVC and PE in a statistical way (Eisenbeis et al., 1999). The past few decades have witnessed great improvements in the techniques used in the manufacturing of different pipe material. One of the best examples showing this can be found in the improvement of the casting method used in the manufacturing of for grey iron pipes. At the beginning, pipes were cast in sand molds in a horizontal order, which makes the thickness of the wall become uneven. It is only after the introduction of the vertical casting technique that the production of walls of the same thickness became possible. This new technique has also helped to make the manufacturing of pipes with thinner walls become possible. The improvements obtained in the centrifugal casting methods has also helped to strengthen pipes and to help the walls to reach a higher consistency of thickness (WRc, 1998).

Seasonal variation

Winter is the season when most of the water distribution networks become the most prone to failures. Andreou (1986) is the first person to find out that it will be easier for pipes of a smaller diameter (those whose length do not exceed 8 inches) to break during winter. After analyzing five water networks in Sweden, Sundahl (1996) found out that among the temperature of air, precipitation and the depth of snow, only the former one would exert an effect on the break rate. In Trondheim, even though the coldness in winter has brought forth a huge amount of frost, the number of reported failures in summer time still overrode that for the winter season (Røstum, 1997). However, Sægrov et al. (1999) found out that the break rate in both summer time and winter time in the United Kingdom was rather high. As the clay soils during the summer season became increasingly drier and kept shrinking, the break rate also kept rising up, whereas during the winter season, usually there would be a great deal of frost, and this is one of the major causes of the high break rate. Another factor contributing to this is the thermal contraction effects. Other than this, it is also found that the mean temperature during the day as well as the amount of rainfall each year have also played a part in the annual break rate over a period of ten years. It is suggested that we ought to use the effects of the climate to find out the factors leading to the failures of pipes. However, since we do not have an idea as to how this factor change over time, it will be really difficult for us to use the effects of the climate as a tool to forecast future failures. In her research, Sundahl (1996) attempted to use a sinus curve to model the changes in the leakage in different seasons. The manager held the view that the change of the failing rate of pipes according to different seasons can offer us help to plan/organize the water network on a daily basis. However, when it comes to the calculation of the future needs for rehabilitation and for making priorities between pipes, the knowledge of the actual day of failure becomes less helpful.

Soil conditions

Soil conditions can not only exert an influence on the rate of external corrosion, but can also affect pipe degradation. In their research, Clark et al. (1982) tried to put pipes in corrosive soil environments and then analyzed their failing rate. They found out that how much of the pipe is laid in corrosive environments has no relation with its breaking rate. Malandain et al. (1998) tried using a geographic information system (GIS) to relate soil conditions to the failing rate for pipes in the water network in Lyon, France. In his analysis of the breaking rate of pipes, Eisenbeis (1994) used ground condition, (which is defined as the presence or absence of corrosive soil) as an explanatory variable.

Previous failures

The braking rate of pipes in the past can help a great deal in the forecasting of future failures. Andreou (1986) used the Cox proportional hazards model to analyze failures in the water network. It is only after the third failure that the failing rate stopped rising, with each failure, and yet the rate still remained to be a high one. The assumption is that at this phase, the pipes have entered a "rapid failing state". It is found out that failures happened in the past can exert a huge effect on the hazard function of the pipes. Eisenbeis (1994) has also spotted a similar pattern. Malaindain et al. (1999) has applied these findings from Andreou and Eisenbeis in a failing rate model. Goulter and Kanzemi (1988) made close observation of the temporal and spatial gathering of water-main breaks, which shows that it is highly likely that failures of a pipe in the past will lead to future failures in its surroundings. Approximately 60% of all of the subsequent failures happened during the first three months after the first failure. This has led us to believe that the damage brought by the repairing work is the culprit behind these subsequent breaks. Possible damages include the rise of pressure brought by pipe-refilling, the change of position of the ground during excavation, the back-filling procedure or the movement of weighty vehicles. Sundahl (1996, 1997) has also pointed out that maintenance work done on the network including repair and replacement after a failure can also lead to a higher failing rate.

Other factors that do not share any correlation with the repair work also play a role in the subsequent failures in the network. Pipes in the same place are usually of the same age, and the materials that they were made from, very often, are also the same. What's more, they are usually constructed and jointed together via the same method. Other than all these, it is also highly possible that both the external and internal factors that can lead to corrosion for these pipes are the same.

Nearby excavation

Excavation work done near the pipelines can exert a negative effect on the bedding conditions, which can cause the pipe to break. Researches conducted in the U.K. (WRc, 1998) indicated that work on closely related services (e.g. gas, electricity) can lead to pipe breaks.

The pressure in static water and the rise of pressure in a distribution system also play a role in pipe breaks. The rise of pressure is usually caused by the opening and closing of water and air valves while the network is under operations. These changes can be seen as one of the causes of break clustering. Andreou (1986) found that when it comes to modelling pipe breaks, it can be useful to take into account the effect of static pressure, but this factor is by no means of huge importance. When Clark et al. (1982) were modelling time to the first break, they used both the absolute pressure and the pressure differential (surge).

Land use

Land use (e.g. traffic zones, places of residence, and commercial areas) is used as a substitute for external loads on pipes. Eisenbeis (1997) used land use over the pipe (i.e. no traffic vs. heavy traffic), as a variable in break models.

Previous papers discussed a lot about the classifications and definitions of factors. However, the availability of factors in data are limited based on real dataset. The factors have higher availability are more likely to be considered in real models and effect more to failure prediction. The frequency of factors using in collected dataset will be discussed in later section.

3.3 Summary of Factors Considered in Different Failure Models

Table 4. Considered Factors Affecting Water Pipes Failure

Classification	Model	References	Response	Age	Length	Diameter	Installation Year	Temperature	Depth	Soil Type	Water Press	Freezing Index	Pipe thickness	Previous number of failure
Deterministic Models	Failure rate	Amarjit Singh (2012)	Failure rate				2							1
		Andreas Scheidegger (2017)	Average number of failure		2									
		Małgorzata Kutyłowskaa (2016)	Average number of failure		2	2								
		Andrew Wood (2009)	Number of failure			2								1
		Hossein Rezaei (2015)	Number of failure	1		1		1			1	1		
		Alex Francisque (2017)	Failure rate		2	2								1
		Katarzyne Pietrucbe (2015)	Number of failure	1	1		1		1	1		1		
		C.Vipulanandan (2012)	Number of failure			2								1

Table 4 (Cont.)

Classification	Model	References	Response	Age	Length	Diameter	Installation Year	Temperature	Depth	Soil Type	Water Press	Freezing Index	Pipe thickness	Previous number of failure
Probabilistic Models	Weibull proportional hazard model	E. Kimutai (2015)	Number of failure			2								1
		Yves Le gat (2000)	Failure rate		1									1
	Cox proportional model	H Shin (2016)	Number of failure	1		1		1			1	1		
	Weibull-Based Failure Models	Lindsay Jenkins (2014)	Average number of failure	1	1	1								
		Stefano Alvisi (2008)	Number of failure	1	1				1					
	Weibull Accelerated lifetime model	André Martins (2013)	Average number of failure		1	1	1							1
	Weibull/Exponential/Exponential model	Babacar Toumbou1 (2013)	Number of failure			2								1
	Weibull/Exponential model	Ben Ward (2016)	Number of failure			2								1

Table 4 (Cont.)

Classification	Model	References	Response	Age	Length	Diameter	Installation Year	Temperature	Depth	Soil Type	Water Press	Freezing Index	Pipe thickness	Previous number of failure
Probabilistic Models	Principal component regression	Zhiguang Niu (2017)	Failure rate		1									1
	Multiple regression model	Pengjun Yu (2013)	Number of failure	1	1				1					
		Mohamed Fahmy (2009)	Number of failure			1	1							
		Yong Wang (2009)	Failure rate		1									1
		Leila Dridi (2009)	Average number of failure	2		2								
		Ahmad Asnaashari (2013)	Number of failure	1	1		1			1				
		Kang Jing (2012)	Number of failure	1	1		1		1	1		1		
	Logistic regression	Boxall (2013)	Number of failure			2								1
	Non-linear regression	B. García-Mora (2015)	Number of failure			2								1

Table 4 (Cont.)

Classification	Model	References	Response	Age	Length	Diameter	Installation Year	Temperature	Depth	Soil Type	Water Press	Freezing Index	Pipe thickness	Previous number of failure
Probabilistic Models	Evolutionary polynomial regression	L. Berardi (2008)	Number of failure	1	1		1		1	1		1		
		D. A. Savic (2009)	Number of failure			2								1
		Seyed Farzad Karimian (2015)	Number of failure	1	1				1					
		Konstantinos Kakoudakis (2017)	Number of failure	1	1		1		1	1		1		
		Qiang Xu (2011)	Number of failure	1	1				1					
		Fulvio Boanoa (2015)	Number of failure	1		1		1			1	1		
	Bayesian method	Kleiner,Yehuda (2012)	Number of failure	1	1		1		1	1		1		
		G Kabir (2015)	Number of failure	1		1		1			1	1		
		Ángela Martínez-Codina (2015)	Number of failure			2								1

Table 4 (Cont.)

Classification	Model	References	Response	Age	Length	Diameter	Installation Year	Temperature	Depth	Soil Type	Water Press	Freezing Index	Pipe thickness	Previous number of failure
Stochastic Models	NHPP	Peter D. Rogers (2009)	Number of failure	1	1				1					
		T. Economou (2008)	Number of failure	1	1				1					
		T. Economou (2012)	Number of failure	1	1				1					
		Li Chik (2016)	Average number of failure	2		2								
		Fengfeng Li (2011)	Number of failure			2								1
		Yehuda Kleiner (2010)	Failure rate		1									1
	Poisson process	Theodoros Economou (2010)	Number of failure				1		1					2
	Linear extended Yule process	Yves Le Gat (2013)	Failure rate		1									1
		Li Chik (2016)	Average number of failure	2		2								

Table 4 (Cont.)

Classification	Model	References	Response	Age	Length	Diameter	Installation Year	Temperature	Depth	Soil Type	Water Press	Freezing Index	Pipe thickness	Previous number of failure
Artificial Intelligence/Machine Learningd	ANN	RaedJafar (2010)	Number of failure	1	1				1					
		M. Tabesh (2009)	Average number of failure		2	2								
		Richard Harvey (2014)	Number of failure	1	1				1					
		Libi P. (2016)	Average number of failure	2		2								
	Genetic programming	Qiang Xu (2011)	Number of failure	1	1				1					
		Wen-zhong Shi (2013)	Failure rate		1									1
	Fuzzy Clustering	Mahmut Aydogdu (2014)	Average number of failure		2	2								

Table 4 (Cont.)

Classification	Model	References	Response	Age	Length	Diameter	Installation Year	Temperature	Depth	Soil Type	Water Press	Freezing Index	Pipe thickness	Previous number of failure
Artificial Intelligence/Machine Learning	Fuzzy Clustering	Małgorzata Kutyłowska	Average number of failure		2	2								
	Dirichlet process mixture of hierarchical beta process model	Peng Li (2015)	Number of failure	1	1				1					
	Moran's I Ripley's K-statistic	Qiang Xu (2012)	Number of failure	1	1				1					
	Grey relational analysis(GRA)	Kang Jin	Number of failure			1		1			1	1		

Number 2 in the form refer to use as group whereas number 1 means it is a covariate in the models. Failure rate defined as λ which is determined from operational data using number of pipe failures in unit time interval divide average pipeline length in a time period and the observation time.

It can be seen from **Table 4** that diameter, length, and age are considered most frequently in the network failure models. Material mostly used as cohorts or groups in the models such as NHPP, Failure rate model, and Weibull distribution model. Diameter and length are easy to quantify and thus are often used as covariates. These features are mostly analyzed as covariates in failure models. Soil type is another common factor that was often included due to its availability. Although soil type is often shown to significantly affect pipe performance, in some of the cases like Berardi (2008), soil type is not found significant. Comparing to other factors, pipe material is usually considered as a cohort, and different models are developed for each material cohort (add those case). For response variable, number of failure account a large percentage.

3.4 Factor Distribution Analysis

In some cases, factors such as diameter and length are considered as covariates in the models while material is considered as cohorts. However, sometimes, especially for failure rate models, all the factors are considered as cohorts and the results of failure rate only apply to certain groups of pipes. Thus, it is difficult to reach conclusion about the whole network failure status based on many independent failure rates for different groups. The distribution of each factor is necessary to be considered because the weight of each diameter and material have different weight in the modeling results. Another important attribute for pipe is the installation year which reflects the variation in failures over time. Since data availability of pipe failure is limited, it would be useful to know the material installation year which has a large percentage in the whole network. The data and figures are presented in **Appendix II**.

.

Chapter 4 Summary of Key Findings in Previous Studies

Most failure prediction models, particularly deterministic, statistical and artificial intelligence models, characterize the relationship between network features and failures of a single network, and predict the number of failures or life span of the network based on these factors. In this section, we summarize the finding in collected papers and extract common conclusion about factors and models as insights for future fitting.

4.1 Factor Effect in Regression Models

Linear models usually had a similar response variable like failure rate or break time which could reflect the degree of failure directly. Most papers provided the result equation, so the coefficient of parameter is easy to obtain, By this way, the results of linear model is analyzed independently in this section. A Figure about linear model results is shown below.

Table 5. Results of Regression Models

References	Result Equation	Response Variable
Genevieve Pelletier (2003)	$\text{Log}_{10}\text{R} = 4.85 - 0.0206\text{A} + 0.000245\text{A}^2 + 0.00281\text{S} - 0.905\text{Log}_{10}\text{L} - 1.40\text{Log}_{10}\text{L}^2 - 1.40\text{Log}_{10}\text{S}$	Failure Rate
	$\text{Log}_{10}\text{R} = 1.83 - 0.911\text{Log}_{10}\text{L}$	Failure Rate
	$\text{Log}_{10}\text{R} = 2.69 - 0.898\text{Log}_{10}\text{L} - 0.745\text{Log}_{10}\text{A}$	Failure Rate
Pengjun Yu (2013)	$\text{R} = 2.096 - 4.4423\text{D} + 3.3571\text{D}^2 - 0.7292\text{D}^3$	Failure Rate
Kang Jing (2012)	$\text{Y} = -6000.741 - 1999.02\text{D} + 17318.428\text{H} + 450.949\text{P}^2$	Leakage Time
Boxall et al. (2013)	$\gamma(\text{D,L,A}) = 0.50247 - 0.00726\text{D} + 0.66252\log\text{L} - 0.03375\text{A} + 0.00016\text{A}^2$	Burst Rate

In the result equation, R refers to the rate of failure; A is age of pipes; D is diameter of pipes; S is soil type; P is water press in pipe; H is depth; L refers to length.

In the case by Pelletier (2003), the pipes of shorter lengths have higher annual break retes than those of longer lengths. The annual break rates of the 100m length of gray cast iron pipes with different diameters versus pipe age. In this network, 100 mm size pipe have the highset annual break rates compared to others for all ages. The 300 and 150 mm size pipe have similar annual break rates. For the network in case by Yu (2013), the models show a negative correlation between the failure rate and diameter when the pipe diameter is less than 100mm while the failure rate is rising when the pipe diameter is greater than 1000mm. So, the pipeline with diameter as 1000mm has the lowest failure rate value. Jing (2012) gives a result that depth has a positive relationship to leakage time and a negative one for diameter. The diameter limited in 50-250 mm. Boxall et al. (2013) indicates that the burst rate only applied in certain material. Diameter, length, age is involved in the equation.

The relationship between annual burst rate, length and diameter for cast iron and asbestos cement pipe groups for each of the two datasets are similar, with slight variation in the coefficient values. Once the models have been derived for a given company or region it is possible to make predictions for every combination of material, diameter, length and age of pipe. These can be used directly to inform investment decision making and planning, or to inform whole life cost decision support procedures and software. It is important to recognize that this kind of burst rate prediction is valid principally for the short term, from perhaps 1 to 5 years. The prediction for a pipe of a given age is for its burst rate in the next year.

4.2 Model Results Review

In this section, the conclusions in the collected papers are summarized and shown in **Appendix III.** Although each case has its own characteristics, the similar conclusion about model performance or factor effect could be common. The common parts in conclusions are extracted and shown in **Table 6**.

Table 6. Extracted Common Conclusions.

Category	Common Conclusions	References
Model	WE model may have a good prediction performance though it does not consider covariates.	Toumbou et al.(2013)
	Extending WE model to WEE model or developing a WE based proportional model would be feasible ways to improve the prediction accuracy.	Francis et al. (2014)
	However, too much covariates covered in proportional model may lead to overfitting.	Davis et al. (2007)
	For failure rate model, modeling the failure in group and individual pipe level would be a good way to avoid the inference that all covariates have the same impact on pipes.	Mahmut Aydogdu (2014)
	Cox-PHM and Poisson process both have their advantages in certain conditions.	García-Mora et al. (2015)
	Although in most situation, Poisson process is used as a comparison for other models or used for a small number of breaks prediction.	Asnaashari et al. (2013)
	Artificial Neural Network are useful for modeling complex problem that a large number of covariates are included and the correlation between covariates are uncertain.	Kabir et al. (2015a)
	Linear models usually have a lot of significant covariates and has accurate prediction when pipe failure history is known. Otherwise, short-term prediction would be more reliable.	Kutyłowskaa et al. (2016)
Material	Ductile pipe has a higher failure rate when the previous number of breaks is zero.	García-Mora et al. (2015)
	After first break, it will decrease the probability of failure, especially for ductile pipe with long length and small diameter.	Rezaei et al.(2015)
	PVC and AC pipes suffered more from cracking which may relate to covariates such as internal pressure, soil deflection and residual pressure.	Kleiner and Rajani (2008)
	Steel and grey cast iron suffered material corrosion which may relate to temperature and humidity	Wood and Lence (2009)
	Time-linear model fits better than time-exponential model for as asbestos cement (AC) and ductile iron (DI), PVC pipes usually has small number of failures and lack of recorded history because of near installation year. So, they can be good predicted by Poisson process.	Aydogdu and Firat (2014)
Diameter	Diameter is a common and efficient group for failure prediction.	Kimutai (2015)
	Smaller diameter (25-50mm) pipe more likely to get damage which may due to pressure fluctuation.	Kleiner and Rajani (2008)
	For pipe has high brittleness, like AC or PVC, failure rate is higher in winter than summer, but this covariate has strong correlation with pipe-laying depth which effect the temperature of pipes.	Martins et al.(2013)
	Generally, high pressure variation will increase failure rate.	M. Tabesh (2009)
	Age, diameter, length, material, buried depth and elevation of pipe were selected as the most critical factors.	Jenkins et al.(2014)
	Pipe diameter and age are the most sensitive factors in two datasets.	Martins et al.(2013)
	For linear, NOPNF has more important weight than in other models.	Karimian et al.(2015)
	Installation site has relationship with many factors.	Jenkins et al.(2014)
	The relationship between burst rate and diameter has been found to increase exponentially with decreasing diameters.	Achim et al. (2007)

Chapter 5 Summary and Conclusion

Unlike previous model review papers that mostly reviewed the model development or improvement in a period of time, this paper focuses on review and summary of contributing factors considered in the models and the associated effects of these factors. Specifically, the characteristics of all the collected cases are summarized to find the distribution and tendency of available networks data; the results of different fitting models are summarized to find common conclusions.

5.1 Conclusions

Based on the review, we reach the following conclusions.

- The distribution of case regions has shown that the United States has concentrated much on network deterioration issue. Prior to Year 2000, the case from Canada and Europe accounted for a majority of total number of cases in papers about failure models because Canada faced the failure problem earlier. However, the increasing number of cases in Asia and North America indicates that some other areas started facing and solving this global common issue. Recent papers applying to the cases in the U.S. would offer more references to future than those applying Canadian cases.
- The analysis has also shown that the number of pipe breaks and the number of pipe segments do not have high correlation. Thus, judging the severity of pipe deterioration based on the number of breaks is not a feasible way.
-

- The summary statistics of models used in the literature has shown the popularity of prediction models. Data driven methods has been used increasingly.

The following tables summarize the recommendations for model selection or validation of results for future studies.

Table 7. Preferred Record Period for Material

Material	Preferred Record Period
CI	All
PVC	After 1950
AC	1950-1970
Steel	After 2000
DI	1900-2000

Table 8. Related Conclusions for Covariates.

Covariates	Related Conclusions
Material	Mostly used as cohorts and the number of type is not necessary to be much
Age	Has negative correlation to failure.
Diameter	Diameter less than 250mm has a negative correlation with failure, while diameter of 1000 or above has a positive correlation with failure.
Length	Has a negative correlation with failure.
Buried depth	Has a positive correlation with failure.
Pipe inner pressure	Not enough conclusion.

Table 9. Preferred Condition for Models.

Models	Preferred Condition
WEE	Small number of covariates
Failure rate model	When the input and output are clear
ANN	With a large number of covariates.
Linear model	When pipe failure history is known; Short term prediction.
Cox-PHM and Poisson process	Use as comparison for models with covariate or non-covariates, respectively.

5.2 Limitation and Future Work

The sample size of collected cases is limited which leads to a limitation for the analysis correlations between model and network characteristics. Less than 20 cases offer the data of factor distributions and the summarized conclusions may have unknown application range, e.g. the model fitting may get influenced by network size. Thus, the conclusions of this paper are not accurate enough to be used as verification for future model fitting.

In addition, this paper only discussed the case information, characteristics distribution and model results separately. The link among them are not explored due to lack of time and data. So it would be a feasible direction to do more research on the characteristics identification in network failures.

References

Aydogdu, M., & Firat, M. (2015). Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM methods. Water resources management, 29(5), 1575-1590.

Alvisi, S., & Franchini, M. (2010). Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system. Civil Engineering and Environmental Systems, 27(1), 1-22.

Asnaashari, A., McBean, E. A., Gharabaghi, B., & Tutt, D. (2013). Forecasting watermain failure using artificial neural network modelling. Canadian Water Resources Journal, 38(1), 24-33.

AWWA. (2001) Dawn of the Replacement Era: Reinvesting in Drinking Water Infrastructure, AWWA Government Affairs Office, Washington, DC

Berardi, L., Giustolisi, O., Kapelan, Z., & Savic, D. A. (2008). Development of pipe deterioration models for water distribution systems using EPR. Journal of Hydroinformatics, 10(2), 113-126.

Boano, F., Scibetta, M., Ridolfi, L., & Giustolisi, O. (2015). Water distribution system modeling and optimization: a case study. Procedia Engineering, 119, 719-724.

Boxall, J. B., O'Hagan, A., Pooladsaz, S., Saul, A. J., & Unwin, D. M. (2007, June). Estimation of burst rates in water distribution mains. In Proceedings of the Institution of Civil Engineers-Water Management (Vol. 160, No. 2, pp. 73-82). London:[Published for the Institution of Civil Engineers by Thomas Telford Ltd.], c2004-.

Chik, L., Albrecht, D., & Kodikara, J. (2016). Estimation of the Short-Term Probability of Failure in Water Mains. Journal of Water Resources Planning and Management, 143(2), 0401607

Dey, P. K. (2003). Analytic hierarchy process analyzes risk of operating cross-country petroleum pipelines in India. Natural Hazards Review, 4(4), 213-221. Data driven water pipe failure prediction: A bayesian nonparametric approach. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 193-202). ACM.

Dridi, L., Mailhot, A., Parizeau, M., & Villeneuve, J. P. (2009). Multiobjective approach for pipe replacement based on bayesian inference of break model parameters. Journal of Water Resources Planning and Management, 135(5), 344-354.

Economou, T., Kapelan, Z., & Bailey, T. (2008). A zero-inflated Bayesian model for the prediction of water pipe bursts. In Water Distribution Systems Analysis 2008 (pp. 1-11).

Economou, T. (2010). Bayesian modelling of recurrent pipe failures in urban water systems using non-homogeneous Poisson processes with latent structure.

Economou, T., Kapelan, Z., & Bailey, T. C. (2012). On the prediction of underground water pipe failures: zero inflation and pipe-specific effects. Journal of Hydroinformatics, 14(4), 872-883.

Fahmy, M., & Moselhi, O. (2009). Forecasting the remaining useful life of cast iron water mains. Journal of performance of constructed facilities, 23(4), 269-275.

Francis, R. A., Guikema, S. D., & Henneman, L. (2014). Bayesian belief networks for predicting drinking water distribution system pipe breaks. Reliability Engineering & System Safety, 130, 1-11.

Francisque, A., Tesfamariam, S., Kabir, G., Haider, H., Reeder, A., & Sadiq, R. (2017). Water mains renewal planning framework for small to medium sized water utilities: a life cycle cost analysis approach. Urban Water Journal, 14(5), 493-501.

García-Mora, B., Debón, A., Santamaría, C., & Carrión, A. (2015). Modelling the failure risk for water supply networks with interval-censored data. Reliability Engineering & System Safety, 144, 311-318.

Harvey, R., McBean, E. A., & Gharabaghi, B. (2013). Predicting the timing of water main failure using artificial neural networks. Journal of Water Resources Planning and Management, 140(4), 425-434.

Jafar, R., Shahrour, I., & Juran, I. (2010). Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. Mathematical and Computer Modelling, 51(9), 1170-1180.

Jenkins, L., Gokhale, S., & McDonald, M. (2014). Comparison of Pipeline Failure Prediction Models for Water Distribution Networks with Uncertain and Limited Data. Journal of Pipeline Systems Engineering and Practice, 6(2), 04014012.

Jing, K., & Zhi-Hong, Z. (2012). Time prediction model for pipeline leakage based on grey relational analysis. Physics Procedia, 25, 2019-2024.

Kabir, G., Tesfamariam, S., Loeppky, J., & Sadiq, R. (2015). Integrating Bayesian Linear Regression with Ordered Weighted Averaging: Uncertainty Analysis for Predicting Water Main Failures. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 1(3), 04015007.

Kakoudakis, K., Behzadian, K., Farmani, R., & Butler, D. (2017). Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering. Urban Water Journal, 14(7), 737-742.

Karimian, S. F. (2015). Failure Rate Prediction Models of Water Distribution Networks (Doctoral dissertation, Concordia University).

Kleiner, Y., & Rajani, B. (2001). Comprehensive review of structural deterioration of water mains: statistical models. Urban water, 3(3), 131-150.Le Gat, Y. (2014). Extending the Yule process to model recurrent pipe failures in water supply networks. Urban Water Journal, 11(8), 617-630.

Kleiner, Y., & Rajani, B. (2010). I-WARP: Individual water main renewal planner. Drinking Water Engineering and Science, 3(1), 71.

Kleiner, Y., & Rajani, B. (2012). Comparison of four models to rank failure likelihood of individual pipes. Journal of Hydroinformatics, 14(3), 659-681.

Kimutai, E., Betrie, G., Brander, R., Sadiq, R., & Tesfamariam, S. (2015). Comparison of statistical models for predicting pipe failures: Illustrative example with the City of Calgary water main failure. Journal of Pipeline Systems Engineering and Practice, 6(4), 04015005.

Kutyłowska, M., & Orłowska-Szostak, M. (2016). Comparative analysis of water–pipe network deterioration–case study. Water Practice and Technology, 11(1), 148-156.

Kutylowska, M. (2015). Modelling of failure rate of water-pipe networks. Periodica Polytechnica. Civil Engineering, 59(1), 37.

Le Gat, Y., & Eisenbeis, P. (2000). Using maintenance records to forecast failures in water networks. Urban Water, 2(3), 173-181.

Li, F., Ma, L., Sun, Y., & Mathew, J. (2014). Group maintenance scheduling: a case study for a pipeline network. In Engineering Asset Management 2011 (pp. 163-177). Springer, London.

Lin, P., Zhang, B., Wang, Y., Li, Z., Li, B., Wang, Y., & Chen, F. (2015, October).

Martins, A., Leitão, J. P., & Amado, C. (2013). Comparative study of three stochastic models for prediction of pipe failures in water supply systems. Journal of Infrastructure Systems, 19(4), 442-450.

Markose, L. P., & Deka, P. C. (2016). ANN and ANFIS Modeling of Failure Trend Analysis in Urban Water Distribution Network. In Urban Hydrology, Watershed Management and Socio-Economic Aspects (pp. 255-264). Springer International Publishing.

Martínez-Codina, Á., Cueto-Felgueroso, L., Castillo, M., & Garrote, L. (2015). Use of pressure management to reduce the probability of pipe breaks: A Bayesian approach. Journal of Water Resources Planning and Management, 141(9), 04015010.

Mehrani, M. J., Jalili, M. R., Abadi, Y. T., Tashayoei, M. R., Hanifi, H., & Yusefpour, R. New equations for Prediction of pipe burst rate in water distribution networks.

National Research Council. (2006). Drinking water distribution systems: assessing and reducing risks. National Academies Press.

Niu, Z., Wang, C., Zhang, Y., Wei, X., & Gao, X. (2017). Leakage Rate Model of Urban Water Supply Networks Using Principal Component Regression Analysis. Transactions of Tianjin University, 1-10.

Pelletier, G., Mailhot, A., & Villeneuve, J. P. (2003). Modeling water pipe breaks—Three case studies. Journal of water resources planning and management, 129(2), 115-123Rajeev, P., Kodikara, J., Robert, D., Zeman, P., & Rajani, B. (2014). Factors contributing to large diameter water pipe failure. Water asset management international, 10(3), 9-14.

Pietrucha-Urbanik, K. (2015). Failure analysis and assessment on the exemplary water supply network. Engineering Failure Analysis, 57, 137-142.Vipulanandan, C., Qiao, W., & Hovsepian, H. (2012). Water Pipeline Failures in the Active Zone. Center for Innovative Grouting Materials and Technology (CIGMAT) 2012.

Rajani, B., & Kleiner, Y. (2001). Comprehensive review of structural deterioration of water mains: physically based models. Urban water, 3(3), 151-164.

Rezaei, H., Ryan, B., & Stoianov, I. (2015). Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. Procedia Engineering, 119, 253-262.

Rogers, P. D., & Grigg, N. S. (2009). Failure assessment modeling to prioritize water pipe renewal: two case studies. Journal of Infrastructure Systems, 15(3), 162-171.

Sargaonkar, A., Kamble, S., & Rao, R. (2013). Model study for rehabilitation planning of water supply network. Computers, Environment and Urban Systems, 39, 172-181.

Savic, D. A., Giustolisi, O., & Laucelli, D. (2009). Asset deterioration analysis using multi-utility data and multi-objective data mining. Journal of Hydroinformatics, 11(3-4), 211-224.

Singh, A., & Adachi, S. (2011). Expectation analysis of the probability of failure for water supply pipes. Journal of Pipeline Systems Engineering and Practice, 3(2), 36-46.

Shin, H., Kobayashi, K., Koo, J., & Do, M. (2016). Estimating burst probability of water pipelines with a competing hazard model. Journal of Hydroinformatics, 18(1), 126-135.

Shi, W. Z., Zhang, A. S., & Ho, O. K. (2013). Spatial analysis of water mains failure clusters and factors: a Hong Kong case study. Annals of GIS, 19(2), 89-97.

Tabesh, M., Soltani, J., Farmani, R., & Savic, D. (2009). Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. Journal of Hydroinformatics, 11(1), 1-17.

Toumbou, B., Villeneuve, J. P., Beardsell, G., & Duchesne, S. (2012). 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, 5(1), 04013006.

Scheidegger, A., Scholten, L., Maurer, M., & Reichert, P. (2013). Extension of pipe failure models to consider the absence of data from replaced pipes. Water research, 47(11), 3696-3705.

Ward, B., Selby, A., Gee, S., & Savic, D. (2017). Deterioration modelling of small-diameter water pipes under limited data availability. Urban Water Journal, 14(7), 743-749.

Wang, Y., Zayed, T., & Moselhi, O. (2009). Prediction models for annual break rates of water mains. Journal of performance of constructed facilities, 23(1), 47-54.

Wood, A., & Lence, B. J. (2009). Using water main break data to improve asset management for small and medium utilities: District of Maple Ridge, BC. Journal of Infrastructure Systems, 15(2), 111-119.

Xu, Q., Chen, Q., Li, W., & Ma, J. (2011). Pipe break prediction based on evolutionary data-driven methods with brief recorded data. Reliability Engineering & System Safety, 96(8), 942-948.

Xu, Q., Chen, Q., & Li, W. (2011). Application of genetic programming to modeling pipe failures in water distribution systems. Journal of Hydroinformatics, 13(3), 419-428.

Yu, P. J., Nie, J. X., Xu, G., Long, Z. H., Wang, Z. H., Tu, Y., ... & Chen, Y. L. (2013). Study of Failure Rate Model for a Large-Scale Water Supply Network in Southern China Based on Different Diameters. In Applied Mechanics and Materials (Vol. 260, pp. 1200-1205). Trans Tech Publications.

Appendix I

Table 10. Information Summary of Collected Cases

#	Title	References	Area	Country	City	Population	Network length (km)	Number of pipe segment	Number of pipe breaks
1	Modeling Water Pipe Breaks—Three Case Studies	Genevieve Pelletier (2003)	North America	Canada	Chicoutimi	64000	352	2096	1719
2			North America	Canada	Gatineau	93000	407	1554	1426
3			North America	Canada	Saint-Georges	20000	155	1806	279
4	Failure Assessment Modeling to Prioritize Water Pipe Renewal: Two Case Studie	Peter D. Rogers (2009)	North America	America	Colorado Spring	400000	2900	1471	1771
5			North America	America	Laramie Water	30000	330	3792	667
6	Development of pipe deterioration models for water distribution systems using EPR	L. Berardi (2008)	Europe	UK	48 water quality zones	19494	173	3669	354
7	Application of Artificial Neural Networks (ANN) to model the failure of urban water mains	RaedJafar (2010)	Europe	France		43000	162	4862	
8	A zero-inflated Bayesian models for the prediciton of water pipe bursts	T. Economou (2008)	North America	Canada	South-Central Ontario			1349	5425
9	On the prediction of underground water pipe failures: zero inflation and pipe-specific effects	T. Economou (2012)	Asia	New Zealand				532	175
10	Integrating Bayesian Linear Regression with Ordered Weighted Averaging: Uncertainty Analysis for Predicting Water Main Failures	G Kabir (2015)	North America	Canada	Calgary	1100000	4281km	49531	
11	Comparative Study of Three Stochastic Models for Prediction of Pipe Failures in Water Supply Systems	André Martins (2013)	Europe	Portugal			367km	11472	1912

Table 10 (Cont.)

#	Title	References	Area	Country	City	Population	Network length (km)	Number of pipe segment	Number of pipe breaks
12	Expectation Analysis of the Probability of Failure for Water Supply Pipes	Amarjit Singh (2012)	North America	US	Island of Oahu		3200		
13	General Model for Water-Distribution Pipe Breaks:	Babacar Toumbou (2013)	North America	Canada	City in Quebec,		185km	1152	
14	Comparison of Statistical Models for Predicting Pipe Failures: Illustrative Example with the City of Calgary Water Main Failure	E. Kimutai (2015)	North America	Canada	Calgary	149552			
15	Estimation of the Short-Term Probability of Failure in Water Mains	Li Chik (2016)	Asia	Australia	Melbourne		376		
16	Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling	M. Tabesh (2009)	Europe	Iran		93719	579		
17	I-WARP: Individual water mAin renewal planner	Yehuda Kleiner (2010)	North America		Western Canada		146.6	1091	
18	Extending the Yule process to model recurrent pipe failures in water supply networks	Yves Le Gat (2013)	North America	US	Mid-Atlantic		627.2	10581	10286
19	GROUP MAINTENANCE SCHEDULING: A CASE STUDY FOR A PIPELINE NETWORK	Fengfeng Li (2011)	Asia	Australia		50000	3640	66405	
20	Data Driven Water Pipe Failure Prediction: A Bayesian Nonparametric Approach	Peng Li (2015)	Asia	China	A	210000			
23	Bayesian Belief Networks for Predicting Drinking Water Distribution	Royce Fransis (2014)	North America	US	Mid-Atlantic	500000	403.4	2598	3686
24	Extension of pipe failure models to consider the absence of data from replaced pipes	Andreas Scheidegger (2017)	Europe	Swizerland	Lausanne				

Table 10 (Cont.)

#	Title	References	Area	Country	City	Population	Network length (km)	Number of pipe segment	Number of pipe breaks
25	Prediction Models for Annual Break Rates of Water Mains	Yong Wang (2009)	North America	Canada	Monton, Laval, Quebec		432km		
26	Estimation of Failure Rate in Water Distribution Network Using Fuzzy Clustering and LS-SVM Methods	Mahmut Aydogdu (2014)	North America		Malatya	550000	440km		1231
27	Comparative analysis of water–pipe network deterioration–case study	Małgorzata Kutyłowskaa(2016)	Europe	Poland	A	40000	10.7		269
28	Estimating burst probability of water pipelines with a competing hazard model	H Shin (2016)	Asia	South Korea			848.1km	26577	1405
29	Study of Failure Rate Model for a Large-scale Water Supply Network in Southern China Based on Different Diameters	Pengjun Yu (2013)	Asia	China	Southern China				
30	Forecasting watermain failure using artificial neural network modelling	Ahmad Asnaashari (2013)	North America	Canada	Toronto		784km		5422
31	Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system	Stefano Alvisi (2008)	North America	Italy	Ferrara	250000	2400km	23000	3472
32	Asset deterioration analysis using multi-utility data and multi-objective data mining	D. A. Savic (2009)	Europe	UK			189	477	89
33	Application of genetic programming to modeling pipe failures in water distribution systems	Qiang Xu (2011)	Asia	China	Beijing		3322.5km	313804	566
34	Spatial analysis of water mains failure clusters and factors: A Hong Kong case study	Wen-zhong Shi (2013)	Asia	China	Hong Kong		643km	84127	
35	Modelling of Failure Rate of Water-pipe Networks	Małgorzata Kutyłowska (2012)	Europe	Poland	A		16km		

Table 10 (Cont.)

#	Title	References	Area	Country	City	Population	Network length (km)	Number of pipe segment	Number of pipe breaks
36	Using Water Main Break Data to Improve Asset Management for Small and Medium Utilities: District of Maple Ridge, B.C.	Andrew Wood (2009)	North America	Canada	Maple Ridge	6000	43.55km		54
37	Water distribution system modeling and optimization: a case study	Fulvio Boanoa (2015)				50000	170km		
38	Time Prediction Model for Pipeline Leakage Based on Grey Relational Analysis	Kang Jing (2012)	Asia	China	North China				
40	Forecasting the Remaining Useful Life of Cast Iron Water Mains	Mohamed Fahmy (2009)	North America	Canada, USA		150000	221		
41	Multiobjective Approach for Pipe Replacement Based on Bayesian Inference of Break Model Parameters	Leila Dridi (2009)							
42	Predicting the Timing of Water Main Failure Using Artificial Neural Networks	Richard Harvey (2014)	North America	Canada	Greater Toronto Area	5500000	5850km	6346	9918
43	Comparison of Pipeline Failure Prediction Models for Water Distribution Networks with Uncertain and Limited Data	Lindsay Jenkins (2014)	North America	USA	southeastern	600000	4800km		
44	Leakage Rate Model of Urban Water Supply Networks Using Principal Component Regression Analysis	Zhiguang Niu (2017)	Asia	China	Tianjin	15.17 mi	5000km		
45	Deterioration modelling of small-diameter water pipes under limited data availability	Ben Ward (2016)	Europe			1.5 mi	15000	800000	60827
46	ANN and ANFIS Modeling of Failure Trend Analysis in Urban Water Distribution NetworkANN and ANFIS Modeling of Failure Trend Analysis in Urban Water Distribution Network	Markose and Deka (2016)	Asia	Indian	Trivandrum	137714			
47	Time Prediction Model for Pipeline Leakage Based on Grey Relational Analysis*	Kang Jing (2012)							

Table 10 (Cont.)

#	Title	References	Area	Country	City	Population	Network length (km)	Number of pipe segment	Number of pipe breaks
48	Model study for rehabilitation planning of water supply network	Aabha Sargaonkar (2012)							
49	Using maintenance records to forecast failures in water networks	Yves Le gat (2000)	Europe	France	Charente-Maritime		1243km	1212	735
50	Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering	Konstantinos Kakoudakis (2017)	Europe	UK					
53	Estimation of burst rates in water distribution mains	Boxall (2013)	Europe	UK				36000	4335
54	Failure Rate Prediction Models of Water Distribution Networks	Seyed Farzad Karimian (2015)	Asia	Qatar	Montreal	1.8mi	5045km	125828	22735
55	New equations for Prediction of pipe burst rate in water distribution networks	Mohammad Javad Mehrani (2015)	Asia		Tehran				
56	Comparison of four models to rank failure likelihood of individual pipes	Kleiner, Yehuda (2012)	North America		A(CI)			1091	
57	Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks	Hossein Rezaei (2015)	Europe	UK		100000	1090	5427	
58	Modelling the failure risk for water supply networks with interval-censored data	B. García-Mora (2015)	Europe	Spain	Mediterranean			25026	1487

Appendix II

Table 11. Diameter Percentage in Cases.

	1982-2003	1982-2003	1982-2003	1993-2005	1993-2003	2000-2006
0-100	2	3	4	5	2.8	50
100-200	82	66	65	30	57	28
200-300	10	25	16	25	19	13
300-400	5	5	15	16	22.4	7
400-600	0	0	0	23	0	2

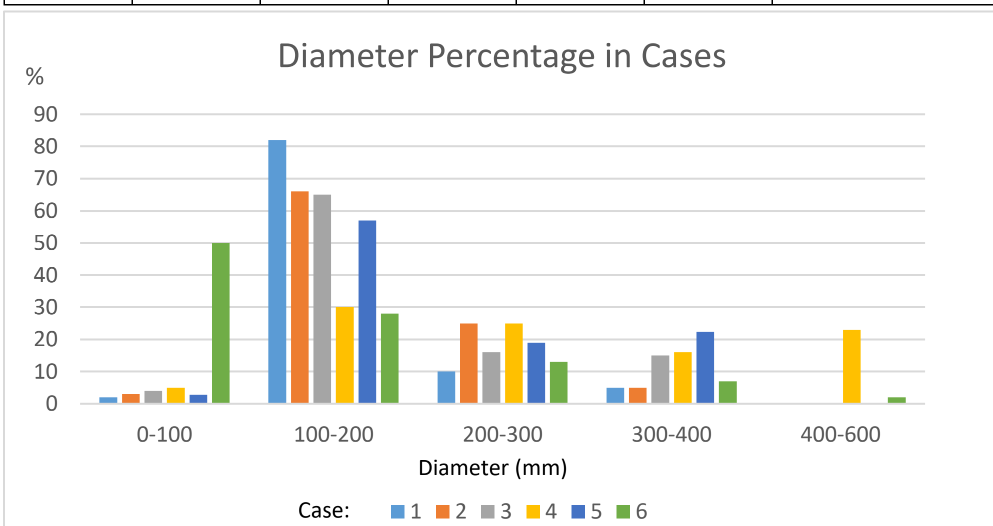

Figure 2. Distribution of Diameters for Each Available Case

Table 12. Installed Pipes Percentage in Cases

1	1950-2013	1950-2013	1982-2003	1982-2003	1982-2003	2000-2006
<1945	0	0	3	0	0	3
1945-1960	46	0	25	20	10	13
1961-1975	9	18	30	45	30	12
1976-1996	19	24	35	35	60	34
1996-2010	26	56	0	0	0	38.5

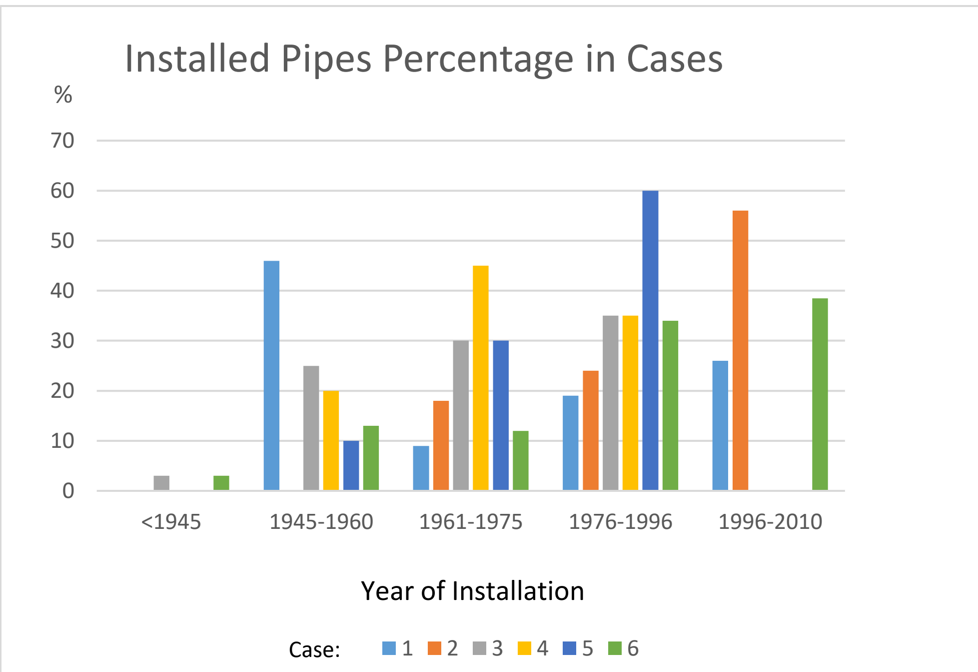

Figure 3. Installed Pipes Percentage in Cases

Table 13. Material Percentage in Cases.

	1940-2010	1972-2015	1992-2003	1992-2003	1993-2003	1995-2005	1999-2012	1999-2012	2000-2006	2000-2006	2003-2013	2006-2012
CI	20	56.5	35	44	20	15	56.6	64.1	2.9	15	69	55
DI	23	26.6	42	35	25	40	0	2	0	0	5	0
PVC	57	5.5	17	25	54	3.2	8.4	7.4	0	17	7	27.4
AC	0	10.5	0	0	0	0	1.8	0.6	31	55	10	1.8
PE	0	0	0	0	0	0	17.2	24	34	0	3	0
Steel	0	0	0	0	0	1.9	8	16	0	15.6	0	0

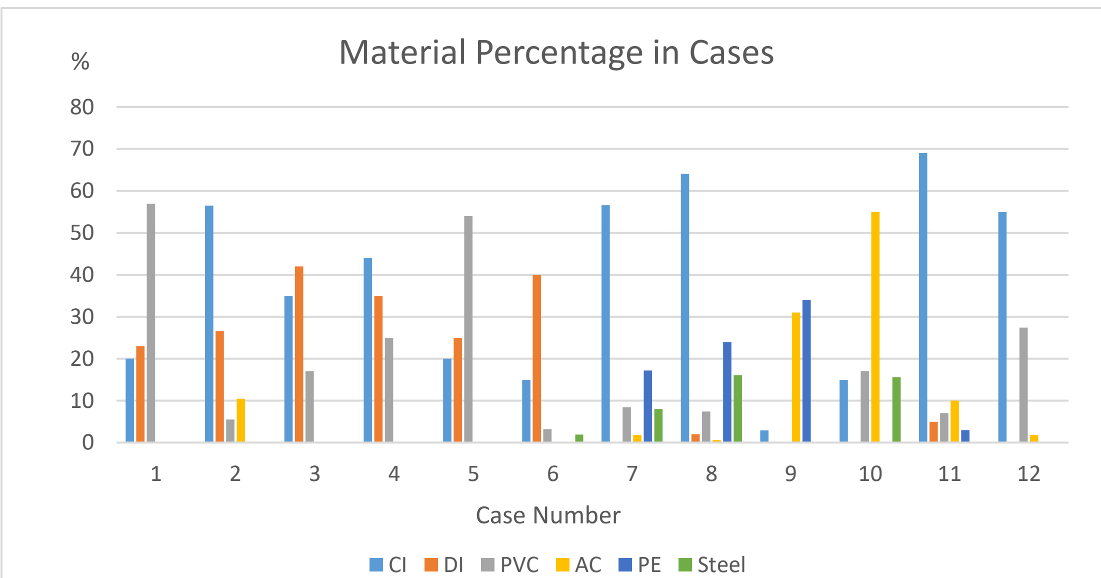

Figure 4. Material Percentage in Cases

Appendix III

Table 13. Model Results Review.

References	Model	Main Conclusions	Type
Scheidegger et al.(2017)	WE	It is obvious that the failure rate of the first-generation ductile pipes is higher	Material
Toumbou et al.(2013)	WE	The WEE mode is not affected by the covariates	Model Comparison
		The effect of pipe diameter grouping is more useful in long term failure prediction	Diameter
Davis et al. (2007)	WE	For pipes made of PVC, the time to brittle fracture for pipes with internal defects are caused by internal pressure, soil deflection and residual stress	Material
Francis et al. (2014)	Data driven	Population density cannot be used to find the relation between pipe age and intensity of water due to its lack of accuracy	Population
García-Mora et al. (2015)	Data driven	Long and small pipes made of ductile cast material will not break easily when they are put under sidewalks	Length, Diameter, Material
Asnaashari et al. (2013)	Failure rate	Both the CP and the CML programs can help to decrease the failure rate	Internal protection
Kutyłowskaa et al. (2016)	Failure rate	Change in the pressure might be one of the causes of the damage of small pipes (25-50mm)	Diameter
		Grey cast iron can be influenced by corrosion; Pipes made of AC or plastic will only be affected by cracks; Steel is exempt from the harm of material corrosion.	Material
		Pipes that are not laid deep into the ground are more likely to break in winter time than in summer time	Temperature
Rezaei et al.(2015)	Failure rate	Change in pressure can lead to failure of the pipe	Pressure
Kleiner and Rajani (2008)	Failure rate	Covariates at both group and pipe levels are analyzed so that the inference that all covariates will exert the same influence on pipes can be avoided	Model Comparison
Wood and Lence (2009)	Failure rate	The time-linear models can help to make the results of the analysis of pipe material groups become more accurate	Material
Aydogdu and Firat (2014)	Failure rate	Pipes with a diameter of 110cm, pipes that are 0-200m long, and pipes aging from 15 years to 20 years are the easiest to break	Length, Diameter, Age
Kimutai (2015)	Weibull Proportioanl Hazard Model; Cox Proportional Hazard Model	Thanks to its accurate estimation of the number of failures, compared with Cox-PHM and Poisson process, WPHM does a much better job at predicting the failing rate of metallic pipes	Model Comparison
		As the failing speed of the pipes becomes increasingly faster, the forecasting made via Cox-PHM becomes less and less accurate	Model Comparison
		Cox-PHM is a better option for the forecasting of young systems.	
		PM is a better option for the forecasting of PVC pipes	Material

Table 13 (Cont.)

References	Model	Main Conclusions	Type
Jenkins et al.(2014)	Weibull Proportioanl Hazard Model; Cox Proportional Hazard Model	If we reduce the number of explanatory variables, then it will become less likely for us to overfill a model	Size
		It can be difficult for us to learn more about the uncertain length of segment pipe from known data	Length
Karimian et al.(2015)	Evolutionary Polynomial Regression (EPR)	Length, diameter, age, material,elevation and the buried depth of pipes were chosen as the most important factors	Length, Diameter, Material
		Among all the factors, age and the diameter of pipes are the most sensitive ones in two of the data sets	Diameter, Age
Achim et al. (2007)	Artificial Neural Network (ANN)	ANNs can help a great deal in the modeling of sophisticated problems and these models can deal with all the effects brought by a wide range of input variables	Model Comparison
		Both time-dependent and other supplementary factors will be incorporated in the analysis for this model	Model Comparison
Kabir et al. (2015a)	Linear Regression Model	CI pipes are more sensitive to the resistance of soil while the DI pipes are more sensitive to the soil corrosivity index	Material
Martins et al.(2013)	linear extension of the Yule process (LEYP); Weibull accelerated lifetime model (WALM)	Neither the linear-extended Yule process nor the Weibull accelerated lifetime model can affect the avoidable breaks	Model Comparison
		The number of past breaks are the priority for both LEYP and WALM	Historical Failure
		Results shown by the other two models are slightly better than the Poisson results	Model Comparison
		Without the effect of past breaks, both LEYP and WALM would perform in as similar way as the Poisson process does	Model Comparison
		Repair work can make a pipe become more prone to breaks	Historical Failure
		Under the circumstances when only a small number of variables are available, the Poisson process can become very good at forecasting the failing rate	Model Comparison
		The shorter the maintenance records are, the better the forecasting done by LEYP and WALM will be	Model Comparison
Le Gat et al.(2013)	Linear	Time-dependent factors including implementation of pipe protection measures, changes in the traffic in the road, the time of frost as well as rainy weather sequence can all be used as references	Other Effect
		The thickness of the walls of pipes are closely related to the break rate	Thickness

Table 13 (Cont.)

References	Model	Main Conclusions	Type
Boxall (2013)	Linear	The reduce of diameters can help to strengthen the correlation between the burst rate and the diameter	Diameter
		Very little changes were found in the correlation between length and the burst rate	Length
		The correlation between age and the burst rate is different if we analyze it in different ways	Age
		Without a dependable age relationship, we would not be able to make long-term forecasting of burst rates.	Model Comparison
Wang et al. (2010)	Linear	The diameter and age of a pipe are the factor that can exert the biggest effect on the condition of the pipe	Age, Diameter
		Since the recharge of electricity, the depth of the trench and the number of roads share no relation to the condition of a pipe, they were not takne into consideration in the final analysis	Environment
Kleiner, Yehuda (2012)	NHPP	The number of past breaks, length as well as the age of the covariates can become very important statistics when the NHPP based model is in use	Age, Length, Historical Failure
		Compared with ductile iron pipes, cast iron mains are more prone to the effects of factor related to the climate	Material
C.Vipulanandan (2012)	Genetic Programming;EPR	Models used for big cities ought to be different from those used for smaller cities	Network Size
Peter D. Rogers (2009)	multiple-criteria decision analysis (MCDA)	Three failures happened in the past will be needed if we are using NHPP to do a single forecasting of pipe break	Model Comparison
L. Berardi (2008)	Evolutionary polynomial regression	Compare with pipes with a larger diameter, when the external pressure becomes really strong, smaller pipes are easier to break	Diameter
T. Economou (2008)	zero-inflated NHPP	Compared with the general NHPP, the zero-inflated version of the NHPP is more suitable for the data and the results it provides is of a slightly higher accuracy even though its performance is still worse than the general one	Model Comparison
T. Economou (2012)	NHPP	For pipes that remained not to break while being watched, the use of the Zero-inflated NHPP will give us a more accurate forecasting	Model Comparison
G Kabir (2015)	Bayesian Linear regression	The mean response forecasting made by the Bayesian regression models is no different from that made by the normal regression model, but predicted response made by the former one is better	Model Comparison

Table 13 (Cont.)

References	Model	Main Conclusions	Type
M. Tabesh (2009)	Artificial neural network	When it comes to the evaluation of the mechanical reliability (availability) values, the ANN pipe failure rate model does a much better job than the Adaptive NeuroFuzzy Inference System (ANFIS).	Model Comparison
Yehuda Kleiner (2010)	NHPP	For pipes with nearly no failures in the past, the aggregated total number of failures per pipe given by the NHPP was over estimated, while the forecasting made by the same model for pipes have failed for many times in the past was underestimated	Model Comparison
Fengfeng Li (2011)	Dirichlet process mixture of hierarchical beta process model	Pipes whose predicted likelihood of breaks in the future do not exceed 0.1 would not be included in future analysis	Model Comparison
Yong Wang (2009)	Multiple Regression model	Short pipes that have broken for more times in a year do not necessarily have more failures than long pipes do	Length
Mahmut Aydogdu (2014)	Fuzzy Clustering; Leaset square support vector machine method (LS-SVM)	LSSVM model results for the sub-regions defined by clustering analysis are better and that the clustering analysis can help to improve the performance of the estimation model and to provide a better result	Model Comparison