**The Thesis Committee for Mohamed Talha Rifaai**
**Certifies that this is the approved version of the following Thesis:**

# Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems

**APPROVED BY**
**SUPERVISING COMMITTEE:**

Polina Sela, Supervisor

Salvatore Salamone

# Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems

**by**

**Mohamed Talha Rifaai**

**Thesis**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Master of Science in Engineering**

**The University of Texas at Austin**

**August 2020**

# Dedication

To my family without whom this chapter of my life would not have been possible.

# Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Dr. Lina Sela for the continuous support during this journey, for the enlightening discussions, and for her patience and guidance. I truly appreciate making herself available more often than not and for always helping me keep a positive and uplifting outlook on my work. My gratitude also goes to Dr. Ahmed Abokifa who was closely involved in this research project. I am very grateful for his assistance, guidance, and supportive collaboration.

Also, I would like to thank Dr. Salamone Salvatore for being too kind and accepting to be a second reader for my thesis. I also extend my gratitude to all of the outstanding faculty and students at the Cockrell School of Engineering at The University of Texas at Austin. It has been truly inspiring and enriching to mingle with forward-looking and outstanding people. I am happy to carry over the energy with me in my future endeavors.

I would certainly like to gratefully acknowledge the U.S. Department of State and Amideast for funding and administering my Fulbright scholarship and making my deeply enriching experience in the U.S. possible and seamless. I also specifically thank my program officer, Emmanuel D. Pimental, and program assistant, Lara Al-Jaibaji, for their support throughout the program.

Last but not least, I feel forever privileged and grateful for the unconditional support of my parents Souad Naciri and Ahmed Nawfl Rifaai, my sister Salma Rifaai, my uncle M'roine Rifaai, and all my extended family members and dear friends.

# Abstract

## Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems

Mohamed Talha Rifaai, MSE

The University of Texas at Austin, 2020

Supervisor: Polina Sela

With an increasingly aging water infrastructure, decision makers are directing their attention to better ways to model and predict asset failure. Failure modeling is a discipline that addresses complex deterioration processes to better inform asset management practices. These deterioration processes and contributing factors have been addressed using a variety of models. One statistical model that has been explored in this thesis is the logistic regression model. The proposed approach consisted of developing a logistic regression model to estimate pipe-level failure probabilities in a flexible time interval. The approach further used the probabilities to estimate a Mean Time to Failure and assign pipe condition scores according to a methodology suggested by Opila and Attoh-Okine [1]. This thesis contributes with a practical and systematic methodology to capitalize on failure records and generate actionable failure probabilities and condition scores to integrate in asset prioritization strategies.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

## INTRODUCTION

In the 2017 Infrastructure Report Card, the American Society of Civil Engineers qualified the drinking water infrastructure in the United States as poor and at risk. In fact, the water infrastructure in the country suffers a poor condition with 240,000 main breaks every year wasting more than two trillion gallons of water [2]. In the face of such deterioration, water utilities struggle to keep pace with repair orders with little financial means and support at hand. To rationalize the use of resources, asset managers need to prioritize assets based on evaluating their condition. Condition assessment in turn can be carried out by inspection or by identifying and characterizing the influence of deterioration factors to estimate a network condition. Although ideal, inspection requires large resources and is limited in practice. Water utilities therefore need to characterize deterioration by analyzing past information to anticipate future failure. However, water utilities have only recently started to diligently record their repair work orders which provides only limited information about individual pipes. As a result, water utilities use a combination of inspection and analysis methods to determine pipe condition, discern trends, and identify priorities.

As failure data is becoming increasingly available, researchers have grown interested in applying physical and statistical methods to discern pipe deterioration patterns and predict failure. While physical models are typically complex and generally need prohibitive information for individual pipes, statistical models have provided a more practical alternative to address asset prioritization on a larger scale. Statistical models are

generally classified into deterministic and probabilistic models [3]. Amongst available probabilistic models, logistic regression has been a classic framework to estimate the probability of failure for each pipe. Several studies employed the framework to assess the predictability of pipe failures using classification metrics to measure the performance of logistic regression models. However, little research has explored applying logistic regression on a flexible time interval and using the outcome probabilities of a logistic regression model to estimate the metric of Mean Time to Failure (MTF). Estimating the MTF provides additional information to use both in failure prediction and condition assessment of the pipes.

## THESIS OBJECTIVES

This thesis aims at providing a practical framework to assess pipe condition based on pipe deterioration factors and to evaluate the predictability of failures using a logistic regression model. The specific objectives of the present work are as follows:

- Provide a review of relevant literature related to pipe deterioration and modeling
- Provide a systematic methodology for pipe condition assessment and failure prediction
- Evaluate the predictive performance of a logistic regression model
- Discern actionable insights for a failure history dataset with limited quality and size
- Evaluate a practical and flexible asset prioritization methodology that accounts for a water utility's rehabilitation strategy

## THESIS CONTRIBUTIONS

The contributions of the present thesis are as follows:

- Developing a logistic regression framework at an individual pipe level with a flexible prediction time interval
- Combining the reliability metric of Mean Time to Failure with the logistic regression outcome to provide estimates of the expected time between failures for each pipe.
- Using a discount rate curve to assign condition scores to individual pipes which can serve as an input to the city's asset management.

## THESIS ORGANIZATION

To address the above-mentioned objectives, the thesis is organized in four chapters. The present Chapter 1 introduces the scope, objectives, and contributions of the thesis. Chapter 2 describes major published research work related to asset deterioration factors and pipe failure models. Chapter 3 presents the proposed pipe condition assessment and failure prediction methods and results. Finally, Chapter 4 concludes with a discussion of results and opportunities for future research.

# Chapter 2: Literature review

## INTRODUCTION

Researchers have long been interested in modeling pipe deterioration processes. Interest in pipe deterioration is certainly as old as the earliest wooden pipes installed two to three centuries ago. Generations of new pipe materials revolutionized the pipe industry, and the structural strength of pipes has been tremendously improved. Although drastically extended, life expectancy of a water distribution system is limited, and a pipe is due to deteriorate over time.

Deterioration factors have been studied both separately and collectively to evince the nature of their influence. While the fundamental causes for deterioration and failure are known to be both mechanical and chemical processes that deplete a pipe's structural capacity, the complexity of the relationship between those processes and a pipe's internal and external environment quickly complicates the task of deterioration modeling. The following section presents some factors and related findings that are commonly discussed in literature. Later, major published models that attempted to characterize deterioration and failure processes are also presented.

## FACTORS CONTRIBUTING TO PIPE DETERIORATION

Pipe deterioration involves complex processes that researchers are still working on thoroughly specifying. A starting point is to determine the underlying causes and factors that can trigger and aggravate deterioration and potentially lead to pipe failure.

Pipe failure is a direct result of applied forces exceeding the structural capacity of a pipe. Applied forces can be external and influenced by the environment of the pipe or internal resulting from interaction with supply water. In fact, applying increasingly higher loads on any pipe material translates as increasing stress and strain that can reach a

fatigue or yield point. Past that point, the material loses its elasticity and can eventually reach a fracture point where a material can no longer withstand the load, and disruption happens. The reaction of a pipe material depends on its structural properties which can decline due to the influence of chemical and physical factors relative to the context where the pipe is installed.

Factors influencing pipe deterioration can be classified as pipe-intrinsic, environmental, or operational. Figure 1 shows common factors within these categories [4]. Some factors are typically addressed in the literature as presented in the following section.

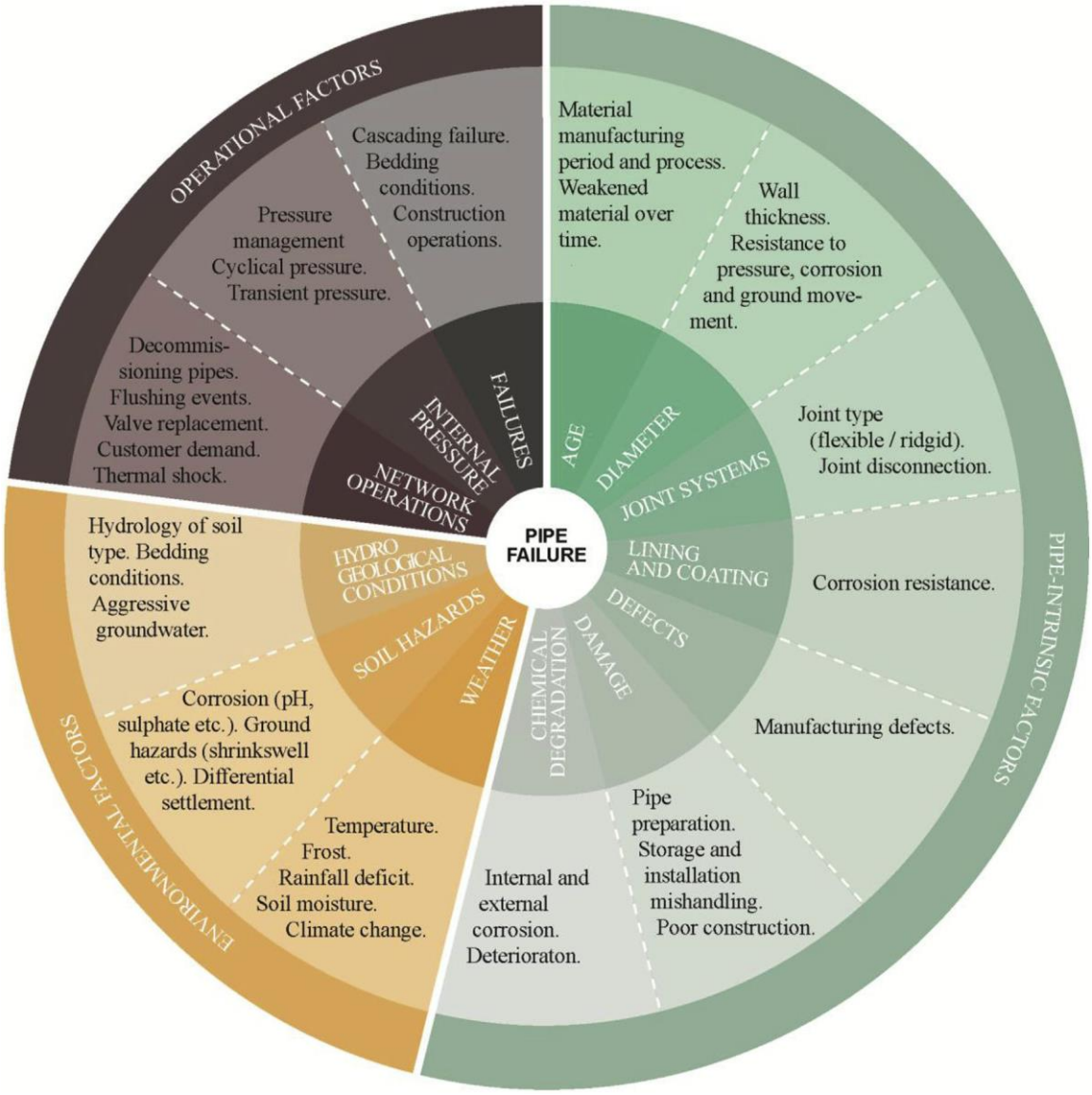


Figure 1.Categories of factors affecting pipe failure

*Note.* Reprinted from “Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks”, by Barton et al. [4]

**Pipe material**

Structural capacity depends first and foremost upon material properties of the pipe. The use of different materials to manufacture pipes has evolved along the years based on available practices and processes. In water distribution networks in the USA and Canada, materials typically found are listed in Table 1 based on a survey of 308 water utilities across both countries [5]. The table shows that a total of 91% of the network consists of CI, DI, PVC, and AC pipes which makes them the main materials presently in operation. These materials have been extensively analyzed and discussed in the literature discussing factors that influence their deterioration.

Table 1: Pipe material types and their fraction of total surveyed network length

| Pipe material | Description | Fraction of total surveyed network length |
|---|---|---|
| CI | Cast Iron | 28% |
| DI | Ductile Iron | 28% |
| PVC | Polyvinyl Chloride | 22% |
| AC | Asbestos Cement | 13% |
| CSC | Concrete Steel Cylinder | 3% |
| Steel | Steel | 3% |
| HDPE | High Density Polyethylene | 0.5% |
| PVCO | Molecularly Oriented PVC | 0.05% |

*Note.* Adapted from "Water Main Break Rates In the USA and Canada: A Comprehensive Study", by Folkman [5]

As described by Rajani & Kleiner [3], iron pipes have been manufactured since the 1880s by pouring molten grey cast iron into a vertical mould. These pipes, named pit cast iron pipes, have been used until the 1930s. In the 1920s, spun cast iron pipes were introduced as an alternative. These pipes were cast horizontally and spun while external

cooling was applied with water. This process provided uniformity and resulted in better structural properties. In 1948, ductile iron was introduced and slowly became popularized for its advantageous properties until it completely replaced cast iron pipe production by 1982. For iron pipes, electro-chemical corrosion is determined as the main cause of failure by formation of corrosion pits. This corrosion is further accelerated by a pipe environment (soil properties, water chemical composition, etc.) and can eventually lead to a pipe break.

As an alternative to iron pipes, AC pipes were introduced and gained currency in the 1950s and 1960s. These pipes offered better operational performance due to lower friction, lower manufacturing costs, and better resistance to corrosion. However, the material is less ductile than DI and therefore less resistant to soil movement [4]. Also, corrosive soils containing acids, alkalis, or sulphates can trigger chemical processes that infiltrate and form products that are detrimental to the microstructure of the material. Acid soils can also corrode reinforcement wiring in prestressed and reinforced concrete, thus diminishing its structural strength [6]. As a result of corrosive soils and mechanical factors, AC pipes can fail through circumferential or longitudinal breaks, joint failure, or chemical degradation [4].

In contrast with iron and AC pipes, PVC pipes provide significantly higher resistance to corrosion, low manufacturing costs, and an ease of installation. These benefits made PVC pipes popularized in the 1970s, and their production has kept rising due to improved manufacturing processes [4]. Despite their highly resistant properties, PVC pipes can still deteriorate through plasticizers' biodegradation, oxidation, and mechanical exertion [6]. In particular, cyclical soil movement and pressure fatigue can cause joint failure or longitudinal fracture in these pipes [4].

Considering this structural difference among pipe materials, researchers have extensively analyzed the influence of a material type on pipe failure rates. In fact, many studies have reported a significant importance of pipe material as a covariate in failure assessment and prediction models [7–9].

**Pipe diameter**

Several studies have suggested that pipe diameter is an important factor in pipe deterioration and have generally inferred an inverse relationship with failure rate [8,10–12]. Bruaset et al. [13] also indicated a negative correlation with both the number of recurring failures and the total number of failures. Also, pipe diameters between 100mm and 200mm were identified as having the worst failure rates. However, most pipes analyzed by the authors were of small diameters. In fact, pipe diameters of 8 inches and less make up two thirds of water mains in many water distribution systems in the USA and Canada while 10 to 12 inches make up an additional 18% of all pipes [5]. This uneven distribution results in a less statistically significant correlation between failure rates and pipe diameter. However, a plausible explanation for smaller diameter pipes having higher failure rates is a reduced structural capacity due to thinner walls and lower joint reliability [13,14]. Also, pipe diameter can have a different influence on deterioration depending on pipe material. In fact, Wang et al. [12] noted for example that pipe diameter had more effect on failure for CI pipes compared to DI pipes.

**Pipe length**

Pipe length is another factor that can influence failure rates. Many studies have only used it as a denominator to failure rates, thus evaluating failure risk per unit of

length [8,11,14]. Evaluating failure rates per unit of length without measuring its effect typically assumes length having a uniform and proportional contribution to failure rates.

However, several studies have suggested a clear effect of pipe length on failure probability. While some suggested that failure rates per unit of length decrease with increasing pipe lengths [12], several others have argued that longer pipes are more vulnerable. A possible reason for longer pipes being riskier could be that they are exposed to more varying environmental and operational factors like traffic load, pressure transients and bedding conditions [9,10]. In particular, Boulos et al. [15] found that water distribution systems with pipe lengths shorter than 2,000 ft are less vulnerable to pressure transients. The suggested reason was that, in shorter pipes, pressure waves are met sooner with junctions, tanks, and similar obstacles that cause wave reflections. These reflections tend to counter the initial effect of a transient and limit its impact.

**Pipe age**

An aging water infrastructure has been widely depicted as the main factor leading to increased pipe failure rates in recent years, but the influence of pipe age can be more complex. In reliability engineering, a common assumption when it comes to a pipe's life cycle is that it follows the "bathtub curve" [3]. The first phase, also called "burn-in" phase, is characterized by a steep decline from a high failure rate. These early failures are usually due to manufacturing defects or construction practices where the pipe's structural integrity is severely undermined and will not withstand normal levels of loads and stress. Once a pipe survives the "burn-in" phase, failure rates are typically constant throughout most of the service life of the pipe, and an occurrence of a failure is considered a random event. This "in-usage" life period is usually followed by a wear-out phase. During this last phase, failure rates rapidly increase as a result of a deteriorating structural capacity.

Since pipes are usually considered repairable units, some models use the bathtub curve assumption to model the inter-break time between two failures. While some researchers evaluate all phases of the bathtub curve, some only consider one or two phases in their models [3,10].

The effect of pipe age on failure rates can also be considered in terms of installation era. In fact, the evolution of practices and industrial processes throughout the years has had a direct impact on failure rates. Barton et al. [4] described how the evolution in regulation and practices in the UK pipeline industry coincide with changes in failure rates throughout a utility's break history. In particular, the author showed how shifting local cast iron production to national and foreign manufacture of spun iron had caused an increase in failure rates due to a production of pipes with thinner walls. Also, the introduction of pipe coating and lining techniques in the 1950s had drastically improved on failure rates in the study. Consequently, some studies explicitly include installation time as a covariate or as a grouping criterion to capture those effects [8,12].

## OVERVIEW OF EXISTING MODELS

Pipe failure models have been extensively developed in the last 40 years to characterize the process of pipe deterioration, evince failure patterns, and anticipate failure events. As described by Kleiner & Rajani [6], models can be first classified into either physical or statistical. Physical models study the mechanical properties of pipes and their environment to determine the nature of the influence and how it impacts a pipe's service life. On the other hand, statistical models evaluate failure patterns on a larger scale by analyzing population-wide relationships between failure and pipe attributes.

**Physical models**

To characterize deterioration processes, physical models evaluate loads acting on pipes and the structural capacity resisting the resulting stresses. As reviewed by Rajani and Kleiner [6], models have approached these mechanical aspects by addressing the effect on deterioration for either a single component separately or multiple components at once, and approaches relied on both deterministic and probabilistic models.

Models focused on individual components attempt to characterize the influence of a single factor on pipe deterioration. A factor that has been commonly analyzed in the literature is frost load. To estimate frost load, empirical methods require several properties including a freezing index, soil backfill properties, and pipe depth. Such analysis suggested in particular that using a backfill soil with lower frost susceptibility compared to the sidewall helps mitigate frost load effect. For pipe soil interaction, models incorporate structural factors and consider that stress has in-plane and longitudinal components. In particular, stress sensitivity analysis supports the finding that smaller pipes suffer increased axial stress [14]. Other models have estimated failure pressure or tensile stress for a steel pipe having a corrosion pit using three-dimensional characteristics of corrosion pits. Alternatively, a pipe can be scored based on a corrosion status index (CSI) depending on the depth of the pit compared to the wall thickness.

In addition to one component, some models simultaneously integrate the effects of multiple factors. Such models can be either deterministic or probabilistic. Deterministic models assume no random variation after determining parameter values through empirical methods. An example of an empirical model uses a power function to predict pipe wall thickness or the pipe hoop stress as a function of corrosion pit depth and pipe age. Another conservative linear model estimates residual life with an assumption of constant growth rate of the corrosion pit depth. Some combined methodologies attempt to

harness various models on a scenario-based approach to determine remaining service life based on corrosion pit measurements and pipe characteristics.

For probabilistic models, instead of assuming determined values for parameters, probability distributions are integrated to capture the uncertainty about accurately identifying parameters and outcomes. In particular, the probability of steel pipe failure can be estimated using a pipe-soil stress model and a power function for the loss of wall thickness. Assumptions of probability distributions for parameters can thus be used to approximate the mean and variance of the failure tensile stress. Also, several models use the residual pipe strength to assist in decision making for inspection schedules. For example, a ratio of the residual strength and the predicted deterministic strength was assumed to follow a log normal distribution. By modeling the deterioration of the strength as a linear birth process, time to failure probabilities can be minimized and thus help identify optimal inspection schedules. Other physical models have alternatively attempted to evaluate the influence of temperature on failure rates using methods like multiple regression analysis. [6]

To characterize structural stress and identify failure points, these physical models in general require detailed pipe-level information. This information can include pipe intrinsic properties related to coating, joint types, and wall thickness. Operational information might also be needed to evaluate how the pipe reacts to pressure transients, chemical water properties, or hydraulic pressure. The environment of the pipe additionally brings another set of components to be integrated. These components can include pipe burial depth, soil properties, temperature, etc. Ideally, a model that perfectly characterizes the physical deterioration process would need to include all such information that influences the process. However, in practice, acquiring detailed information about an individual pipe and its environment is a costly endeavor. Most

water utilities typically only have general information about their pipe network, which drastically limits the potential of physical models. Nevertheless, physical models can be very useful when detailed analysis is needed for a portion of the network or select pipes to understand and model a deterioration behavior of specific interest.

**Statistical models**

As opposed to physical models which are limited in application due to the level of detailed and costly information required for each pipe, statistical models provide an alternative to assess the condition of large distribution systems.

Kleiner and Rajani [3] presented an overview of statistical models discussed in the literature which the authors classified into three categories: deterministic, probabilistic single-variate, and probabilistic multi-variate models. This review was later updated in 2012 and included another level of classification related to the type of deterioration [16]. The type of deterioration referred to whether a statistical model was interested in breakage frequency, survival analysis, or condition rating. According to the review, deterministic models have generally used grouped data and included time exponential and time linear approaches to estimate the number of breaks or the age at failure. In contrast, probabilistic models have incorporated uncertainty in determining model parameters to analyze the probability of failure, life expectancy, or failure clustering.

While the authors presented a comprehensive review of statistical models and underlying assumptions, a unified perspective was still needed to compare the models. To that end, Scheidegger et al. [17] presented a review of statistical models on a comparable basis by formulating the models into their failure rates representation and assessing their

predictive power. Prior to presenting the comparison, common model assumptions are first introduced.

***Model assumptions***

To specify a statistical model, assumptions about the properties of the model are usually required. In general, statistical models consider failure events as a stochastic process where pipes can fail at any time, repairs are immediate, and failure counts are unbounded. These failure counts are typically assumed to follow a Poisson distribution, thus describing Poisson point processes. To characterize such stochastic processes, a failure rate $\lambda$ is defined as:

$$\lambda\big(t|H(t)\big) = \lim_{\Delta \to 0^+} \frac{Prob(failure\ in\ [t, t+\Delta]|H(t))}{\Delta} \qquad (1)$$

Where H(t) represents the failure history at time *t.* It follows that the failure rate represents a probability density function per unit of time and reflects the propensity for failure at a given time *t*. If the failure rate is constant, the Poisson point process is characterized as homogeneous. Otherwise, the model is defined as non-homogeneous.

Also, the life cycle of a pipe is typically assumed to follow the "bathtub" curve. At the beginning of the pipe's life cycle, the failure rate starts high with a quick decline during its "burn-in" phase. The high rate is typically due to installation risks related to construction mishandling, pipe defects, and similar infant break factors. Then, the failure rate stabilizes at a low value during the "in-usage" phase which makes up most of a pipe's life cycle. Towards the end of the pipe's life cycle, the failure rate increases steadily as part of the "wear-out" phase. During that phase, the structural integrity of the pipe is undermined, and the pipe is increasingly more prone to failure under the same conditions. When a repair follows a pipe's failure, the pipe's structural capacity is restored, and a new life cycle begins. Incorporating the entirety of a pipe's life cycle is

often a complex task. Therefore, statistical models typically depict one or all phases of the cycle [3].

To specify the relationship between a dependent variable and independent variables, parameters of statistical models need to be estimated. If a statistical model estimates factors based on observed failure events and uses the factors to predict future failures, predictions are considered conditional. Else, predictions are considered unconditional if a model does not account for past failure. Two common approaches use failure history information to estimate parameters and calibrate a model: Maximum Likelihood Estimation (MLE) and Bayesian Inference (BI). Both methods estimate parameters by formulating a likelihood function that describes the probability of observing available data given those parameters.

Also, observed data have censorship characteristics that need to be accounted for in a statistical model to reduce estimation bias. These characteristics refer to right censorship, left truncation, and absence of replaced pipe data [17]. Right censorship represents a situation where events are not recorded after the latest failure in the dataset. In particular, if pipes are still in service, they are right censored. For left censorship, failure information prior to the earliest observation in the dataset is not included. Throughout the recording period, there might also be times when data about replaced pipes are absent for various reasons including missing, deleted, or corrupted data records. As different likelihood functions can be derived based on different censorship characteristics, considering these characteristics prior to model calibration is essential to reducing bias.

***Published models***

Statistical models have been presented in the literature via different formulations and following different assumption. Following a systematic approach to comparing published models, Scheidegger [17] presented a review of the main statistical models in their failure rate representation. Table 2 presents those models with their failure rate formulation, main characteristics, and assumption.

Table 2: Published statistical models as described by Scheidegger [17]

| **Model reference** | **Failure assumption** | **Failure rate** | **Model main characteristics** |
|---|---|---|---|
| Eisenbeis [18] | Weibull-distributed inter-break times | $\lambda(t, t_n, n) = \begin{cases} \theta_1 (t)^{\theta_1 - 1}, n(t) = 0 \\ \theta_2 \left(t - t_{n(t)}\right)^{\theta_2 - 1}, 1 \leq n(t) < 4 \\ \theta_3 \left(t - t_{n(t)}\right)^{\theta_3 - 1}, 4 \leq n(t) \end{cases}$ | - Deterioration factors are integrated through proportional hazards approach (pipe properties, external conditions, 15 years past failure<br>- The failure rate restarts at zero at repair times (deterioration due to repairs is not modeled) |
| Gustafson and Clancy [19] | Generalized Gamma-distributed inter-break times | $\lambda(t, n) = \begin{cases} \theta_1 \theta_2 (\theta_2 t)^{\theta_1 - 1}, n(t) = 0 \\ \theta_{n(t)+2}, 1 \leq n(t) < 10 \\ \theta_{13}, 11 \leq n(t) \end{cases}$ | - The model uses exponential distributions after the first failure for simplification<br>- The constant rates past the first inter-break time do not reflect deterioration over time<br>- Model calibration requires a large dataset to estimate rates up to the $11^{th}$ failure |

Table 2, Continued

| | | | |
|---|---|---|---|
| Pelletier [20] | Weibull-Exponential distributions | $\lambda(t, t_n, n) = \begin{cases} \theta_1\theta_2(\theta_2 t)^{\theta_1-1}, n(t) = 0 \\ \theta_3\theta_4\left(\theta_4(t-\mathrm{t}_1)\right)^{\theta_3-1}, n(t) = 1 \\ \theta_5, n(t) = 2 \\ \theta_6, n(t) \geq 3 \end{cases}$ | - Fewer parameters to estimate require a smaller dataset size than previous model |
| Røstum [21] | Weibull distribution. Non-homogenous Point Process | $\lambda(t) = \theta_1\theta_2 t^{\theta_2-1}$ | - It is assumed that the failure rate increases with age<br>- The model does not account for the influence of previous failures (NHPP process is memoryless) |
| Watson et al. [22] | Homogenous Point Process | $\lambda = \theta_1$ | - The model is too simple to model the complexity of pipe deterioration<br>- Calibration relied on Bayesian inference with a Gamma distributed failure rate prior |
| Economou et al. [23] | Weibull distribution | $\lambda(t) = \theta_1 t^{\theta_1-1}$ | - Zero-inflation is modelled per pipe but the effect on failure rate is unclear<br>- Random effects were estimated via Bayesian inference |

Table 2, Continued

| | | | |
|---|---|---|---|
| Le Gat [24] | Yule process | $\lambda(t,n) = \big(1 + \theta_1 n(t)\big)\theta_2 t^{\theta_2 - 1}$ | - The model yields a binomial distribution for the likelihood of failure and predictions |
| Kleiner and Rajani [25] | Non-homogenous Point Process with year-long piece-wise constant rates | $\lambda(t) = \theta_1 t^{\theta_2}$ | - The model accounts for seasonal effects through time-varying covariates.<br>- Known previous failures are incorporated as a covariate which defies the NHPP memoryless assumption |
| Scheidegger et al. [26] | Weibull distribution for first inter-break time and exponential for subsequent times | $\lambda(t,n) = \begin{cases} \theta_1\theta_2(\theta_2 t)^{\theta_1 - 1}, n(t) = 0 \\ \theta_3, 1 \le n(t) \end{cases}$ | - Failure rate is not influenced by previous failures past the first failure and does not represent deterioration over time.<br>- The model is practical for small datasets since a limited number of parameters is included |

# Chapter 3: Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems

## INTRODUCTION

Water distribution systems are critical lifelines of modern infrastructure. Maintaining a reliable and efficient water distribution system is crucial to supporting all spheres of human activity. To organize maintenance efforts once the components of the distribution system wear out, deterioration modeling has always been a concern in asset management. As water distribution systems keep needing higher investments for rehabilitation in what has been described by the American Water Works Association as the replacement era, researchers have been increasingly interested in more accurately modeling deterioration processes. Recent advances have been made in collecting more detailed and higher quality pipe-level information on asset attributes and failure history, and models are further developed to incorporate more complexity and adaptability. Yet, often times, water utilities are confronted with a persistent lack of granular information on assets and faulty failure records. Limited information availability requires practical failure models that can provide actionable outcomes for water utilities to integrate in their asset management practices.

One set of models that provides practical applicability is logistic regression. Logistic regression models are used to estimate the probability of an event based on a set of variables. Several studies have used logistic regression to reveal the influence of certain factors on asset deterioration [27–29]. In particular, age has been determined as a factor that increases the likelihood of deterioration and failure, and pipe diameter has been identified as a factor that inversely influences the likelihood of failure [28].

Logistic regression has also been used to evaluate the predictability of pipe failures [30]. By setting a threshold on an estimated failure probability, logistic regression models have been used as classifiers, and the performance of the logistic model has been evaluated for several use cases. A common measure of this predictive performance has been the receiver operating

characteristic curve [31,32]. In some cases, binary logistic regression has performed better than other commonly used models [33]. In other cases, models like Artificial Neural Networks and Random Forest performed closely higher [34,35]. Other researchers have used the logistic regression model to rank pipes per their likelihood of failure [29,36,37]. Cooper et al. [38] further associated the likelihood of failure with spatial information on consequences to assign risk scores to individual water mains.

Yet, the predictive performance of logistic regression models remains to be fully explored [39]. A logistic regression model estimates the probability of failure in a limited time interval, and researchers typically reduce the time interval to a period small enough to include at most one failure [40]. However, a given failure may result from unusual stress and not reflect a deteriorating trend. It may then be useful to observe failure history in a more flexible period to cover longer deterioration trends. Also, little research has used the logistic failure probability to generate an estimate of the expected remaining service life as a metric for condition assessment. Opila and Attoh-Okine [1] suggested a methodology for calculating the Mean Time to Failure (MTF) using failure probability . The authors further converted MTF estimates into condition scores according to a flexible scale. Ultimately, the obtained scores incorporated various factors in addition to a water utility's risk attitude.

This chapter explores the applicability of the methodology using the logistic regression model and further expands the scope to allow for a flexible choice of the time interval of the dependent variable. The intended contribution of this chapter is threefold: (1) assess the performance of a logistic regression model with a flexible time-interval choice, (2) provide a measure of the Mean Time to Failure based on a specified logistic regression model, and (3) assign and evaluate pipe condition scores based on the MTF measure.

## METHODS

The proposed approach for pipe failure prediction and condition scoring involves five steps: (1) data collection and processing, (2) developing a logistic regression model to estimate

the probability of a pipe failure in a given time period, (3) estimating the mean time to failure for each pipe, (4) assigning scores to each pipe according to a condition scoring method reflecting a water utility's attitude towards risk, and (5) evaluating model performance. The main steps of the proposed approach are illustrated in Figure 2.

To demonstrate the methods, the proposed approach was applied on a dataset of 4,153 water distribution pipes with 6,769 failure events covering a time period from 2000 to 2019.

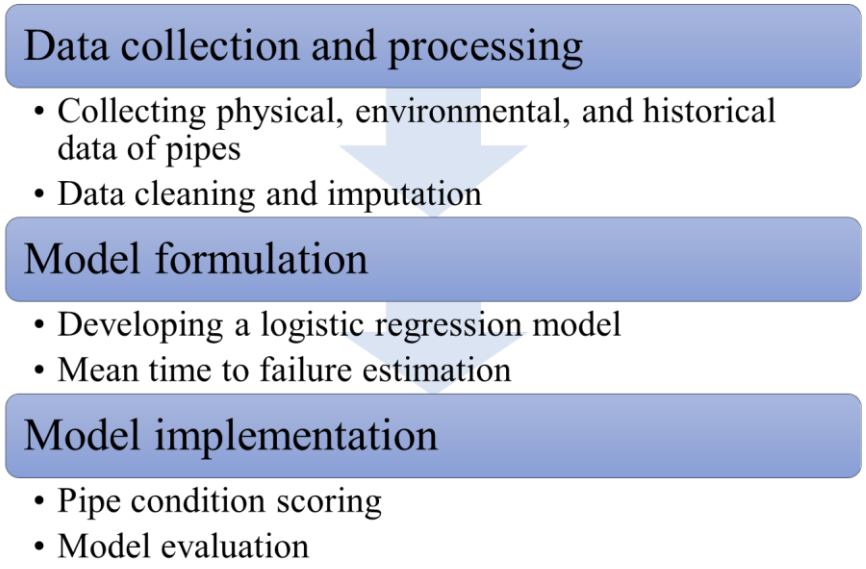


Figure 2: Research methodology

## Logistic regression model

### *Model formulation*

The first step of the modeling approach is to estimate the probability of pipe failure using physical, environmental, and historical information of individual pipes. To estimate failure probability, the approach relies on a logistic regression model with a flexible prediction time interval.

In fact, the approach allows for a flexible selection of a prediction period of interest consisting of a selected number of years *T*. One to several years can be chosen as a T-year period depending on the resulting model performance and the water utility's preference.

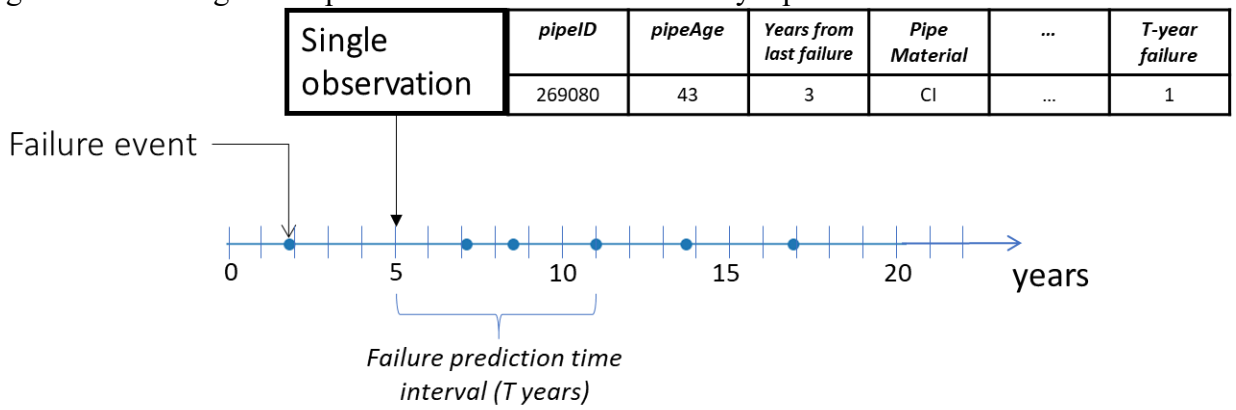


Figure 3. Diagram of the covariates for an individual pipe

Figure 3 illustrates the timeline of an individual pipe, a sample of its covariates, and the prediction time period *T*. The covariates are characteristics of an individual pipe measured at the beginning of a *j*th T-year period with $j = 1,2, \dots, n_i$ and $n_i$ is the total number of T-year periods covered by a single pipe *i*'s timeline. These covariates include pipe age, material, diameter, soil conditions, as well as time elapsed since last failure. In section 3.1, a detailed description of the pipe information used in this work is provided. For an individual pipe *i*, the logistic regression model estimates the probability of the pipe failure event $Y_{ij}$ occurring in a *j*th T-year period given a set of pipe covariates represented using a vector $X_{ij}$. Each covariate influences the probability of failure according to a coefficient $\beta = (\beta_1, \beta_2, \dots, \beta_p)$. Eq. (2) represents a mean probability $P_f$ of a failure event $Y_{ij}$ for a single pipe *i* in the *j*th T-year period, and $X_{ij}$ is a vector of pipe covariates.

$$\mu_{ij} = P_f\left(Y_{ij} = 1 \middle| X_{ij}\right) = \frac{1}{1+e^{-X_{ij}{}'\beta}} \quad (2)$$

To specify the relationship between covariates and the response variable, regression coefficients $\beta$ are to be estimated. For deterministic models, regression coefficients are typically estimated by maximizing a likelihood function. An important assumption in calculating a

likelihood function is independence of observations [41]. However, since the dataset included observations of a same pipe for different periods, the longitudinal nature of observations could potentially induce a correlation across pipe failure response, thus violating the sensitive assumption of independence.

In fact, the raw data is restructured by non-overlapping periods of T years. For each *j*th T-year period, a single response $Y_{ij}$ indicates whether a pipe *i* failed at least once during the *k*th T years, i.e., $Y_{ij} = 1$, or not, $Y_{ij} = 0$. Hence, a same pipe appears as a data point *j* times in the model. These repeated measures might create a correlation in the dataset that is similar to the standard autocorrelation often exhibited in many time series (e.g., hydrological time series and spatial environmental data) where samples are not spaced enough in time or space. Correlated samples might not provide an accurate representation of the population. For example, 10 uncorrelated samples from a population might provide the same accuracy as 100 correlated observations.

However, in statistical models, the purpose is to characterize a population when only a sample of the population is available. Hence, the measurements taken of the population - pipe failures in the present study - need to be a reliable representation of the population. It follows that there is a need to account for a potential correlation between samples.

To account for a possible correlation between outcomes for each individual pipe, a Generalized Estimating Equations (GEE) method estimates regression coefficients by incorporating within-cluster effects through their population-average [42]. A cluster in the present dataset refers to a single pipe with multiple observations. This method also prevents the need to explicitly specify a probability model of the correlation structure.

Let $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})'$ represent the response vector of the *i*th pipe consisting of $n_i$ observations and $\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{in_i})'$ refers to the mean vector of failure probability for pipe *i*. Let $V_i$ be the variance-covariance matrix for $Y_i$ and is defined as $V_i = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$ where $A_i = Diag\{\mu_{i1}, \mu_{i2}, \dots, \mu_{in_i}\}$ and $R_i(\alpha)$ is known as the working correlation structure. $R_i(\alpha)$ is a square matrix of elements $Corr(Y_{ij}, Y_{ik})$ and size $n_i \times n_i$ and is defined based on one of several

commonly used types of covariance structures. $R_i(\alpha)$ also depends on a parameter $\alpha$ estimated iteratively based on the number of regression coefficients *p* and the response residuals defined as $e_{ij} = (y_{ij} - \mu_{ij})/\sqrt{\mu_{ij}}$ using the current value of the coefficient $\beta$ [43]. Table 3 details these matrix elements and parameter estimation for the independent, exchangeable, and autoregressive correlation structures used in this study.

Table 3. Correlation matrix elements for common working correlation structures for a pipe *i*

| Correlation structure | $Corr(Y_{ij}, Y_{ik})$ | Parameter estimator |
|---|---|---|
| Independent | $Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$ | - |
| Exchangeable | $Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$ | $\hat{\alpha} = \frac{1}{N' - p} \sum_{i=1}^{K} \sum_{j \neq k} e_{ij} e_{ik}$ <br> $N' = \sum_{i=1}^{K} n_i (n_i - 1)$ |
| Autoregressive AR(1) | $Corr(Y_{ij}, Y_{i,j+m}) = \alpha^m,$ <br> $m = 0,1, \ldots, n_i - j$ | $\hat{\alpha} = \frac{1}{K_1 - p} \sum_{i=1}^{K} \sum_{j \leq n_i - 1} e_{ij} e_{i,j+1}$ <br> $K_1 = \sum_{i=1}^{K} (n_i - 1)$ |

Despite the existing difference among correlation structures, estimates of the regression coefficients are asymptotically consistent despite a misspecification of the correlation structure [44].

For *K* pipes and *p* covariates, regression coefficients $\beta$ can be estimated by solving the GEE in Eq. (3):

$$For\ j = 1, \ldots, p: \quad \sum_{i=1}^{K} \frac{\partial \mu_i}{\partial \beta_j} V_i^{-1} (Y_i - \mu_i) = 0 \qquad (3)$$

To decide upon the goodness of fit of a logistic model based on a specified correlation structure that accounts for potential correlation from multiple observations from each pipe, the

Quasi-likelihood under the Independence model Criterion (QIC) is used [45]. Unlike likelihood-based methods such as the Maximum-Likelihood (ML), GEE-based models do not explicitly specify a likelihood function. However, the QIC metric provides an alternative to the commonly used Akaike Information Criterion (AIC) metric to compare the goodness of fit for different GEE models, such that a GEE model with a lower QIC value is considered a better fit for the dataset.

***Selection of covariates***

An important step of developing a logistic regression model is the selection of covariates. Covariate selection can improve a model's interpretability, avoid overfitting by diminishing the effect, filter out covariates with low relevance without compromising model accuracy, and even improves prediction performance for new observations.

In this study, covariate selection is carried out in two steps. First, Least Absolute Shrinkage and Selection Operator (LASSO) regression reduces the number of covariates based on their contribution to the performance of the logistic regression model. Secondly, a Recursive Feature Elimination (RFE) method is performed to further reduce the number of covariates.

LASSO regression is a statistical tool that performs variable selection by shrinking less significant regression coefficients to zero [46]. Coefficient shrinkage is possible by integrating an additional term to the error minimization such that the goal of lasso regression is to solve Eq. (4):

$$\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{N} \| logit(\mu) - X'\beta \|_2^2 + \lambda \|\beta\|_1 \right\} \qquad (4)$$

Where *logit* is the logistic function defined as $logit(P_f) = \ln\left(\frac{P_f}{1-P_f}\right)$, $\lambda$ is a regularization parameter that balances between two objectives: minimizing the sum of squared error between the fitted and observed failures (first term) and regularization (second term). $\| \ \|_2$ is the Euclidian norm *l2* defined such that $\| logit(\mu) - X'\beta \|_2^2 = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (logit(\mu_{ij}) - X_{ij}{}'\beta)^2$. $\| \ \|_1$ is the *l1* norm defined such that $\|\beta\|_1 = \sum_{i=1}^{p} |\beta_i|$. The $l1$ regularization penalizes a model with many covariates. The rationale for including the $l1$ penalty is that it modifies the estimation to achieve sparsity by eliminating the predictors that explain the response variable the least. It also prevents

overfitting the model. By cross-validating over $\lambda$ values, the value that yields the lowest objective function was chosen for the present dataset.

An additional step of covariate selection was performed using RFE. The goal of RFE is to select covariates by recursively considering a decreasing number of covariates [47]. First, a logistic regression model is trained on the set of covariates selected after the LASSO regularization step, and the statistical significance of each covariate is obtained through p-values for each covariate's coefficient. The covariate with the highest p-value is eliminated from the current set of covariates. This procedure is repeated on the resulting subsets until the highest p-value is below a specified cutoff of 0.10. The final subset of covariates was then used to develop the final logistic regression model that estimates pipe failure probability for a given T-year period.

The outcome of the logistic regression model provides an estimate of the probability of a pipe failure in a T-year period by integrating the effects of the correlation structure and selected physical, environmental, and historical information. Then, a discrete decision about the state of the pipe can be made by setting a discrimination threshold on a given failure probability of a pipe. If the failure probability exceeds this threshold value, a pipe is expected to fail, i.e., the failure outcome is equal to 1. If the estimated probability is below the designated threshold value, the pipe is expected to survive, i.e., the failure outcome is equal to 0.

***Estimating Mean Time to Failure (MTF)***

The developed logistic regression model estimates failure probabilities for each pipe, which provides a measure of criticality for a given T-year period. While such a measure can assist a water utility in defining maintenance priorities for a planning period, it does not provide a direct measure of the expected time to failure. To estimate the remaining time to pipe failure, the proposed approach relies on calculating the Mean Time to Failure (MTF). MTF is a reliability parameter typically used to account for the expected life expectancy in the design of

products [40]. For repairable systems, MTF refers to the time between failures, i.e., inter-arrival time, and it can be estimated as the arithmetic mean of the survival probability over time:

$$MTF = \int_{t_0}^{\infty} P_s(t)dt \qquad (5)$$

Where $t_0$ denotes the pipe's repair time and $P_s(t)$ is the survival function defined as the probability that a pipe will survive past a time t.

Assuming an independence across failure events for a given pipe, the probability that a pipe survives past a time t was approximated by the product of the probabilities that the pipe survives during each of the successive T-year periods leading to the time t, each T-year survival event being conditional on the pipe surviving up to the beginning of the T-year period. The MTF can, thus, be approximated following Eq. (6):

$$MTF = T \cdot \sum_{n=0}^{\infty} \prod_{k=0}^{n} P_s(k) \qquad (6)$$

Where $P_s(k)$ refers to the probability of survival during the kth T-year period with $k = 1, 2, \dots, \mathrm{n}_i$. Since the event "at least one failure" is the complement of a survival event, the probability of failure in a T-year period $P_f$, as estimated by the developed logistic regression model, was used to calculate the MTF following Eq. (7).

$$MTF = T \cdot \sum_{n=0}^{\infty} \prod_{k=0}^{n} \left(1 - P_f(k)\right) = T \cdot \sum_{n=0}^{\infty} \prod_{k=0}^{n} \frac{1}{1 + e^{X(k)^T \beta}} \qquad (7)$$

Where $X(k)$ represents the vector of covariates measured at the beginning of the *k*th T-year period. This method converts the failure probabilities in a limited time interval to a measure of expected time to the next failure. The MTF is a direct measure that can be used by water utilities to decide whether to include pipes in repair and improvement projects.

***Condition scoring***

The first outcome of the proposed approach was a T-year probability of pipe failure, and the second yielded estimates of the mean time to next failure. The third step assigns pipe condition scores to facilitate the water utility's risk assessment and prioritize maintenance,

replacement, and decide on project scope. Furthermore, the scoring approach is flexible to the utility's risk attitude and the granularity of scores it desires.

The condition scoring method, as suggested by Opila & Attoh-Okine [1], uses the economic concept of discount rate to assign condition scores to pipes based on MTF estimates. According to its economic interpretation, a discount rate typically implies the extent to which future benefits are valued, where a higher discount rate implies a lower present value of money accrued in the future, compared to a higher present value of money with a lower discount rate. In this study, a discount rate $d$ refers to a factor that penalizes shorter times to failure and resulting in a more critical condition score of pipes. Given a maximum desired criticality score $S_{max}$, a discount rate $d$, and the MTF of a pipe, a pipe's condition score can be determined as:

$$S = \frac{S_{max}}{(1+d)^{MTF}} \quad (8)$$

This condition scoring method assigns a single score to a pipe, which lumps the impact of various environmental and physical covariates, pipe failure history, as well as a utility's attitude towards risk and decision scale. Higher scores indicate higher criticality, and higher discount rates suggest that fewer pipes will have high scores for a given MTF, thus reflecting a lower level of rehabilitation priority [1]. Figure 4 illustrates the condition scoring proposed in Eq. (8) as a function of the calculated MTF proposed in Eq. (7). Based on the curve, scores can be assigned to pipes on either a continuous (solid line in Figure 4) or a discrete (dashed line in Figure 4) scale. In the present study, scores were assigned using a discrete scale, which allowed to categorize pipes into a finite number of groups for a practical input for asset management.

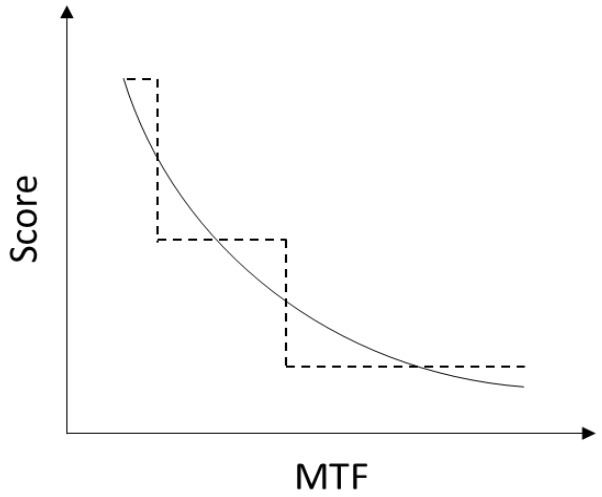


Figure 4. Condition scoring curve

**Model evaluation**

The proposed framework includes a logistic regression model for pipe failure prediction based on estimated failure probabilities and a condition scoring method using the concept of MTF. In order to evaluate the accuracy of the proposed framework, several classification and error metrics were employed.

To evaluate the performance of the logistic regression model, a confusion matrix is used. As shown in Figure 5, a confusion matrix summarizes the performance of a classification model by showing both discrepancy and agreement between true labels and predicted labels [48]. Before computing the confusion matrix, predictions are obtained by converting failure probabilities to a binary outcome (i.e., pipe failure or pipe survival) by setting a probability threshold. Following the confusion matrix terminology, correctly predicted labels are either True Positives (TP) or True Negatives (TN), and incorrectly predicted labels are either False Positives (FP) or False Negatives (FN). Here, positives represent pipe failure event and negatives represent pipe survival.

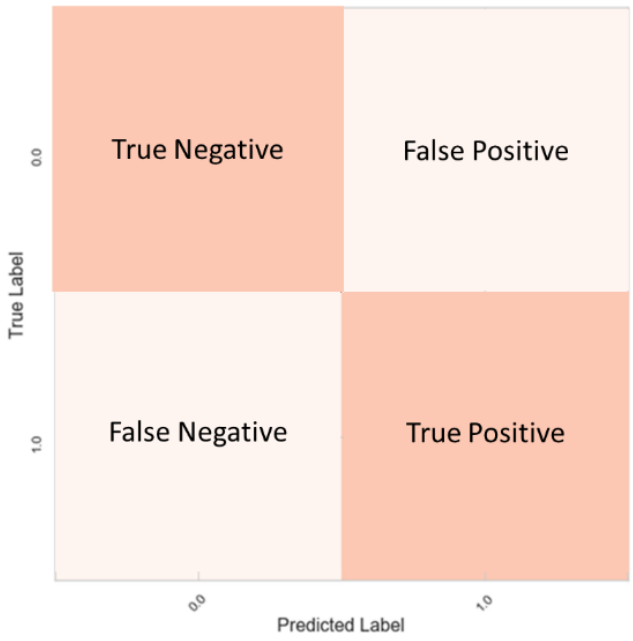


Figure 5. Confusion matrix

Additionally, based on the confusion matrix, several performance metrics were calculated. Table 4 shows the calculated metrics and their definitions.

Table 4. Classification metrics and definitions

| Classification metric | Definition |
|---|---|
| Precision | $\frac{TP}{TP + FP}$ |
| Recall or True Positive Rate | $\frac{TP}{TP + FN}$ |
| False Positive Rate | $\frac{FP}{TN + FP}$ |
| Accuracy | $\frac{TP + TN}{TP + TN + FP + FN}$ |
| Matthews Correlation Coefficient (MCC) | $\frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$ |

Since predictions are made based on a chosen probability threshold, the defined classification metrics can only be comprehensively interpreted if a threshold value is justified. To decide upon the choice of a probability threshold, Receiving Operating Characteristic (ROC) and

Precision-Recall curves are common tools to analyze the impact of a varying threshold on model performance [31,48,49]. As shown in Figure 6, a ROC curve is a graphical tool that plots True Positive Rate (TPR) values versus False Positive Rate (FPR) values for a varying threshold, where $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{FP+TN}$. A high TPR indicates the rate of correctly predicted pipes that are expected to fail and a low FPR indicates the rate of pipes whose failure was incorrectly predicted by the model. Hence, the goal is to achieve a high TPR and a low FPR. A performance metric associated with a ROC curve is the Area Under the Curve (AUC). The closer AUC is to 1, the better the model is at correctly predicting the true events and simultaneously minimizing false predictions.

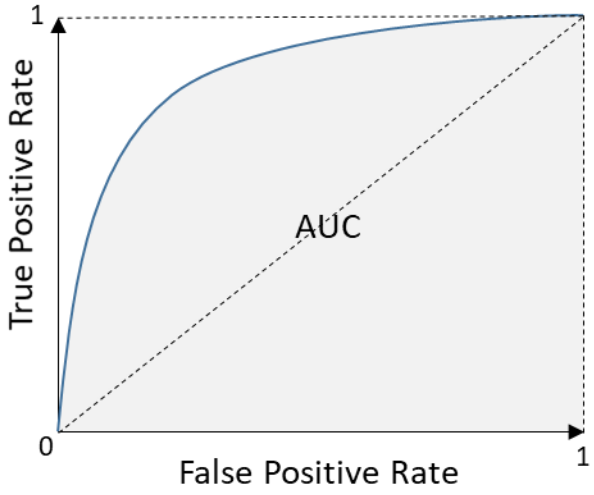


Figure 6. ROC curve

While a ROC curve allows to visualize how well a classifier captures true labels, ROC curves can be influenced by imbalanced true and positive events. When the number of negative events is much greater than the number of positive events (as typically occurs for pipe failure data where a majority of pipes do not exhibit failures), the FPR can be artificially suppressed making it more difficult to assess model performance. Instead, Precision-Recall curve performs better for imbalanced datasets, where precision indicates the fraction of pipes identified by the model that are expected to fail that indeed experience failure, and recall indicates the sensitivity

of model prediction [49]. A tradeoff applies between precision and recall as the probability threshold varies. When the probability threshold is low, the number of unidentified failure events is expected to decrease, thus having higher recall values. However, the number of events incorrectly classified as failures will increase as well, thus decreasing the model's precision. As the probability threshold increases, fewer relevant events will be identified (i.e. lower recall), however the confidence (i.e., precision) of correctly identified events will be greater. It is useful to plot precision and recall curves against the threshold settings, as illustrated in Figure 7, thus visualizing how different threshold levels specifically influence both curves. Visualizing the precision and recall tradeoff curves allows the water utility to directly set the probability threshold to achieve a desired level of performance.

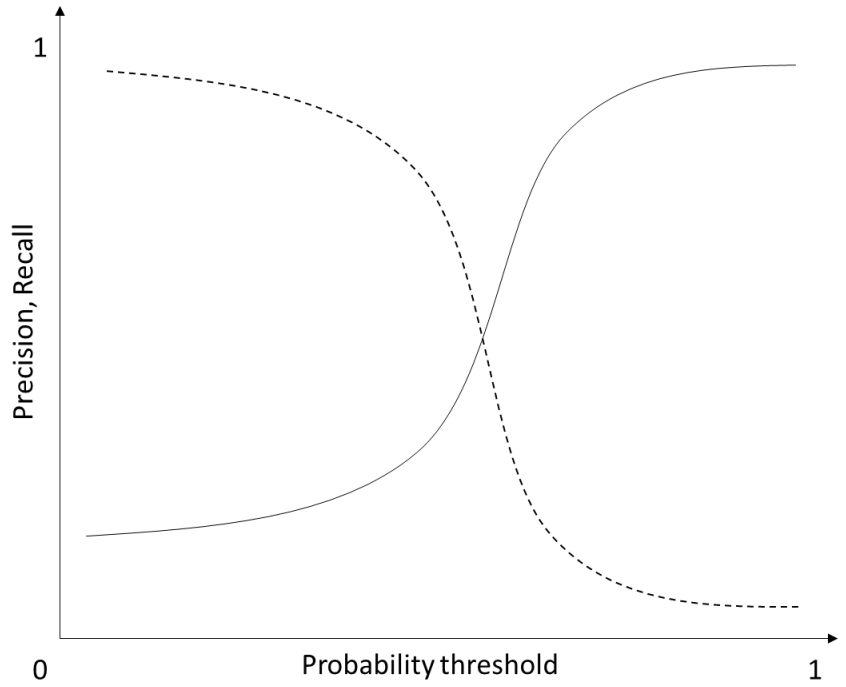


Figure 7. Precision (solid) & Recall (dashed) vs Threshold

Classification metrics listed in Table 4 and ROC and Precision-Recall curves are useful to improve failure predictability and, in turn, the MTF and condition scoring by determining the probability threshold. For MTF calculation and condition scoring, results can be evaluated

against the observations by comparing the MTF to the actual time to failure for pipes that failed more than once in the observation period by using qualitative and quantitative measures such as histograms, boxplots, and the Root Mean Square Error (RMSE).

## APPLICATION & RESULTS

The proposed framework is demonstrated using the information provided by the City of Austin, which included data about pipe characteristic, location, and failure history. All models developed in this work were implemented in Python 3.7, and some preliminary data processing was executed in ArcGIS Pro 2.4.0.

### Data description and preprocessing

The studied drinking water distribution system consists of 244,830 pipe segments with a total network length of 5,202.1 miles. A unit in the pipe network was either a whole pipe or only a segment of a whole pipe. Because the utility's records did not explicitly make such a distinction, the present study refers to both types as pipes. Out of the total number of pipes on file, past failure was only recorded for 4,425 pipes. These repaired pipes account for a total 6,989 recorded repair events spanning from 2000 to 2019. A repair event is typically triggered by a reported leak and refers to an intervention from a utility's maintenance team to restore a pipe into service.

Prior to considering pipe attributes, the dataset had to be screened for duplicates and other inconsistencies. In addition to removing duplicates, a portion of failure events was not stored in a readable format, which practically resulted in a total 6,769 failure events from 4,153 pipes. Finally, the dataset with failure history had a total network length of 336.48 miles representing 6.5% of the entire network length.

Figure 8 shows the annual failure rate per the length of the entire network from 2001 to 2018. The first year 2000 and last year 2019 were excluded from this figure as failure data collection may not have been complete. Across the 2001 to 2018 period, pipes had in average

7.35 failures per 100 miles per year with a standard deviation of 3.29 years. Break rates mostly fluctuated between 5 and 11 failures per 100 miles per year. A 2018 survey of water utilities in the USA and Canada reported an average failure rate of 14 breaks per 100 miles per year, which was compared to other sources reporting failure rates ranging from 21 to 30 breaks per 100 miles per year [5]. This report also refers to typical industry targets of 11 to 15 breaks per 100 miles per year. This suggests that the failure rate calculated based on the dataset provided by the city of Austin was low. Because the failure records in the dataset only consisted of pipes representing 6.5% of the entire network, another portion of the network must have suffered past failures. Also, as can be seen in Figure 8, unusually low failure rates were recorded in 2001 and 2002 with no provided explanation. Despite years with unusual rates, the entire pipe failure dataset was considered in the analysis. Excluding outliers was not warranted since individual events could not be directly associated with any identified variability in trends. Also, rejecting some events might influence potential correlations across the pipe network since a pipe failure might have an impact on adjacent pipes or other parts of the network.

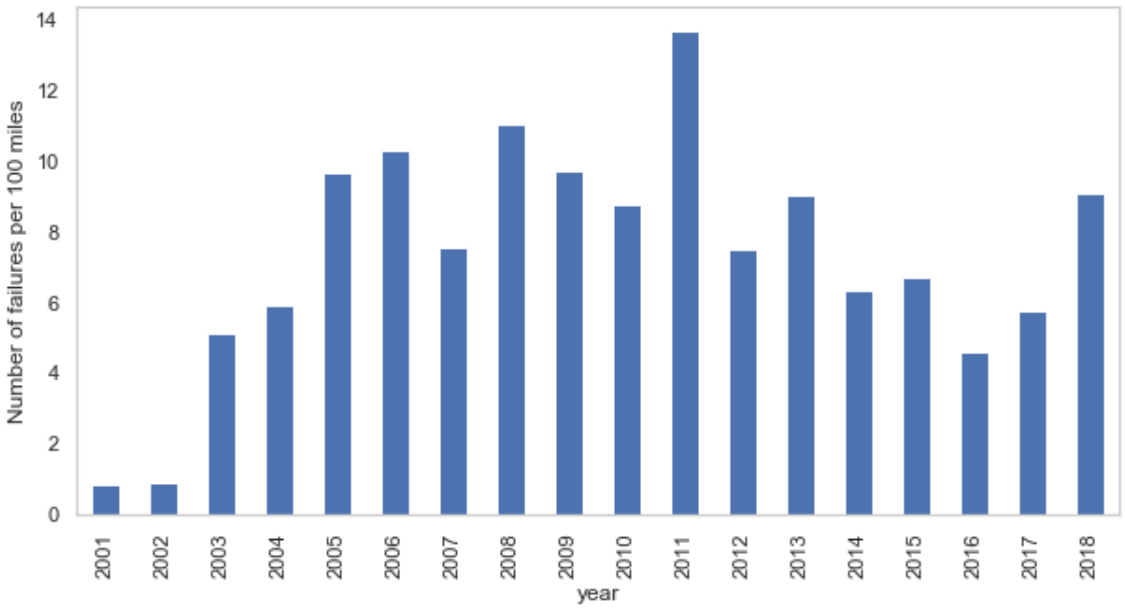


Figure 8. Failure rate per year

Relevant attributes that were provided with the dataset included pipe length, diameter, age, material, and pressure zone. Physical, environmental, and historical information used in this analysis are briefly summarized as follows.

*Pipe material*. The majority of pipes consist of cast iron (CI) pipes (71.2% of pipe length) followed by ductile iron (DI) (6.1%), Polymerizing Vinyl Chloride (PVC) (5%), and Asbestos Cement (AC) (13.7%). Other pipe materials included concrete steel cylinder, polybutylene, and copper, which comprised less than 4%. More than half of the pipes had only one past failure and 77.3% had either one or two past failures in the 20 years observation period.

*Pressure zones*. Pipe attributes included six main pressure zones, North (NO), Central North (CN), North West (NW), South (SO), Central South (CS), South West (SW), and Others. CS, CN, and N pressure zones included 73.6% of the pipes with recorded past failures.

The analyzed dataset thus consisted primarily of CI pipes with one or two past failures and around the central area of the city. On the other hand, only 25.5% of the entire 5,202.1 miles of the network consisted of CI pipes, and the central south, central north, and northern pressure zones make up only 35.4%. Therefore, this difference between the studied sample and the total population needs to be considered when interpreting results.

*Pipe age*. As common with pipe records, approximately 12% of pipes were missing pipe age. The age of the remaining pipes was approximated using spatial interpolation based on radial basis function [50]. The age of CI pipes was further adjusted based on our discussions with the water utility following the changes in installation practices. As suggested by the water utility, CI pipe installation ceased in the early 1980s. A cutoff was therefore defined such that estimated installation dates for CI pipes that were dated after 1980 (approximately 3% of all the pipes) were instead approximated by assigning an age value from the geographically nearest pipe that was installed before 1980. This approximation assumed that those CI pipes were installed in the same year as the nearest pipes that were installed before 1980. Such an assumption is reasonable

considering that rehabilitation efforts typically target several pipes in a given geographical area for cost considerations.

Figure 9 shows the distribution of resulting pipe ages by pipe length and material. Newer pipes consist mostly of DI and PVC, and older pipes consist mostly of CI and AC pipes.

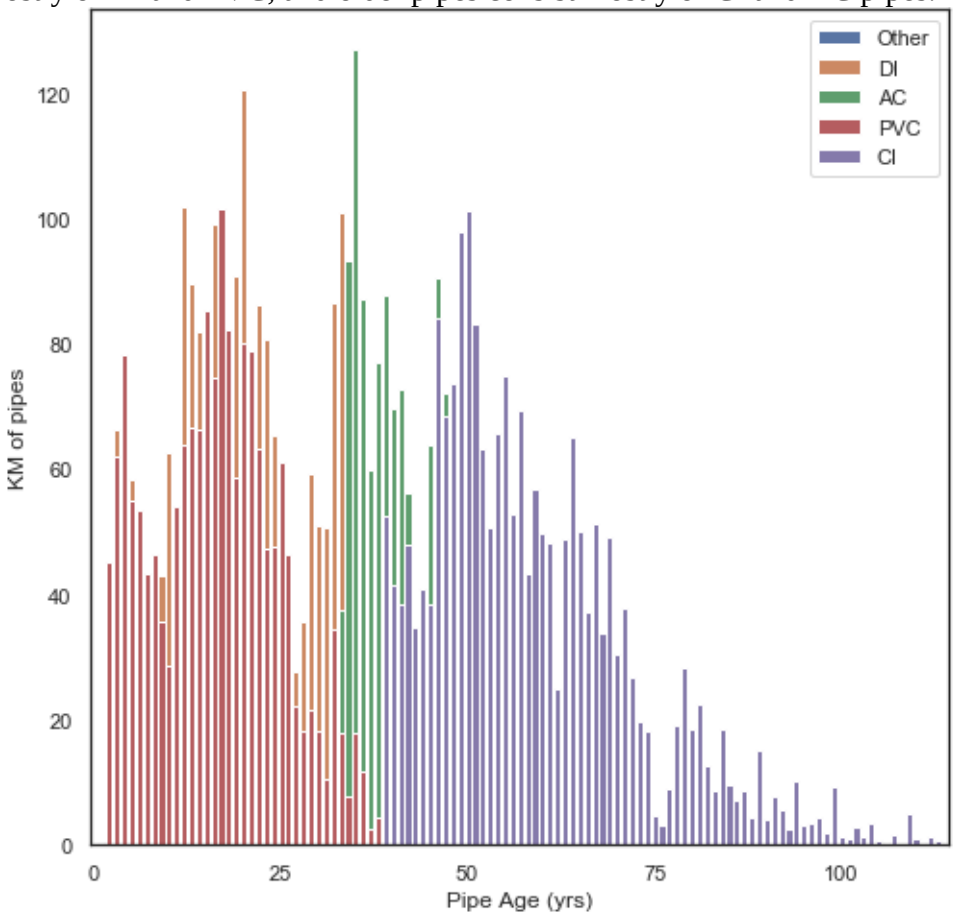


Figure 9. Distribution of pipe age and material

*Soil and land use*. Soil information was extracted from the Soil Survey Geographic (SSURGO) Database as provided by the National Cooperative Soil Survey. The database is made publicly available by the United States Department of Agriculture (USDA) [51]. Soil attributes included the dominant soil order, which is defined in accordance with USDA soil taxonomy [52]. The dominant soil order refers to a soil classification that lumps soil properties like depth, structure, and moisture. Additionally, land attributes were assigned to pipes with information on

road type and land use as potential covariates [53]. Annual precipitation was also considered as a model covariate and was provided as an average rainfall associated with soil information.

Table 5 summarizes the considered covariates in terms of categories, units, and important values. Overall, 15 different continuous and categorical covariates were considered in the regression model.

Table 5. List of covariates per category

| **Category** | **Covariate** | **Alias** | **Unit** | **Important values** |
|---|---|---|---|---|
| Pipe characteristic | Length | pipeLength | Ft | Mean: 419.18; Std: 424.18 |
| | Diameter | pipeDiameter | In | Mean: 7.13; Std: 4.64 |
| | Age | pipeAge | Years | Mean: 45.17; Std: 18.74 |
| | Material | pipeMaterial | – | CI; DI; AC; PVC; Other |
| Failure history | Number of past failures | NOPF | Break | Mode: 0; Mean: 0.51 |
| | Years from last failure | upTime | Years | Mean: 5.58; Std: 3.80 |
| Soil attribute | Elevation | terrainElevation | Ft | Mean: 611.41; Std:103.34 |
| | Concrete corrosion potential | corrosionConcrete | – | Low/Moderate/High |
| | Steel corrosion potential | corrosionSteel | – | Low/Moderate/High |
| | Saturated hydraulic conductivity | Ksat | inch/hr | Mean: 20.75; Std: 30.44 |
| | Mean annual precipitation | rainFall | mm | Mean: 876.81; Std: 79.70 |
| | Dominant soil order | soilOrder | – | Mollisols/Vertisols/Entisols/… |
| Land attribute | Land use | landUse | – | Commercial/Residential/Office |
| | Road type | roadType | – | Interstate; Minor Arterials; Private Road; |
| Operational | Pressure zone | pressureZone | – | CN; NO; CS; SO; NW; SW; Other |

## Logistic regression results

### *Model selection*

For model training, 75% of the dataset with observations from 2000 to 2014 was selected and used to develop the logistic regression model. The remaining records from 2015 to 2019 were held out for testing and validation. For the response variable, a period of 5 years was chosen as a T-year response window. In other words, the output of the regression model estimates the failure probability of a pipe in the next 5 years. The choice of the T-year period was based on three criteria: (a) a high resulting AUC score after developing the logistic regression model, (b) a period that offers practical implementation for the utility's asset management, and (c) a period that reduces imbalanced classification [54]. Table 6 shows that a period of 5 years scored the highest AUC compared to other periods. Also, the water utility's Capital Improvement Program follows a 5-year planning window, according to which a budget is allocated for pipe rehabilitation. It follows that a measure of pipe failure risk that covers the allocation period ensures a coherent approach to rehabilitation. In terms of data imbalance (i.e., the number of failure events versus the survival events), the shorter the T-year period is, the more imbalanced the dataset becomes. Preprocessing the dataset with a 5-year response variable yielded 32% failure events versus 68% survival events, which considerably reduced class imbalance. Consequently, a 5-year period was chosen for its practical application and the higher predictive accuracy that it provided.

Table 6. AUC scores for time interval selection

| Time interval (years) | AUC |
|---|---|
| 1 | 52% |
| 2 | 68% |
| 3 | 69% |
| 4 | 67% |
| 5 | 70% |
| 6 | 67% |

To estimate the effects of covariates, the logistic regression model used the GEE with an independence covariance structure. In fact, when compared to an exchangeable correlation, the independence structure provided a better fit as shown in Table 7, whereas the model failed to converge with an autoregressive covariance structure.

Table 7. Goodness of fit and covariance structures

| Covariance structure | QIC |
|---|---|
| Independent | 13,721.76 |
| Exchangeable | 13,859.32 |

The goodness of fit with an independent covariance structure suggests that failure events across pipes do not display a significant correlation in the present dataset. Additionally, estimates of covariates effects are still consistent despite a possible misspecification of the correlation structure [44]. Therefore, the final model estimated coefficients and failure probabilities based on an independent covariance structure.

***Effects of covariates***

The initial set of covariates was included into a LASSO regression model that was cross-validated across a range of continuous values for the regularization parameter $\lambda$. LASSO regression reached an optimum at $\lambda = 0.03$, thus filtering out 22 continuous and categorical covariates. The 25 retained covariates were recursively modeled into a GEE logistic regression model with an independent covariance structure, and variables with

the highest p-value were filtered out until the highest p-value of a subset was below a 0.10 cutoff. As an exception, despite its low statistical significance in the dataset, pipe age was retained considering its proven importance in the literature [4,8,12]. The resulting subset of covariates is shown in Table 8 with corresponding coefficients and p-values.

Table 8. Covariates effects as estimated by the LR model

| Covariate | Description | Coefficient | p-value |
|---|---|---|---|
| Intercept | Intercept | -0.83 | <0.01 |
| pipeAge | Pipe age | 0.04 | 0.15 |
| pipeDiameter | Pipe diameter | -0.07 | <0.01 |
| pipeLength | Pipe length | 0.20 | <0.01 |
| NOPF | Number of past failures | 0.15 | <0.01 |
| upTime | Years from last failure | 0.87 | <0.01 |
| landUse_residential | Residential land use | 0.07 | <0.01 |
| pipeMaterial_CI | CI pipe material | 0.08 | <0.01 |
| pressureZone_NW | North-West pressure zone | -0.09 | <0.01 |
| soilOrder_Vertisols | Soil order: Vertisols | 0.08 | <0.01 |

For pipe material, only the CI type was retained, which suggests that other material types did not provide sufficient statistical significance to count towards the final subset of covariates. In fact, over 70% of the studied dataset consisted of CI pipes. The consideration of a larger representation of other materials should allow for their analysis with more certainty in terms of impact on failure. Also, despite an expected high influence of steel and concrete corrosivity covariates, their values were only available for a portion of the dataset which might have led to their exclusion from significant covariates.

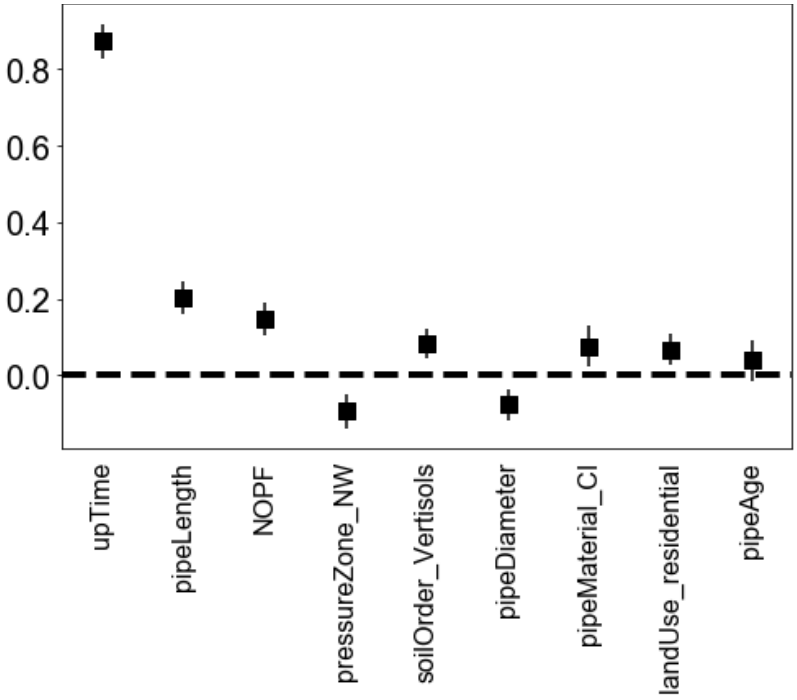


Figure 10: Logistic regression coefficients plot

When coefficients are ranked from most to least influential as in Figure 10, covariates related to failure history show some of the highest effects on pipe failure. The number of years from last failure (upTime) appears as the most influential attribute, thus suggesting that the more time elapses from a previous break, the more likely a pipe is to fail within the next 5-year period. This correlation is illustrated in Figure 11. A possible explanation for this effect is that a longer period without failure might indicate a longer exposure to internal and external factors affecting a pipe until its structural integrity is restored again. This interpretation supports the “in-usage” and “wear-out” phases of the bathtub failure rate curve assumption where the failure rate is expected to rise until a failure occurs [3].

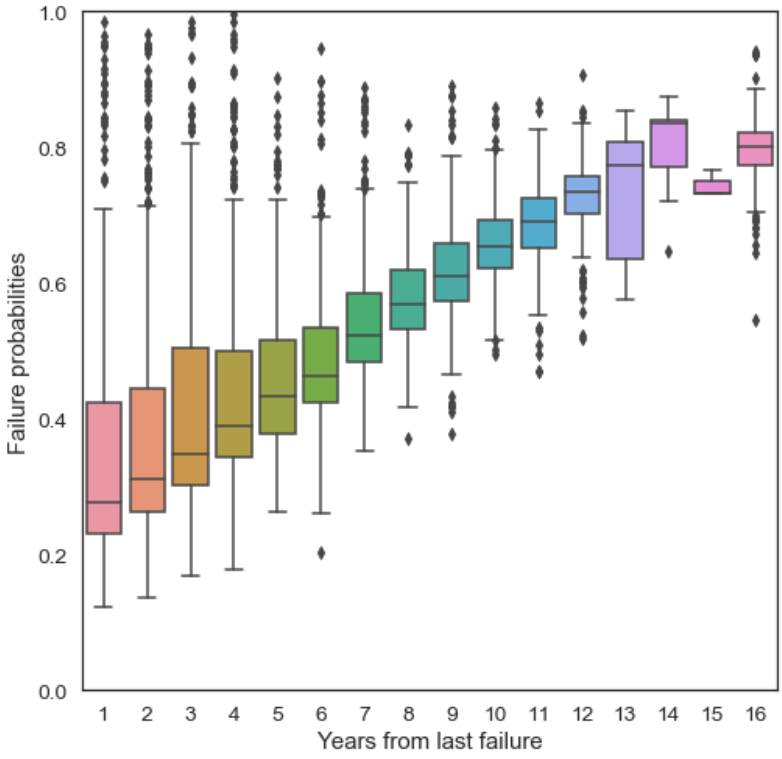


Figure 11. Failure probabilities versus the time from last failure.

*Note*. Whiskers are 1.5 times the interquartile range, any data point beyond is considered an outlier.

Additionally, the more total previous breaks are recorded at a pipe level, as integrated by the number of previous failures (NOPF) covariate, the higher the pipe failure probability is. This observation also coincides with the conceptual failure rate 'bathtub' model, in which the failure rate increases as the number of previous failures increases [17]. A rich failure history of a pipe could suggest a structural integrity that is undermined by repeated repairs.

In terms of pipe characteristics, covariates' importance was generally consistent with previous research findings. Pipe length has been associated with higher failure probability [9,10,15]. Beyond an additional exposure directly correlated to pipe length,

longer pipes could be more exposed to varying environmental conditions and more sensitive to effects like pressure transients [15]. Also in consistence with literature findings, smaller pipes inversely affect failure probability such that pipes with small diameters are associated with thinner walls which translate into a lower structural strength [8,10–12].

***Model evaluation***

In order to define a discrimination threshold for the developed logistic regression model and make predictions, the ROC curve is first generated for the test data, as shown in Figure 12. The corresponding AUC is 0.68, thus suggesting a reasonable discrimination strength for predicting pipe failures. By setting a discrimination threshold, the model can be positioned at a specific point along the ROC curve. As can be seen on the curve, a plausible discrimination threshold could be set at 0.75 so that the TPR is 60%, just before the slope is sharply reduced. However, while the ROC curve evinces the discrimination strength of the model, it is insensitive to the balance of the dataset and gives no measure of precision. In fact, TPR and FPR do not treat a misclassified event equally in percent terms. It might be tempting to seek an additional 10% of TPR by conceding 20% of FPR (by adjusting the probability threshold from 0.69 to 0.53), but a marginal increase in the FPR, which is twice the marginal increase in the TPR, could result in a number of false alarms that is much higher than twice the additional number of correct predictions.

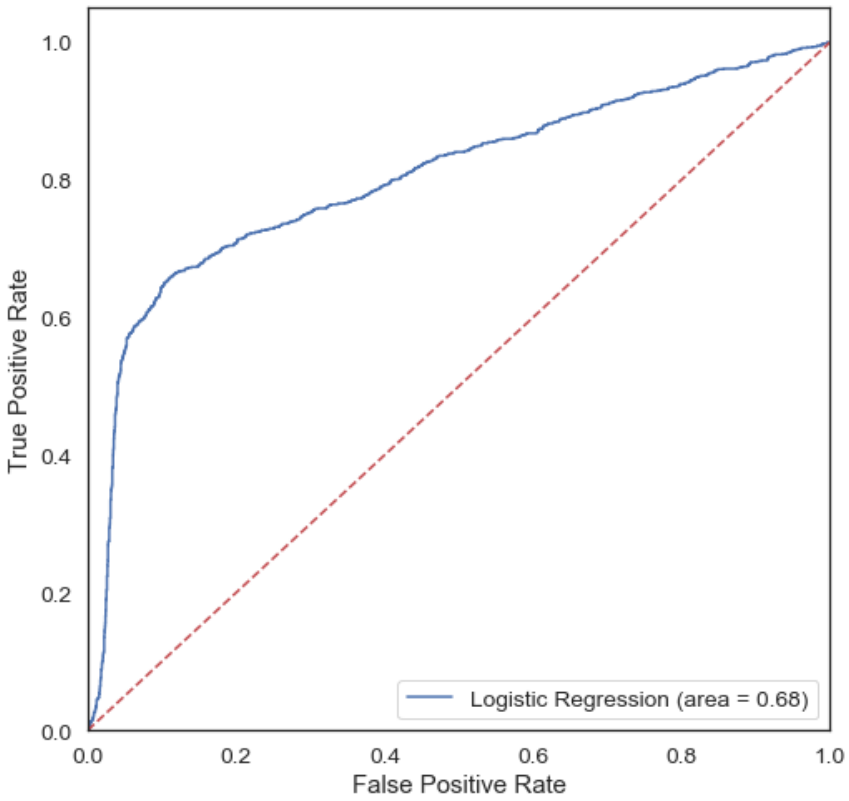


Figure 12. ROC curve of the logistic regression model

To account for the model's precision, the precision-recall versus discrimination threshold curves are plotted in Figure 13. The precision-recall curves can be visually used to control for the correct proportion of total predictions based on threshold values. While the objective is to maximize both precision and recall, the two metrics are conflicting, and the level of compromise needs to be determined. A choice of a discrimination threshold should be determined based on an acceptable level of performance for each metric. Acceptable levels may be determined per the priorities of the water utility. For example, a water utility might want to account for the fact that missing a true failure event is worse than having a false alarm. In fact, because the loss in recall is typically more costly than a loss in precision, setting a recall level that is higher than precision

could be warranted. In this study, no such preference was expressed by the utility, so the chosen probability discrimination threshold (0.69) was determined as the intersect of precision and recall such that both metrics are at 67%. By defining such a threshold, 67% of true failure events were correctly predicted by the model, and 67% of predicted failures corresponded to true failure events.

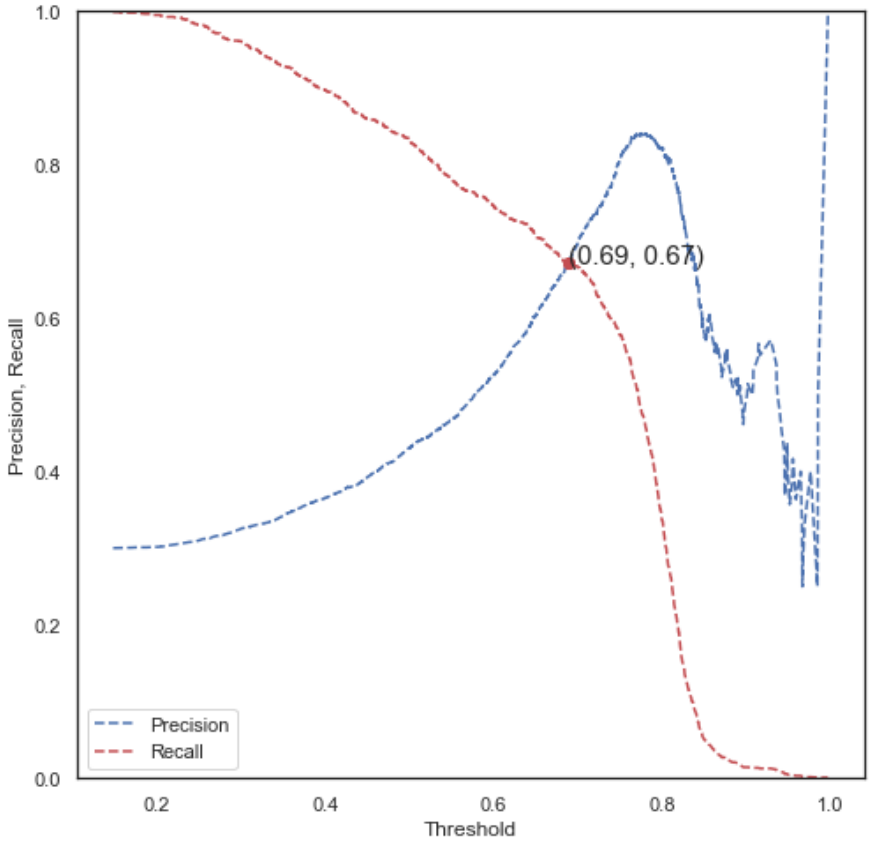


Figure 13. Precision-recall versus discrimination threshold

Using the designed discrimination threshold of 0.69, the confusion matrix is computed for the test set in Table 9. According to this confusion matrix, the model accuracy was calculated at 80%, and the MCC was equal to 0.53. An MCC equal to 1

reflects a perfect prediction, a 0 value represents a random prediction, and -1 reflects an inverse prediction. The model's predictive strength was therefore satisfactory.

Table 9: Confusion matrix with the 0.69 probability threshold

| | Predicted non-failure | Predicted failure |
|---|---|---|
| True non-failure | 2,526 | 411 |
| True failure | 411 | 845 |

***Predictions***

Following the 69% probability threshold, predictions can be made for future failures. A pipe age can be set at a year of interest, and predictions would be made for a time interval starting in that year. By setting the year at 2019, the specified logistic regression model generates failure and survival predictions for the period from 2019 to 2023 as shown in Figure 14.

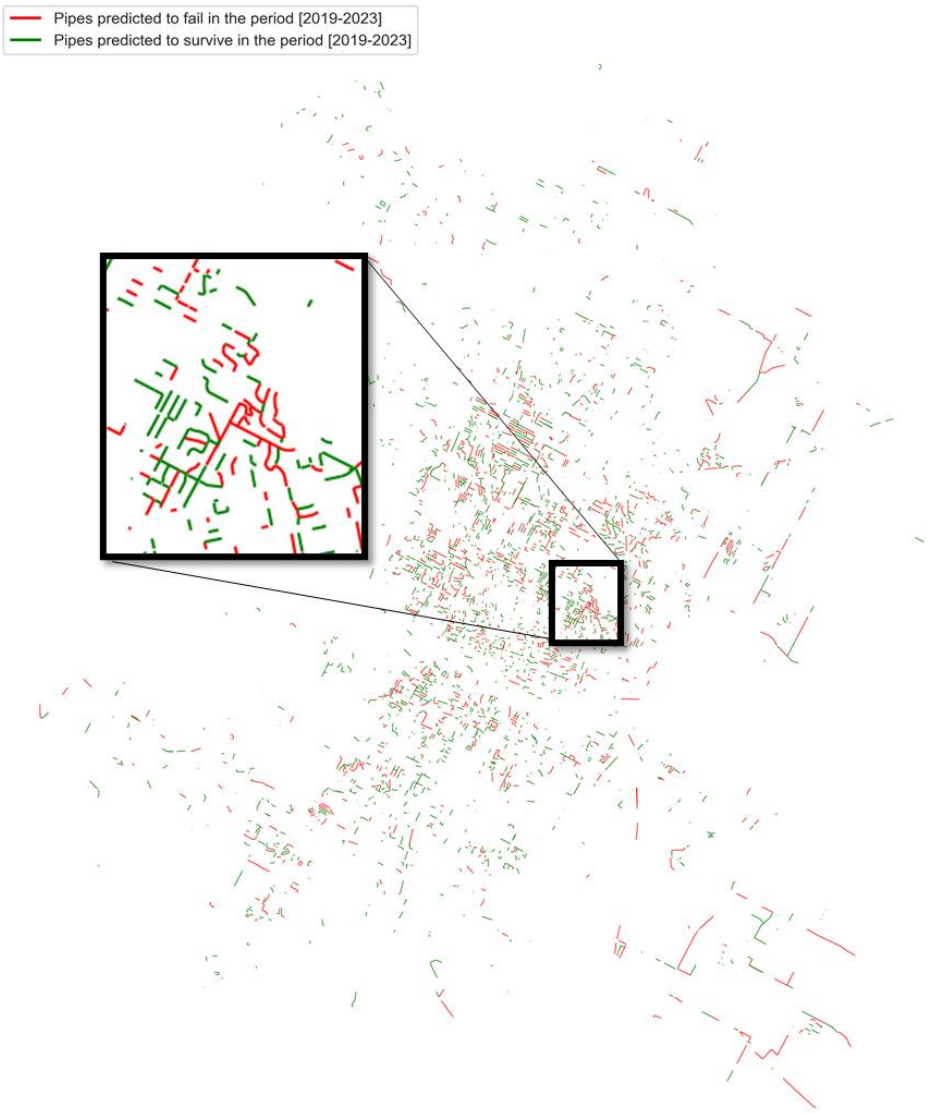


Figure 14: Map of predicted failures in the period from 2019 to 2013

According to the predictions, 44.6% of the network length corresponds to pipes that are predicted to fail at least once during the designated period. This proportion is roughly in line with the overall failure rate in the dataset. 6,769 failure events spanning 20 years are equivalent to 1,692.25 failures in a 5-year period. The studied portion of the network represents 4,153 pipes, which therefore yields an average 40.7% 5-year failure

rate assuming a constant rate across the network. It can also be noted that there is no distinguished spatial concentration of failure predictions despite a few scattered areas.

Because the water utility in this study plans its Capital Improvement Programs on a 5-year basis, such 5-year predictions incorporate information from deterioration factors to provide a practical measure of risk for the decision-making process of rehabilitation efforts.

**Mean Time to Failure and condition scoring**

Logistic regression provided failure probabilities for limited time intervals. The MTF equation allowed to further use those probabilities to compute the expected times to failure given the selected covariates of each pipe (as listed in Table 8). Figure 15 shows how the obtained values evolve over time from the previous failure for the entire data set. As can be seen, the expected time to failure is shorter as the time from last failure increases. Also, the MTF average values decrease from around 6 years to below 1 year with decreasing standard deviations. Low uncertainty associated with shorter MTF values for longer elapsed times since last failure reflect pipes with a higher failure probability. It is noteworthy to mention that MTF values do not exceed 12 years, which is induced by a high failure rate in the dataset. In fact, the dataset that was used in this study consisted of only pipes with at least 1 failure event in a 20-year observation period. Consequently, MTF calculations do not reflect normal expected pipe life expectancies in the entire network, but instead give an expected time between failures for pipes with characteristics and a failure history similar to those in the studied dataset.

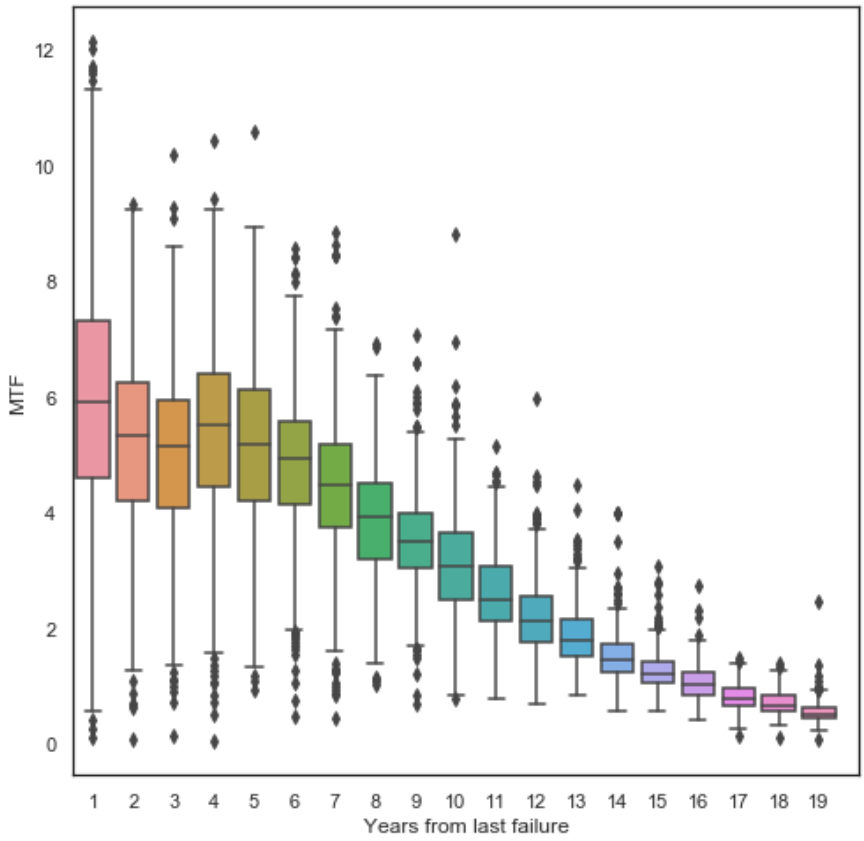


Figure 15. Mean time to failure versus time from last failure.

*Note*. Whiskers are 1.5 times the interquartile range, any data point beyond is considered an outlier.

To validate the estimates, an error was measured as the difference between MTF values for the entire studied dataset and the actual time between failures. The error was calculated for 1,222 pipes that failed in at least two different years, so that the actual time between failures could be measured. As a result, the error had a near normal distribution with a mean equal to 0.54 years and a standard deviation of 3.10 years. The Root Mean Square Deviation (RMSD) associated with the error was equal to 3.29 years. Although an MTF value was in average off by more than 3 years, the near zero mean suggests a

tendency towards correct predictions. A larger sample could potentially reduce the deviation and lead to more accurate MTF estimates.

Finally, the MTF estimates that include useful information about the likelihood of failure were incorporated into a pipe scoring method. Figure 16 shows the condition scores of the pipes as a function of MTF based on the entire studied dataset. These scores were computed using the pipe scoring equation (Eq. 7) by setting the maximum criticality score of 5 and for different discount rates. The maximum criticality score represents an instant failure and was defined such that the scoring scale matched the water utility's scoring scale, ranging from 1 to 5, respectively indicating low and high criticality levels. The choice of the discount rate should reflect a water utility's attitude towards risk, and maintenance and replacement strategy. As can be seen, a higher discount rate leads to a decreased condition score for a given MTF, thus reflecting a propensity to delay rehabilitation efforts by increasing the portion of pipes with low scores.

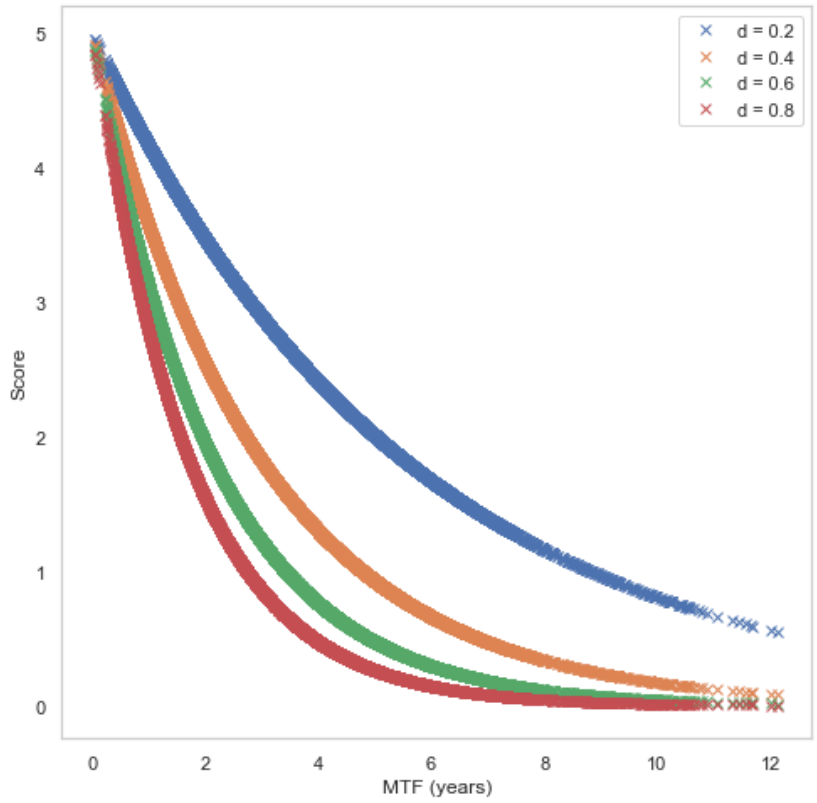


Figure 16. Condition scores as a function of MTF for different discount rates

In order to facilitate rehabilitation, pipes are typically categorized into separate classes by assigning discrete scores instead of continuous ones. Figure 17 shows this categorization as a stepwise pipe scoring curve using a discount rate of 0.2 assigned to each pipe in the studied system.

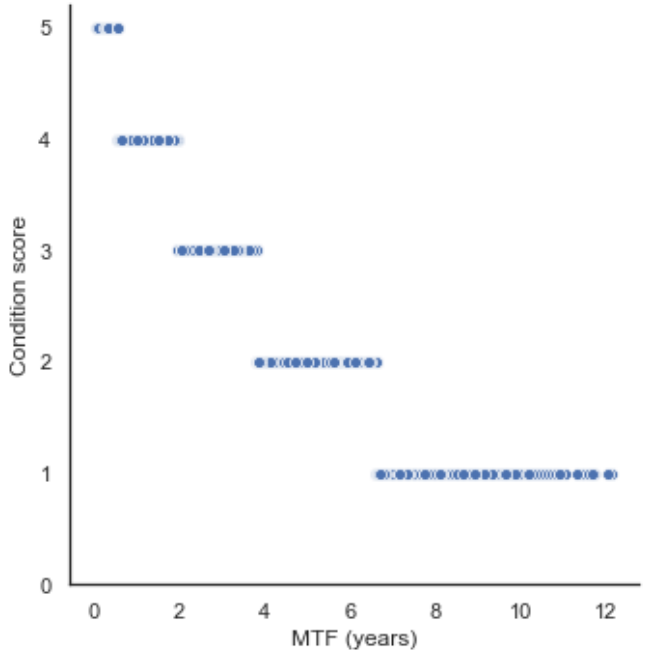


Figure 17. Stepwise scoring curve using a 0.2 discount rate

To evaluate the scoring method, scores were assessed against time to next failure. Only 1,222 pipes failed more than in one year, and true times between failures were measured. Pipes with higher condition scores in general took less time to fail again. This result suggests that assigned condition scores can give a plausible measure of the criticality for the pipes' condition. An advantage of using pipe scores is the ability to capture the likelihood of failure as inferred from the dataset without specifically estimating time to failure. In fact, this scoring method incorporates pipes covariates, probability of failure, as well as utilities' preferences in a simple and easily interpretable single metric that can be used to rank pipes and prioritize rehabilitation efforts.

While condition scores incorporate how deterioration factors influence failure probability for each pipe, they do not provide a measure of the consequence of failure. Risk assessment methods typically include both criticality and consequence scores when

prioritizing asset management. Yet, to assign an integrated risk score, an advantage of the described condition scoring method is its linear scale. As explained in [1], because condition scores are considered a present value of a future failure event based on a chosen discount rate, the scoring scale can be considered linear. For example, a pipe with a condition score of 4 is twice as critical as a pipe with a score of 2. A risk score can thus be simply obtained by multiplying the assigned condition score by a consequence score. The resulting risk score can eventually be used to rank pipes per risk level [1].

As in MTF calculation, assigned scores were also updated each time an annual failure was recorded. By using the last assigned scores, a water utility can visualize the criticality of its pipes. Figure 18 displays a map of the city's water distribution network by categorizing pipes based on the latest scores assigned in the dataset. This condition scoring map can be easily integrated in any spatial and hydraulic software, e.g., ArcGIS, WaterGems, InfoWater, KYPipe, which are commonly used by water utilities and shared among different divisions involved in pipe condition assessment, including operations, planning and management, and asset management.

By analyzing the proportions of network length per condition score, it is noted that 8.4% of the studied pipe network's length has a score of 5, 29.6% has a score 4, 28.2% has a score of 3, 29.5% with a score of 2, and 4.4% has a score of 1. These proportions depend partly on the chosen discount rate. By increasing the discount rate, more pipes would have lower scores, and vice versa. If for example a water utility only has a budget to inspect 50% of the pipes with two levels of priorities, a discount rate could be chosen so that score 5 and 4 pipes make up 50% of all scores. Also, out of the portion of the network having a score of 5, 88.9% of the length consisted of pipes with 15 to 19 years elapsed from last failure. This proportion is consistent with the inferred

covariates' effects which suggested that a longer time from last failure leads to higher failure probability.

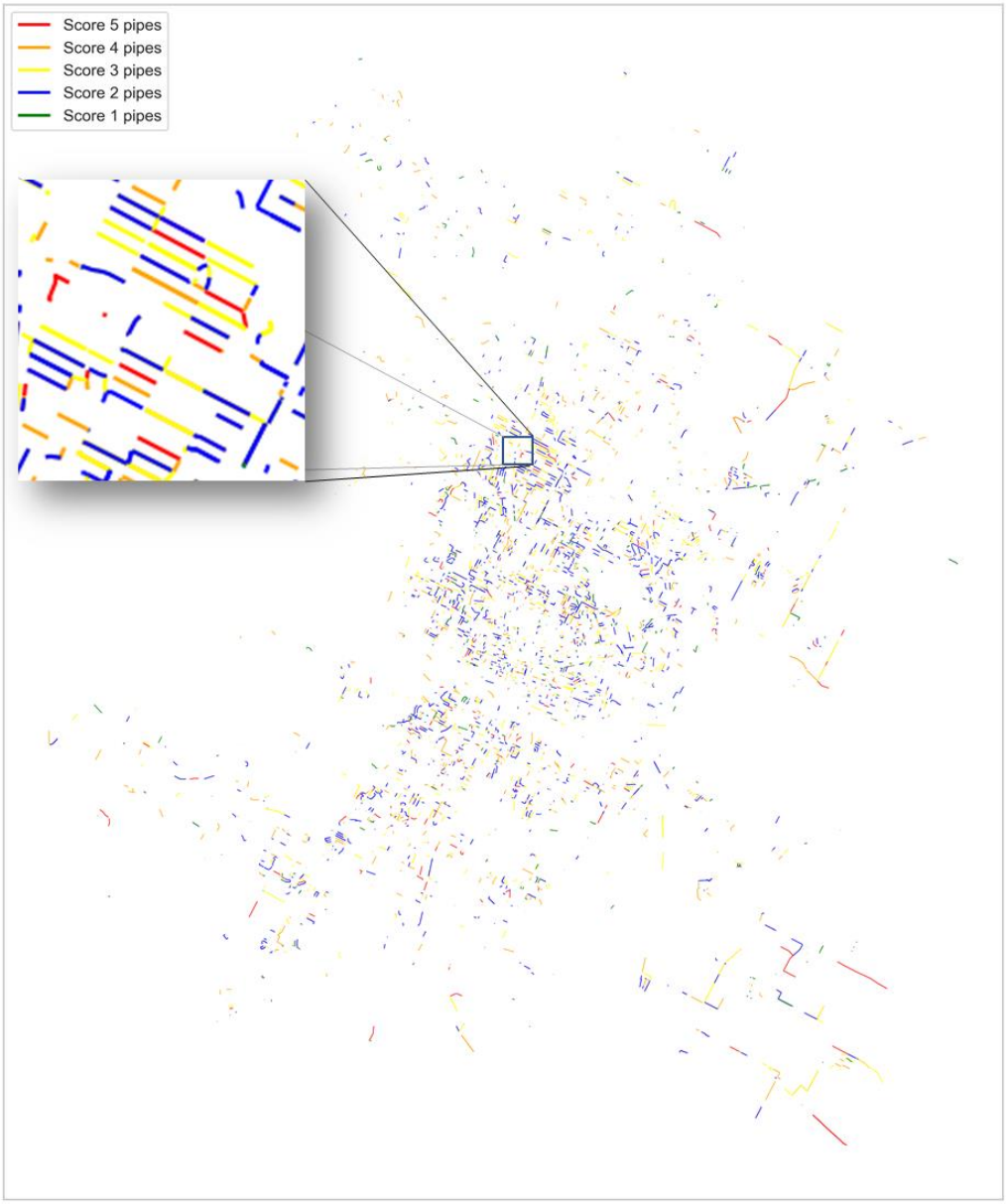


Figure 18. Pipe condition scores as of 2019

**Spatial analysis**

As a final step in the analysis, the focus was on whether the pipe failures exhibited any spatial correlation for higher failure rates. Spatial Autocorrelation Analysis (SAA) was employed to uncover the locations of hotspot/coldspot pipes that encountered a statistically significantly higher/lower number of failures than other pipes in the system. The implemented SAA approach included two main steps, namely global and local SAA. In the first step, the global Moran's I index [55] was computed to reveal the degree of spatial clustering in the pipe failure data. This was followed by computing the local Moran's I index [56] for each individual pipe to reveal its hotspot/coldspot classification. This approach was recently used in [57] to reveal spatial patterns in a water system using aggregated data. In this work, a similar approach is adopted to reveal spatial patterns in individual pipes.

***Global Spatial Autocorrelation Analysis***

The global index of Moran's I was first implemented to determine whether the number of failures recorded for each pipe in the database is spatially autocorrelated on a global scale. If existing, this spatial correlation would reflect either a strong clustering behavior, in which neighboring pipes have similarly high or low number of failures, or a strong dispersion that takes place when neighboring pipes have a vastly dissimilar number of previous failures.

The global Moran's I index was calculated as [58]:

$$I = \frac{N}{W}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(f_i - \bar{f}\right)\left(f_j - \bar{f}\right)}{\sum_{i=1}^{N}\left(f_i - \bar{f}\right)^2} \qquad (8)$$

where $f_i$ and $f_j$ are the number of failures recorded for each two pipes $i$ and $j$ in the pipe failure database, $N$ is the number of pipes for which previous failures have been

observed, $\bar{f}$ is the average number of failures per pipe across all $N$ pipes, $w_{ij}$ is the pairwise spatial weight that represents the spatial relationship between the number of failures recorded for pipes $i$ and $j$; and $W$ is the sum of all the pairwise spatial weights $W = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$. The magnitude of $I$ reflects the strength of the spatial autocorrelation between the number of failures per pipe (the greater the magnitude of $I$, the stronger the autocorrelation), and the sign of $I$ reflects the nature of this clustering ($I > 0$ indicates clustering and $I < 0$ indicates dispersion).

In addition to revealing the strength of the spatial pattern, it is also important to verify whether this pattern is statistically significant compared to what could result from spatial randomness. To this end, the value of the global $I$ was converted into a standardized Z-score, which was calculated as $Z = (I - \mu[I])/\sigma[I]$, where $\mu[I]$ and $\sigma[I]$ are the mean and standard deviation of the distribution of $I$ values representing spatial randomness. The greater the magnitude of the standardized $Z$, the more statistically significant the observed spatial pattern since this indicates that the calculated $I$ is very far from the expected value under spatial randomness, which means that the pipe failures are spatially distributed in a unique, non-random manner. The sign of $Z$ also indicates the type of this pattern, whether it is spatial clustering ($Z > 0$) or spatial dispersion ($Z < 0$). The distribution of $I$ values under spatial randomness (represented by $\mu[I]$ and $\sigma[I]$) was generated using a numerical random permutation approach. This was done by randomly shuffling the number of failures per pipe $f_i$ across all pipes for a number of $k$ permutations ($k$ = 999 in this study) and recalculating the value of $I$ for each permutation to generate the distribution.

***Local Spatial Autocorrelation Analysis***

After revealing the type of the spatial patterns exhibited by the pipe failures and validating that the patterns are statistically significant compared to spatial randomness, the local index of Moran's I was computed for each pipe to reveal its hotspot/coldspot classification. Hotspots/coldspots are defined here as pipes with an above/below average number of previous failures that are surrounded by neighboring pipes with a similarly large/small number of previous failures. To reveal this classification, the local Moran's index for each pipe ($I_i$) was calculated as in Eq. 9 [56]:

$$I_i = \frac{f_i - \bar{f}}{S^2} \sum_{j=1, j \neq i}^{N} w_{ij}\left(f_j - \bar{f}\right) \qquad (9)$$

where $S^2$ is the variance of the number of failures for all pipes. For each pipe, the value of the local $I_i$ reveals whether the number of failures recorded for this pipe is correlated with the number of failures recorded for neighboring pipes. A positive correlation ($I_i > 0$) indicates that the pipe is part of a spatial cluster of similarly high/low number of failures, while a negative correlation ($I_i < 0$) indicates that the pipe is a spatial outlier that is surrounded by neighboring pipes with a dissimilar number of failures.

Upon identifying the pipes that belong to spatial clusters ($I_i > 0$) and the ones that are spatial outliers ($I_i < 0$), the hotspot/coldspot classification was revealed by comparing the number of failures for the pipe $f_i$ and its spatial lag $l_i$, which is the weighted average of the number of failures for neighboring pipes ($l_i = \sum_{j=1, j \neq i}^{N} w_{ij} f_j$), with the average number of failures across all pipes $\bar{f}$. Pipes with an above average number of failures ($f_i > \bar{f}$) that are also surrounded by neighboring pipes with an above average number of failures ($l_i > \bar{f}$) are classified as a hotspots. Conversely, pipes with both ($f_i < \bar{f}$) and ($l_i < \bar{f}$) are classified as coldspots. On the other hand, pipes with a high number of failures ($f_i > \bar{f}$) but are surrounded by neighboring pipes with a low number of

failures ($l_i < \bar{f}$) are considered high-low outliers. Conversely, pipes with ($f_i < \bar{f}$) and ($l_i > \bar{f}$) are classified as low-high outliers.

The statistical significance of the local $I_i$ computed for each pipe was tested by using a conditional permutation approach similar to the one used to test the statistical significance of the global $I$. This was done by generating a reference distribution of $I_i$ for each pipe by holding the $f_i$ value fixed and randomly shuffling the values of $f_j$ across the rest of the pipes for $k$ permutations. A p-value was then computed for each pipe as $p_i = (m + 1)/(k + 1)$, where $m$ is the number of instances from the generated distribution that are greater in magnitude than the computed $I_i$ index for the pipe. The smaller the value of $p_i$, the higher the statistical significance of the observed spatial pattern. The p-values were corrected for multiple comparisons by means of the False Discovery Rate (FDR) method of Benjamini and Hochberg, (1995), and a significance level of 0.05 was considered.

***Spatial Weights***

The pairwise spatial weights ($w_{ij}$) needed to calculate both the global $I$ and the local $I_i$ are determined using the following procedure: (I) the pairwise Euclidian distances were computed between the centroids of each pair of pipes in the dataset ($R_{i,j}$); (II) Using a cutoff threshold distance ($R_t$), any two pipes whose ($R_{i,j} < R_t$) were considered neighbors; (III) For neighboring pipes, the pairwise spatial weight ($w_{ij}$) was computed as the ratio between $R_t$ and $R_{i,j}$ ($w_{i,j} = R_t/R_{i,j}$), while for non-neighboring premises, the pairwise spatial weight was set to zero; (IV) the pairwise spatial weights were row-standardized so that the standardized spatial weights for each premise sum up to unity. A threshold distance of 1000 ft was used in this study, which is nearly equivalent to the

95th-percentile of the lengths of the pipes in the database. This threshold distance is also representative of the scope of pipe repair/replacement projects for the utility under study, which typically span 2-4 residential blocks at a time.

***Identifying spatial autocorrelation***

By analyzing the spatial distribution of failure events, the global SAA revealed a statistically significant clustering in the pipe failure data ($I = 0.06, Z = 5.3$). The local hotspot-coldspot analysis revealed the classification of 125 hotspot pipes (i.e., High-High (HH)) and 39 coldspot pipes (i.e., Low-Low (LL)), while the number of spatial outliers was 182 High-Low (HL) pipes and 166 Low-High (LH) pipes. The remaining 3,707 pipes had a non-statistically significant spatial pattern ($p_i > 0.05$). The locations of the spatial clusters made up of HH and LL pipes, and outliers consisting of HL and LH pipes, are depicted in Figure 1.

On average, hotspot pipes experienced 3.4 failures/pipe, which is almost twice the average number of failures per pipe across all pipes in the database (1.65 failures/pipe). Similarly, high-low pipes experienced 2.6 failures/pipe, which is around 60% more than the average number of failures per pipe across all pipes. On the other hand, both the coldspot and the low-high pipes experienced -on average- only one failure during the 20-year observation period.

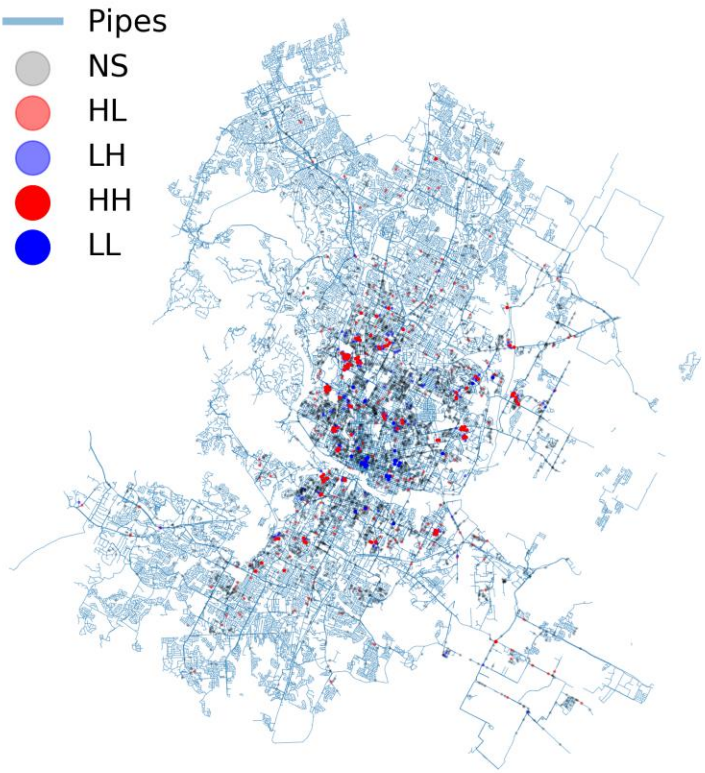


Figure 19. Hotspot-coldspot classification

The hotspot-coldspot classification showed a relative clustering of failure probability such that hostpot pipes consisting of HH and HL labels seemed to indicate a general tendency towards higher failure probabilities as opposed to LH and LL pipes. This relationship is illustrated in Figure 20.

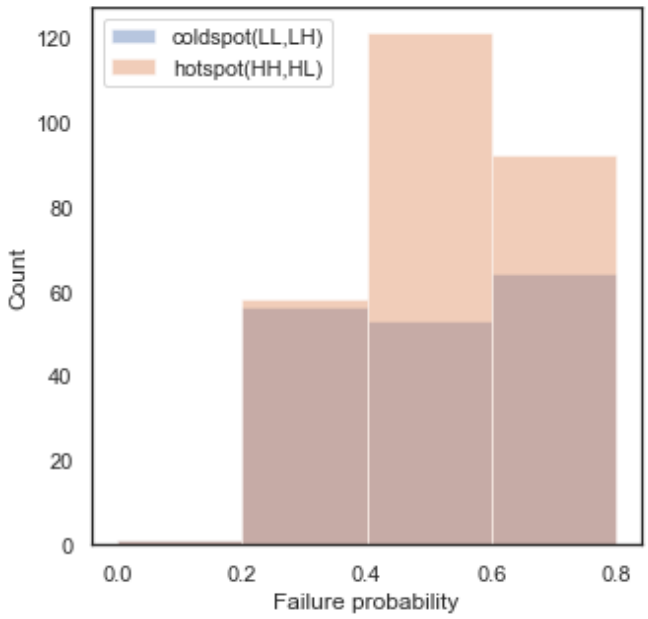


Figure 20. Histogram of failure probability per hotspot-coldspot classification

When analyzed per spatial clustering, the histogram of scores is plotted per hotspot-coldspot classification as shown in Figure 21, it can be seen that scoring criticality is not systematically higher for hotspot pipes. In particular, score 3 represents the mode for coldspot pipes, whereas score 2 is the mode for hotspot pipes. Also, no hotspot pipe was assigned the highest score of 5 which consisted mostly of pipes with a failure probability of at least 0.9. This contrast might question whether condition criticality for a pipe spatially correlates with failure history. However, hotspot-coldspot classification was only significant for a small portion of the pipes with low Moran's I value. Having a failure history covering a longer period could potentially provide a larger basis to evaluate the correlation of spatial patterns with pipe-level failure as well as include this information in the prediction model.

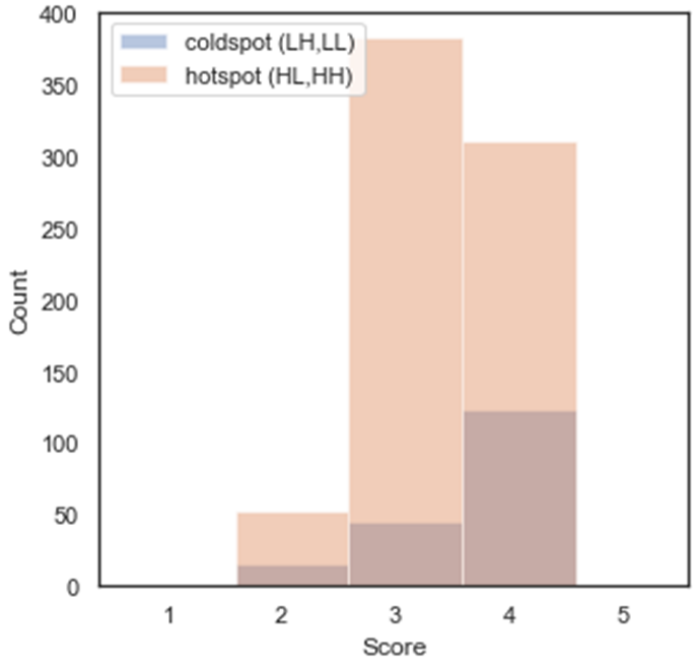


Figure 21. Histogram of scores per hotspot-coldspot classification

# Chapter 4: Conclusions

To address pipe failure modeling, this thesis first presented a review of major published research as related to the scope of this work. Relevant research on identifying factors that influence deterioration was presented. Typically, pipe age, length, diameter, and material have been widely documented as statistically significant contributors in the pipe deterioration processes, but a thorough characterization of the influences is still needed. To further specify the relationship between these factors and failure outcomes, this thesis reviewed a range of published approaches and deterioration models categorized as physical or statistical. Physical models study the interaction on a pipe level and require detailed attributes in an attempt to fully specify the extent of the relationship between a set of factors and a condition. These models provide more of an in-depth characterization of concepts like corrosion and structural strength but often require costly information. On the other hand, statistical models study deterioration patterns on a higher level by investigating relationships across an entire dataset. These models only capture patterns observed across a network and assume that past records fully inform future observations, but they provide practical network-wide information for water utilities to act on.

In addition to a review of previous research, the intended contribution of this thesis was to develop and apply a systematic approach to capture the criticality of pipes in a water supply system using GEE logistic regression and to assign practical condition scores for asset management prioritization. A pipe network dataset was first preprocessed to define a T-year failure response variable and extract covariates that provide information on soil, traffic, land use, failure history, and operational attributes. A GEE logistic regression model was then specified with reasonable accuracy in estimating the

probability of recording at least one failure in a 5-year time interval. Beyond a measure of a period specific criticality for pipes as provided by the logistic classifier, the MTF metric estimated the expected inter-failure times. The estimates were used to apply a flexible scoring approach to discriminate pipes based on their criticality. The pipe scoring provided condition metrics with a reasonable ability to predict poor condition.

The presented promising results would still need further validation with larger datasets. An accuracy of 80% was achieved by the logistic classifier, but specifying the model on failure records covering a period longer than 20 years might mitigate the uncertainty related to the described performance metrics. Also, the MTF calculations use a fundamental assumption that past trend perpetuates. Because failure history is used only from the last 20 years, the model does not provide a full simulation of a pipe's life cycle. As a result, accuracy is bound to decline as predictions are made farther into the future. Also, uncertainty underlying the logistic regression model is accumulated as the MTF calculations integrates probabilities infinitely into the future. The choice of the time-interval in the logistic model is also a factor that influences this uncertainty. It follows that failure probabilities generated by the logistic regression model are conceptually generated with a higher performance compared to pipe scores. However, failure probabilities only provide information on a period specific condition, whereas pipe scores attempt to additionally capture a practical measure of the service life. These limitations in the application of this methodology might justify for a water utility to choose between using probability outcomes or pipe scores depending on the applications. For example, a water utility that prepares a 5-year rehabilitation plan could use the 5-year failure probabilities as a measure of criticality. However, using 5-year failure probabilities might not suffice in integrating criticality in a long-term rehabilitation strategy, and the suggested pipe condition scores would then be more relevant.

The suggested framework demonstrates that useful results can be inferred using a GEE logistic model on a dataset covering a limited time interval and suffering potential censorship. Overall, the proposed methods provided two practical outcomes: (1) a predictive logistic regression model to help prioritize rehabilitation for a specific time interval that is determined based on the quality of the dataset and on the utility's preference, and (2) an integrated condition scoring model to estimate pipe criticality. Future research could further assess the performance of the presented model by using larger and high-quality datasets as they become available. Also comparing the logistic regression model to other statistical and data-driven models could provide further analysis of the performance. Beyond a classical performance evaluation, this thesis intended to provide a flexible framework that can adapt to real world complexity that water utilities have to contend with. Research has shown that deterioration patterns can be region-specific, and results may differ per local conditions [5]. So, developing models that not only deliver good performance but also allow for flexible application is to be further explored.

# References


[1] Opila MC, Attoh-Okine N. Novel Approach in Pipe Condition Scoring. J Pipeline Syst Eng Pract 2011;2:82–90. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000081.

[2] ASCE. Infrastructure Report Card. 2017.

[3] Kleiner Y, Rajani B. Comprehensive review of structural deterioration of water mains: Statistical models. Urban Water, vol. 3, Elsevier; 2001, p. 131–50. https://doi.org/10.1016/S1462-0758(01)00033-4.

[4] Barton NA, Farewell TS, Hallett SH, Acland TF. Improving pipe failure predictions: Factors effecting pipe failure in drinking water networks. Water Res 2019;164:114926. https://doi.org/10.1016/j.watres.2019.114926.

[5] Folkman S. Water Main Break Rates In the USA and Canada: A Comprehensive Study. 2018.

[6] Rajani B, Kleiner Y. Comprehensive review of structural deterioration of water mains: Physically based models. Urban Water, vol. 3, Elsevier; 2001, p. 151–64. https://doi.org/10.1016/S1462-0758(01)00032-2.

[7] Gould SJF, Boulaire FA, Burn S, Zhao XL, Kodikara JK. Seasonal factors influencing the failure of buried water reticulation pipes. Water Sci Technol 2011;63:2692–9. https://doi.org/10.2166/wst.2011.507.

[8] Mackey T, Cashman A, Cumberbatch R. Identification of factors contributing to the deterioration and losses in the water distribution system in Barbados. Paris: UNESCO. - Google Search. 2014.

[9] Sattar AMA, Ertuğrul ÖF, Gharabaghi B, McBean EA, Cao J. Extreme learning machine model for water network management. Neural Comput Appl 2019;31:157–69. https://doi.org/10.1007/s00521-017-2987-7.

[10] Berardi L, Giustolisi O, Kapelan Z, Savic DA. Development of pipe deterioration models for water distribution systems using EPR. J Hydroinformatics 2008;10:113–26. https://doi.org/10.2166/hydro.2008.012.

[11] Jun HJ, Park JK, Bae CH. Factors Affecting Steel Water-Transmission Pipe Failure and Pipe-Failure Mechanisms. J Environ Eng 2020;146:04020034. https://doi.org/10.1061/(asce)ee.1943-7870.0001692.

[12] Wang Y, Zayed T, Moselhi O. Prediction models for annual break rates of water mains. J Perform Constr Facil 2009;23:40–6. https://doi.org/10.1061/(ASCE)0887-3828(2009)23:1(47).

[13] Bruaset S, Sægrov S. An Analysis of the Potential Impact of Climate Change on the Structural Reliability of Drinking Water Pipes in Cold Climate Regions. Water 2018;10:411. https://doi.org/10.3390/w10040411.

[14] Kettler AJ, Goulter IC. ANALYSIS OF PIPE BREAKAGE IN URBAN WATER DISTRIBUTION NETWORKS. Can J Civ Eng 1985;12:286–93. https://doi.org/10.1139/l85-030.

[15] Boulos PF, Karney BW, Wood DJ, Lingireddy S. Hydraulic Transient Guidelines for Protecting Water Distribution Systems. J Am Water Works Assoc 2005;97:111–24. https://doi.org/10.1002/j.1551-8833.2005.tb10892.x.

[16] Zheng Liu, Yehuda Kleiner, Balvant Rajani. Condition Assessment Technologies for Water Transmission and Distribution Systems. 2012.

[17] Scheidegger A, Leitão JP, Scholten L. Statistical failure models for water distribution pipes - A review from a unified perspective. Water Res 2015;83:237–47. https://doi.org/10.1016/j.watres.2015.06.027.

[18] Eisenbeis P. Modelisation statistique de la prevision des defaillances des conduites d'eau potable 1994.

[19] Gustafson J, conference DC-P of the A annual, 1999 undefined. Modeling the occurrence of breaks in cast iron water mains using methods of survival analysis n.d.

[20] Pelletier G. Impact du remplacement des conduites d'aqueduc sur le nombre annuel de bris. 2000.

[21] Røstum J. STATISTICAL MODELLING OF PIPE FAILURES IN WATER NETWORKS. Fakultet for ingeniørvitenskap og teknologi; 2000.

[22] Watson TG, Christian CD, Mason AJ, Smith MH, Meyer R. Bayesian-based pipe failure model. J Hydroinformatics 2004;6:259–64. https://doi.org/10.2166/hydro.2004.0019.

[23] Economou T, Kapelan Z, Bailey T. A zero-inflated bayesian model for the prediction of water pipe bursts. Proc. 10th Annu. Water Distrib. Syst. Anal. Conf. WDSA 2008, 2009, p. 724–34. https://doi.org/10.1061/41024(340)61.

[24] Le Gat Y. Une extension du processus de Yule pour la modélisation stochastique des événements récurrents : application aux défaillances de canalisations d'eau sous pression. Doctorat AgroParisTech - Sciences de l'eau (Option Statistique); 2009.

[25] Kleiner Y, Rajani B. I-WARP: Individual Water mAin Renewal Planner. Drink Water Eng Sci 2010;3:71–7. https://doi.org/10.5194/dwes-3-71-2010.

[26] Scheidegger A, Scholten L, Maurer M, Reichert P. Extension of pipe failure models to consider the absence ofdata from replaced pipes. Water Res 2013;47:3696–705. https://doi.org/10.1016/j.watres.2013.04.017.

[27] Ana E, Bauwens W, Pessemier M, Thoeye C, Smolders S, Boonen I, et al. An investigation of the factors influencing sewer structural deterioration. Urban Water

J 2009;6:303–12. https://doi.org/10.1080/15730620902810902.

[28] Ariaratnam ST, El-Assaly A, Yang Y. Assessment of Infrastructure Inspection Needs Using Logistic Models. J Infrastruct Syst 2001;7:160–5. https://doi.org/10.1061/(ASCE)1076-0342(2001)7:4(160).

[29] Friedl F, Möderl M, Rauch W, Schrotter S, Liu Q, Fuchs-Hanusch D. Failure Propagation for Large-Diameter Transmission Water Mains Using Dynamic Failure Risk Index 2012:3082–95.

[30] Yamijala S, Guikema SD, Brumbelow K. Statistical models for the analysis of water distribution system pipe break data. Reliab Eng Syst Saf 2009;94:282–93. https://doi.org/10.1016/j.ress.2008.03.011.

[31] Debón A, Carrión A, Cabrera E, Solano H. Comparing risk of failure models in water supply networks using ROC curves. Reliab Eng Syst Saf 2010;95:43–8. https://doi.org/10.1016/j.ress.2009.07.004.

[32] Robles-Velasco A, Cortés P, Muñuzuri J, Onieva L. Prediction of pipe failures in water supply networks using logistic regression and support vector classification. Reliab Eng Syst Saf 2020;196:106754. https://doi.org/10.1016/j.ress.2019.106754.

[33] Salman B, Salem O. Modeling Failure of Wastewater Collection Lines Using Various Section-Level Regression Models. J Infrastruct Syst 2012;18:146–54. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000075.

[34] Laakso T, Kokkonen T, Mellin I, Vahala R. Sewer Condition Prediction and Analysis of Explanatory Factors. Water 2018;10:1239. https://doi.org/10.3390/w10091239.

[35] Sousa V, Matos JP, Matias N. Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition. Autom Constr 2014;44:84–91. https://doi.org/10.1016/j.autcon.2014.04.004.

[36] Kleiner Y, Rajani B. Comparison of four models to rank failure likelihood of individual pipes. J Hydroinformatics 2012;14:659–81. https://doi.org/10.2166/hydro.2011.029.

[37] Fuchs-Hanusch D, Günther M, Möderl M, Muschalla D. Cause and effect oriented sewer degradation evaluation to support scheduled inspection planning. Water Sci Technol 2015;72:1176–83. https://doi.org/10.2166/wst.2015.320.

[38] Cooper NR, Blakey G, Sherwin C, Ta T, Whiter JT, Woodward CA. The use of GIS to develop a probability-based trunk mains burst risk model. Urban Water 2000;2:97–103. https://doi.org/10.1016/S1462-0758(00)00047-9.

[39] Wilson D, Filion Y, Moore I. State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. Urban Water J 2017;14:173–84. https://doi.org/10.1080/1573062X.2015.1080848.

[40] Birolini A. Reliability engineering: Theory and practice, Seventh edition. Springer Berlin Heidelberg; 2014. https://doi.org/10.1007/978-3-642-39535-2.

[41] Myung IJ. Tutorial on maximum likelihood estimation. J Math Psychol 2003;47:90–100. https://doi.org/10.1016/S0022-2496(02)00028-7.

[42] Hardin JW, Hilbe JM. Generalized Estimating Equations, Second Edition. 2012.

[43] Wang M. Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments 2014. https://doi.org/10.1155/2014/303728.

[44] Liang K-Y, Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models. vol. 73. 1986.

[45] Pan W. Akaike's Information Criterion in Generalized Estimating Equations. vol. 57. 2001.

[46] Tibshirani R. Regression Shrinkage and Selection via the Lasso. vol. 58. 1996.

[47] Louw N, Steel SJ. Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination. Comput Stat Data Anal 2006;51:2043–55. https://doi.org/10.1016/j.csda.2005.12.018.

[48] Giraldo-González MM, Rodríguez JP. Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks. Water (Switzerland) 2020;12:1153. https://doi.org/10.3390/W12041153.

[49] Davis J, Goadrich M. The Relationship Between Precision-Recall and ROC Curves. Proc. 23 rd Int. Conf. Mach. Learn., Pittsburgh, PA: ACM Press; 2006. https://doi.org/10.1145/1143844.

[50] Wright GB. Radial Basis Function Interpolation: Numerical and Analytical Developments 2003.

[51] USDA. Web Soil Survey. Nat Resour Conserv Serv United States Dep Agric 2015:http://websoilsurvey.nrcs.usda.gov/.

[52] Staff SS. Soil taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys. U.S. Department of Agriculture Handbook 436. 2nd edition. United States Dept. of Agriculture, Naturel Resources Conservation Service; 1999.

[53] TxDOT Open Data Portal n.d. http://gis-txdot.opendata.arcgis.com/ (accessed June 11, 2020).

[54] Japkowicz N, Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intell DATA Anal 2002:429--449.

[55] Moran PAP. Notes on Continuous Stochastic Phenomena. Biometrika 1950;37:17. https://doi.org/10.2307/2332142.

[56] Anselin L. Local Indicators of Spatial Association—LISA. Geogr Anal

1995;27:93–115. https://doi.org/10.1111/J.1538-4632.1995.TB00338.X.

[57] Abokifa AA, Sela L. Identification of Spatial Patterns in Water Distribution Pipe Failure Data Using Spatial Autocorrelation Analysis. J Water Resour Plan Manag 2019;145:04019057. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001135.

[58] Goodchild MF. Spatial Autocorrelation. Concepts and Techniques in Modern Geography 47. Geo Books 1986;47.

[59] Y Benjamini YH. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995;57:289–300.