Sammendrag

Hyppige rørfeil og vannlekkasje utgjør en stor trussel for vannverkenes arbeid med å sikre en kontinuerlig og tilstrekkelig vannforsyning. Hydraulisk trykk som drivkraft i vanndistribusjonsnett regnes som en viktig driver ved rørbrudd og vannlekkasje og blir sett på som en viktig variabel for å bestemme forventet levetid for rør. Forventet levetid for rør og rørfeil henger sammen gjennom en kompleks sammenheng mellom en rekke fysiske og miljømessige faktorer. Når det gjelder lekkasje, er den hovedsakelig relatert til størrelsen på åpningen eller sprekken, trykket, jordforholdene og elastisiteten til rørmaterialet. I denne studien ble disse relasjoner trykk– forventet levetid – lekkasje undersøkt ved hjelp av en real-verden case-studie i Bergen. Flere regresjonsmodeller ble utviklet for å undersøke sammenhengen mellom trykk og forventet levetid for rør. Et stort datasett, som inkluderer servicerør og hovedrør som er rammet av ett eller flere brudd tidligere i Bergen, ble samlet inn for modellutvikling og validering. I tillegg ble det satt opp en hydraulisk modell for å undersøke sammenhengen mellom trykk og vannlekkasje ved å regulere trykket ved hjelp av ulike reguleringsmetoder. Data fra SCADA-systemet ble samlet inn, og komponentene for minimum nattforbruk (MNF) ble beregnet for å utvikle og kalibrere denne hydrauliske modellen for et utvalgt district metered area (DMA) i Bergen. Eiendoms- og adressedata ble brukt for en optimal fordeling av forbruket, og lekkasje ble modellert som en trykkavhengig komponent. Resultatene fra denne studien viser en direkte sammenheng mellom på den ene siden trykk og lekkasje, og trykk og forventet levetid på røret på den andre. Disse funnene kan forbedre effektiviteten til algoritmer for prioritering av rørutskifting og renovering. De kan også hjelpe til med å implementere passende trykkstyringsmetoder for å redusere lekkasje.

Abstract

Frequent pipe failures and water leakage pose a major threat to the water utilities work in securing a continuous and adequate water supply. Hydraulic pressure as a driving force in water distribution networks is considered an important driver in pipe breakage and water leakage and is seen as a major variable in determining the life expectancy of pipes. Pipe-life expectancy and pipe failures are related through a complex correlation between a host of physical and environmental factors. As for leakage, it is mainly related to the size of the opening or crack, the pressure, the soil conditions, and the elasticity of the pipe material. In this study, these relationships pressure – life expectancy – leakage were investigated using a real-world case study in Bergen. Several regression models were developed to investigate the relationship between pressure and pipe-life expectancy. A large data set, that included the services and mains that experienced one or more breaks in the past in the city of Bergen, was collected for model development and validation. In addition, a hydraulic model was set up to investigate the relationship between pressure and water leakage by regulating the pressure using different management methods. Data from the SCADA system were collected, and the minimum night flow (MNF) components were calculated to develop and calibrate this hydraulic model for a selected district metered area (DMA) in Bergen. Property and address data was used for an optimal distribution of consumption, and leakage was modelled as a pressure-dependent component. The results from this study demonstrate a direct relationship between, on the one hand, pressure and leakage, and pressure and pipe-life expectancy, on the other. These findings can improve the efficiency of pipe replacement and renovation prioritization algorithms. They can also aid in deploying appropriate pressure management methods to reduce leakage.

Acknowledgments

First and foremost, I am deeply grateful to my supervisor Dr. Franz Tscheikner-Gratl and second supervisor Dr. Marius Møller Rokstad for their valuable guidance, support, and insightful feedback throughout the entire process.

Furthermore, I would like to thank Bergen Vann for making it possible to study beside my work and thank you to all my colleagues. Special mention to my colleagues Tor-Arne Torsvik and Dalibor Vlaski. Thank you for the exiting and useful discussions.

In addition, I would like to thank my beloved wife Diana. You have been amazing; I couldn't have done this without you. Finally, I cannot forget to thank my family and friends for their encouragement and support in the past few years.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAS	Asbestos cement
APP	Average pressure point
CDP	Critical demand point
CPP	Critical Pressure point
DEM	Digital elevation model
DMA	District metered area
GA	Genetic algorithm
HDPE	High-density polyethylene
LDPE	Low-density polyethylene
MCU	Copper
MGA	Galvanized steel
MNF	Minimum night flow
PE	Polyethylene
PMA	Pressure managed area
PRV	Pressure reduction valve
PVC	Polyvinyl chloride
SCADA	Supervisory control and data acquisition
SJG	Grey cast iron
SJK	Ductile iron
WDN	Water distribution network

1 Introduction

Global freshwater demand is increasing at a rate of one percent per year due to population growth, and a global water deficit of 40% is expected by 2030 (UNESCO, 2021). Approximately 71% of freshwater is used in agriculture, 12% in industry and 17% for domestic uses (World Bank, 2017). According to Liemberger and Wyatt (2018), an average of 30% of global freshwater withdrawal is lost on its way to consumers. They estimate this 30% to amount to 126 billion m³ per year. The increasing demand and water loss pose significant challenges to water utilities and threaten their ability to secure an adequate and continuous water supply.

To address this challenge, many research studies in the past two centuries have focused on finding better ways to manage water resources and on deploying more efficient leakage management measures. Lambert and Hirner (2000) introduced a water balance model which basically categorises water consumption into authorised consumption and water losses and then into subcategories such as apparent and real losses. The water balance model has been used worldwide by water utilities to investigate and assess non-revenue water in their water distribution systems (WDN). Further, minimum night flow (MNF) was established as a key performance indicator in leakage detection work. The MNF was divided into various components to be able to differentiate legal consumption from the water lost to leakage (Fantozzi & Lambert, 2010). Also several equations to calculate each component based on easily available data from the utility systems were developed (Lambert, 2009). In addition, Lambert (2000) and May (1997) identified four basic methods to reduce the avoidable leakage component of real losses, which itself is a component of non-revenue water. These methods are **pressure management**, speed and quality of repairs, active leakage control and pipeline and **asset management**.

This thesis will try to quantify the impact of different pressure management practices in a selected pressure-managed DMA in the city of Bergen in Norway. Even though Bergen doesn't face a scarcity issue like many other utilities, it has a leakage rate of 40% of the total water production (Bergen Vann, 2022). This leakage affects the utility significantly when it comes to the costs associated with production, distribution, and storage on top of its environmental impact. At the same time, Bergen must meet a national goal of no more than 20% leakage by 2030 or earlier where Bergen plans to meet this goal by 2028 (Vann og avløpsetaten, 2020, p. 16).

Different approaches to estimating the pressure-leakage relationship have been done in the past and can be categorised into experimental and theoretical groups. Experimental approaches are either made in a real WDN by testing different types of pressure reduction valves (PRV) with different control profiles and monitoring the flow-pressure data in real-time (Fontana et al., 2018) or in a lab using a loop network with a supply pipe (Hoțupan et al., 2019). Theoretical approaches usually represent a real WDN and use a calibrated hydraulic model to run simulations with different scenarios and PRV types and settings (Berardi et al., 2015). Other theoretical approaches may use hydraulic models for totally virtual WDN's or represent real WDN models created long time ago, so they don't represent the current situation in these, like using water distribution systems from the database developed by the Kentucky Infrastructure Authority (Jolly et al., 2014).

The results of theoretical approaches should always be taken with high caution: No matter how extensive the conducted calibration of the hydraulic model is, there will be a degree of error. Error may be caused by inaccuracies in the data entered in the model, such as pipe lengths, diameters, roughness, valve openings, demand estimation and allocation, and others. Theoretical models do however enjoy cost and time advantages over field experiments.

As mentioned, pressure management is only one facet of the leakage problem. It overlaps with the broader field of asset management which tries to avoid more significant impact and harmful consequences like frequent pipe breaks. Such pipe failures can occur because of internal, external or a combination of both internal and external factors (Ghorbanian et al., 2016). Breaks are usually followed by significant financial consequences due to the costs associated with the repair work, environmental consequences such as energy and water loss, and reputational consequences caused by water supply disruptions and traffic delays (Berardi et al., 2008). They also hinder the utility’s ability to improve and develop its networks. In addition, pipe breaks strain the operational capacity of the water utility. In most cases, they delay scheduled activities and require extra labour especially if there are no readily available alternatives for water supply and a need to deploy emergency water tanks arises.

Unlike the water leakage rate, there is no global database or statistics for the rate of pipe breaks or pipe-life expectancy. The main reasons behind this may be that most water utilities prioritise reducing leakage over reducing pipe breaks because of water scarcity, high production, transport and distribution costs or other challenges. In addition, the leakage rate is usually used as an indicator of utility performance and as a national goal, traditionally linked to a deadline. Bergen, which is the case study in this thesis, has a burst rate of 120 to 165 breaks per 1000 km per year (Bergen Vann, 2015-2020). This burst rate is comparable to burst rates from other countries. For example, in the UK it is between 155-185 (Farrow et al., 2017) and in the US it is between 130-170 (Folkman, 2018). Nevertheless, the interruptions and the costs associated with the minimum rate of 120 are significant and preventative measures should be considered.

The largest survey in this field was conducted by Thornton and Lambert (2007) where they managed to collect data from 112 systems in 10 countries. The survey data shows the impact of the reduction in maximum pressure on reducing new pipe breaks. Other forms of pressure indicators have been observed when reviewing previous studies, such as pressure fluctuations (Rezaei, 2017), mean pressure and pressure range (Jara-Arriagada & Stoianov, 2021), maximum pressure (Moslehi & Jalili Ghazizadeh, 2020) and average maximum and average minimum pressure (Ghorbanian et al., 2016). Each of these indicators based on the results of the studies may explain the pressure-pipe break to some extent. However, there is still no general agreement on which indicator is preferable and provides the best explanation.

In addition, water pressure can be found in two different types in the WDNs: steady-state and dynamic. Steady-state pressures are stable over time. In contrast, dynamic pressure is a pressure caused by pressure transients or water hammers, which happen when sudden changes occur in the network, such as valve openings and closings, pump restarts, hydrant flushings, human mistakes, or equipment failure. Dynamic pressures are generally excluded in most of the previously published studies that investigate the pressure-break relationship because of their complex nature (Jara-Arriagada & Stoianov, 2021; Ghorbanian et al., 2016; Moslehi & Jalili_Ghazizadeh, 2020). For the same reasons, this thesis will only consider the steady-state type of pressure.

Different research goals can be noticed when reviewing some of the previous studies of the pressure-break relationship. Some studies tried to estimate the probability of pipe breaks (Jara-Arriagada & Stoianov, 2021; Moslehi & Jalili_Ghazizadeh, 2020; Martínez-Codina et al., 2015), others the rate of pipe breaks (Ghorbanian et al., 2016; Wang et al., 2009) and others still the pipe-life expectancy (Thornton and Lambert 2007). It's also noticeable that the studies on pipe-life expectancy are limited in number and size. The lack of large-scale research on the impact of pressure on pipe-life expectancy considering other influencing factors leads to the necessity of conducting more studies that highlight this impact and determine its size to increase the life of the pipes. On the other hand, the life expectancy indicator may be easier to benefit from compared with other pipe break indicators, especially in pipe rehabilitation and replacement prioritisation algorithms.

The approaches to achieving the goals mentioned above were categorised by Jara-Arriagada & Stoianov (2021) into mechanistic and data-driven. Mechanistic approaches involve extensive field investigations that are costly and very time-consuming. On the other hand, data-driven approaches like statistical, classification and regression analyses provide good results in a short period at negligible cost.

Beyond the effects of pressure, Wang et al. (2009) have highlighted a host of physical and environmental correlates of the life expectancy of pipes and their break rates. Most of the prior research appears to tend to include as many of these factors as possible, including one or several pressure indicators when building pipe break prediction or deterioration models (Robles-Velasco et al., 2020; Wang et al., 2009; Jara-Arriagada & Stoianov, 2021). However, there is always a limitation on how many factors can be included, mainly because of data quality and availability issues.

Pipe-life expectancy as a term has no universal definition, but since pipes are products the term can follow one common definition for product life expectancy. "Product life expectancy is the duration of the life of a product starting from acquisition and ending at the moment of replacement" (Van Nes & Cramer, 2006). Based on this general definition, the life expectancy of a pipe in this thesis refers to the period between pipe installation and pipe burst, and this period can be extended due to repairs and renovations.

Consequently, this thesis will address two questions: (1) what effects modulating water pressure has on avoidable water leakage, and (2) whether and to what extent maximum zonal pressure in conjunction with other parameters is correlated with the life expectancy of pipes in a water distribution system.

2 Methodology

2.1 Avoidable water leakage

The methodology for investigating the pressure leakage relationship will follow the steps shown in Figure 1. It starts with the collection of necessary data from utility systems. A real pressure-managed DMA model will then be built and calibrated using field-verified pipe roughness data, and pressure and flow data from the utility's SCADA system. Next, MNF components' equations will be used to calculate the MNF components. After that, the calculated demands will be distributed and assigned to the addresses of consumers. Different hydraulic scenarios and pressure regulation methods will be considered for the simulation and investigation of the impact of pressure management on avoidable water leakage. The use of real SCADA data, MNF equations, and accurate demand allocation results in a representative hydraulic model of the real WDN that provides results with reasonable accuracy. An overview of the simulation results will then be presented and discussed.

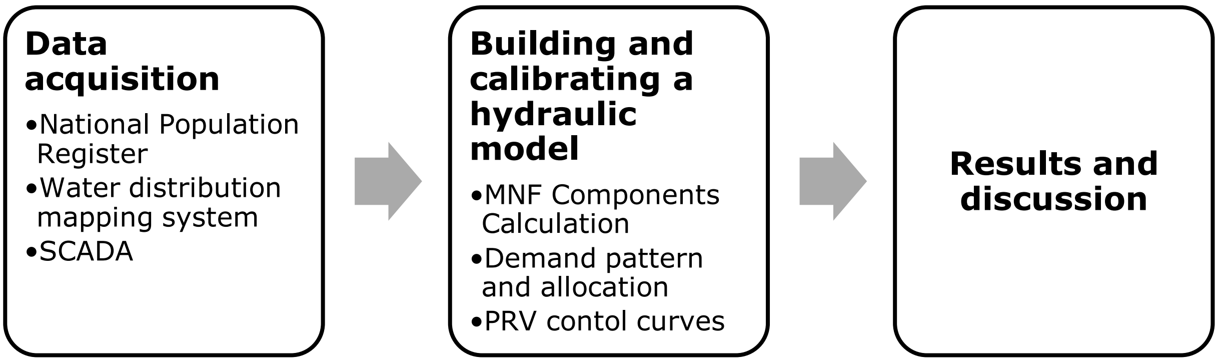


Figure 1 Main steps of the applied methodology for the pressure-leakage relationship investigation

2.1.1 Data acquisition

The following data will be collected:

- Data about the water mains and services (length, diameter, roughness, etc.)
- Data about the water pumps, storage tanks and PRVs
- Flow and pressure data
- The addresses of the domestic and non-domestic properties and their numbers

2.1.2 Minimum night flow

MNF is widely recognized and used as an important indicator in water leak detection work. It is used to determine the size of and to monitor the development of existing leaks, and to detect new leaks.

To maximize the benefit of using MNF, the International Water Association divided MNF into four components to make it easier to focus on avoidable leakage. As shown in Figure 2, these are customer night use, customer night leakage, unavoidable background leakage, and avoidable leakage. IWA has also developed several equations for the calculation of these components (Table 1).

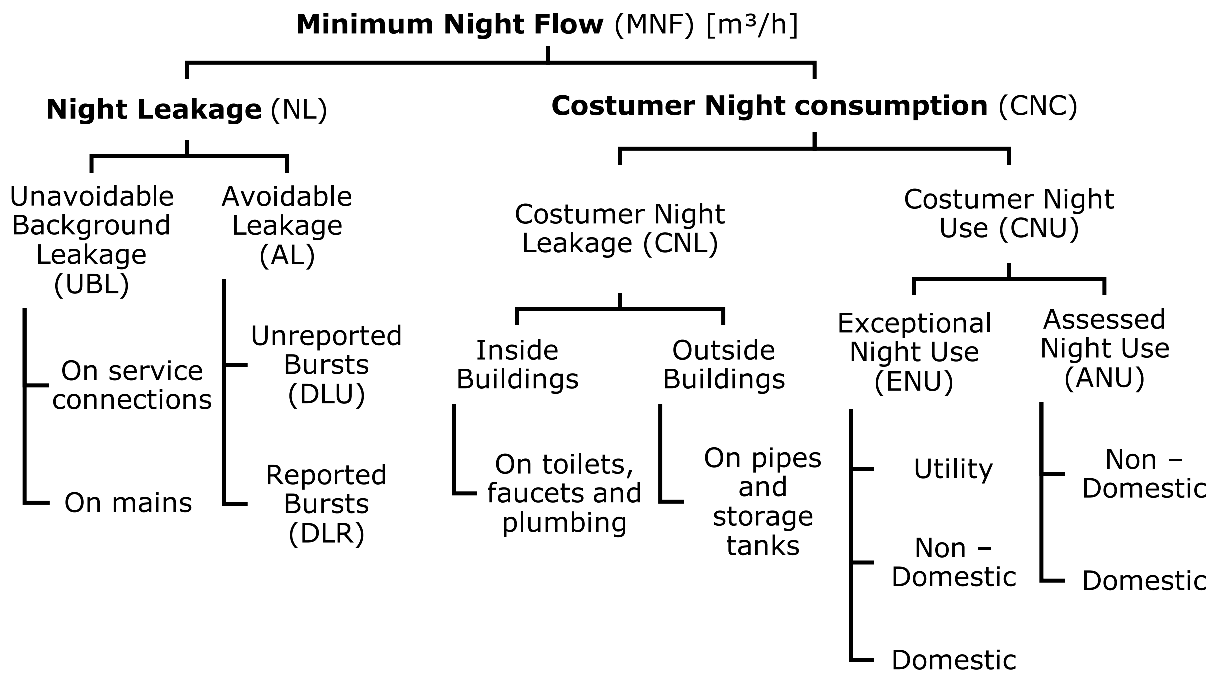


Figure 2 MNF components (Fantozzi & Lambert, 2010)

Table 1 MNF components' equations (Lambert, 2009; Maggs, 2007)

Equation	Variables
$ANU_{Non-Domestic} = NDP * 8$	NDP = number of non-domestic properties
$ANU_{Domestic} = DP * R * 0.06 * CS$	CS = average toilet citern size [liter], DP = [Number of domestic properties], R= average number of residents per property
$ANU = ANU_{Non-Domestic} + ANU_{Domestic}$	ANU = Assessed Night Use
$CNU = ANU + ENU$	CNU = Costumer Night Use
$UBL = (20 * Lm + 1.25 * NC + 0.0333 * Lc) * (\frac{P}{50})^{N_1}$	UBL = unavoidable background leakage [l/h], Lm = length of water mains [km], Nc = number of service connections, lp = average distance from curb stop to customer meter [m], Lc = total length of private pipe, Nc x lp [m], P = average pressure [m], N1= the leakage exponent
$MNF_{Theoretical} = UBL + CNU + CNL$	
$AL = MNF_{Observed} - MNF_{Theoretical}$	AL = Avoidable leakage

2.1.3 PRV control curves

Prior research used several methods to create and optimize PRV control curves (Vicente et al., 2016). AbdelMeguid (2011) categorized these methods into three main categories: linearization methods, nonlinear programming and evolutionary computing and genetic algorithms. The linearization methods and nonlinear programming optimize the outlet pressure of the PRV to minimize pressure-dependent leakage, while the evolutionary computing and genetic algorithms optimize the outlet pressure of the PRV to minimize the pressure head at the network critical point (AbdelMeguid and Ulanicki, 2010). However, these methods share the optimisation problem of minimising background leakage with the minimum allowable pressure set as a common constraint, but they use different approaches to solve it.

This thesis will include two types of PRV control curves: time-modulation and flow-modulation curves. The time-modulation curve will be created empirically by using Mike Urban + as a hydraulic simulator to find PRV settings that maintain constant pressure at the critical point in the DMA over time. The flow-modulation curve will be created and optimized using the curve fitting toolbox in MATLAB and Mike Urban + as a hydraulic simulator. There will be no connection between MATLAB and Mike Urban +, so the hydraulic simulation results will be manually entered into the curve fitting toolbox in contrast to what AbdelMeguid and Ulanicki (2010) did in their application of genetic algorithms where the GA toolbox was connected to EPANET. The curve fitting toolbox has many built-in equations that can be used to optimize the flow modulation curve, such as Polynomial, Fourier, Gaussian, Sum of Sine, and others. A ‘Sum of Sine’ equation with three terms as expressed in Equation (1) is chosen.

$$\boldsymbol{PRV\ Setting} = \boldsymbol{h_x(t)} = \boldsymbol{a_1} * \mathbf{sin}(\boldsymbol{b_1} * \boldsymbol{q_x(t)} - \boldsymbol{c_1}) + \boldsymbol{a_2} * \mathbf{sin}(\boldsymbol{b_2} * \boldsymbol{q_x(t)} - \boldsymbol{c_2}) + \boldsymbol{a_3} * \boldsymbol{sin}(\boldsymbol{b_3} * \boldsymbol{q_x(t)} - \boldsymbol{c_3}) \quad \textbf{(1)}$$

Here, $\boldsymbol{h_x(t)}$ **=** PRV outlet pressure [m] at the PMA entry, $\boldsymbol{q_x(t)}$ **=** the flow [l/s] at the PMA entry at time $\boldsymbol{t}$, and coefficients $\boldsymbol{a_1, b_1, c_1, \ldots}$ are coefficients of the flow modulation curve

It’s worth mentioning that the optimization of the modulation curves will only consider a single-feed scenario for the sake of simplicity and because the chosen DMA has a single inlet with closed valves at the boundaries with adjacent DMAs.

2.2 Pipe-life expectancy

I start by collecting the necessary data from the utility systems for the parameters that are expected to have an impact on the life expectancy of pipes. Significance analysis will then be performed to check if these independent variables are associated with the response. Then the best subset analysis will try to find the best combination of the independent variables to be used in a regression analysis and thus eliminate those that do not provide any additional information. After that, several regression models will be built using these independent variables. The models will be evaluated and compared using statistical measures to identify the best model. The best model will then be used to investigate the impact of pressure on pipe-life expectancy. See Figure 3.

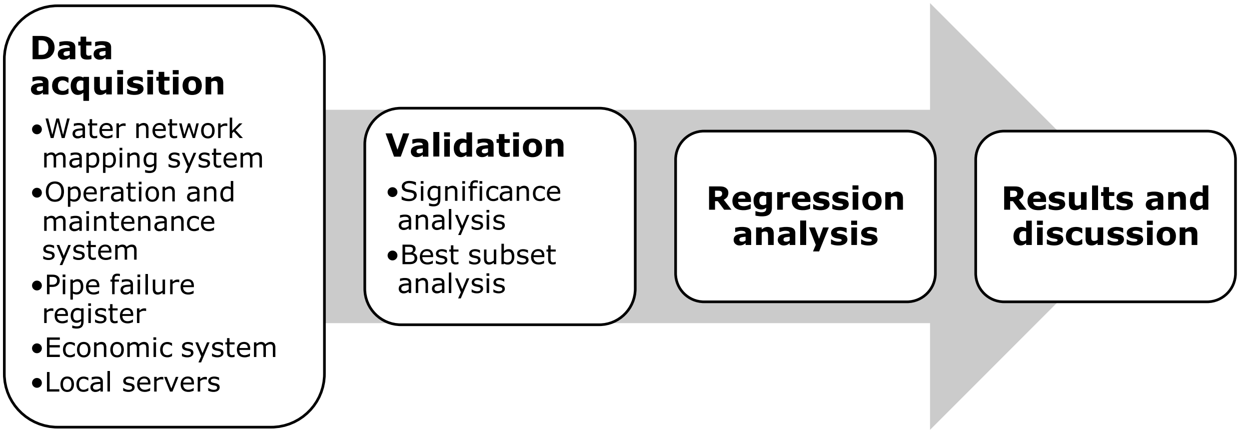


Figure 3 Main steps of the applied methodology for the pressure-pipe-life expectancy relationship investigation

2.2.1 Data acquisition

Data about the pipe's material, breaks, length, diameter, year of construction and piezometric head will be collected. Detailed information about the data acquisition process is provided in the case study.

2.2.2 Significance analysis

A significance analysis can help in determining whether there is a relationship between an independent variable or variables and a dependent variable. There are many methods that can be used to conduct the analysis. In this thesis, the test of fixed effect terms (Kenward-Roger approximation) is chosen.

2.2.3 Best subset analysis

The use of too many variables to create a regression model increases the probability of including variables that may be unimportant for the performance of the model and may offer a fake improvement leading to an overfitted model. Overfitting happens when the model extensively fits the current data but cannot validate future data with the same precision, if at all. Therefore, it is crucial to investigate how the different independent variables (predictors) affect the model. Several methods such as best subset regression, forward stepwise regression, and Lasso regression can help conduct such investigation. In comparison between methods, Hastie, Tibshirani & Tibshirani (2020) have concluded that neither best subset analysis nor Lasso dominate the other and that best subset analysis and forward stepwise regression have similar performance. Therefore, all these methods are appropriate to perform the investigation.

Best subset regression is chosen, and it will be used in this thesis to decide which independent variables to include in the regression analysis i.e. selecting the independent variables that increase R^2 and adjR^2 and decrease s and Mallows' Cp values.

2.2.4 Statistical measures

This section presents an overview of the statistical measures used to evaluate the results from the best subset analysis, the significance analysis, and the regression analysis. Statistical programs like Minitab and MATLAB are used to calculate these measures. However, their equations are included to show how they are calculated.

2.2.4.1 **R²**

R-squared or as sometimes called the "coefficient of determination" is a statistical measure commonly used as a measure of goodness of fit in regression models. In other words, R^2 indicates how well the model fits the data i.e. how close its predictions of the dependent factor (response) are to the observed data. R^2 can also explain how much the variability of the dependent factor is correlated with another related independent variable or variables. R^2 value always varies between 0.0 and 1.0, where 0.0 indicates that the model doesn't explain any of the variability in the dependent factor while 1.0 explains all the variability in dependent factor. Thus, higher R^2 value indicates a better data fit, but this is not always the case. A justified causal relationship between the response and the predictor is necessary to consider R^2 value as a measure of goodness of fit. In this thesis, it is expected that there is a causal relationship between the regression predictors and the response. Therefor R^2 is considered an important measure to evaluate the results. R^2 can be calculated using the following equation:

$$\boldsymbol{R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}} \qquad \textbf{(2)}$$

Where, $\boldsymbol{Y_i}$ = observed value, $\boldsymbol{\hat{Y}_i}$ = predicted value of $\boldsymbol{Y_i}$, $\boldsymbol{\bar{Y}}$ = mean value of the observations

Other uses of R^2 like measuring the explanatory power of the regression predictors and using R^2 as an indicator of the correctness of the regression model were criticized by Moksony (1999). Moksony considers the use of R^2 for these purposes irrelevant and misleading for several reasons; R^2 does not provide any substantive explanation but only statistical explanation, R^2 does not necessarily indicate a real causal relationship, and R^2 value is based on the observations that are used in the regression model training which means this value can change for different set of observations. However, this thesis does not employ R^2 for any of the uses criticized by Moksony.

2.2.4.2 **R² adjusted**

Adjusted R^2 as R^2 also indicates how well the model fits the data, but unlike R^2 the adjusted R^2 take into consideration the number of independent variables (k) as shown in Equation 3. R^2 value increases by adding more independent variables while adjusted R^2 decreases. More specifically adjusted R^2 decreases when the add parameters are useless and increases when these are useful. This is what makes adjusted R^2 preferred by researchers as a measure of goodness of fit instead of R^2 or along with R^2 (Lewis-Beck, Bryman & Liao, 2004). In this thesis, adjusted R^2 will be used to ensure that the increase in R^2 value is due to the addition of useful independent variables.

$$\mathrm{adj}\boldsymbol{R^2} = \boldsymbol{1} - \left[\frac{\boldsymbol{(1 - R^2)(n - 1)}}{\boldsymbol{n - k - 1}}\right] \qquad \textbf{(3)}$$

Where, n = sample size, k = number of independent variables

2.2.4.3 **R² predicted**

The predicted R^2 is mainly used to avoid overfitting when training a regression model. It can also be used to investigate the prediction performance of a regression model for new observations. It is calculated by removing a single observation from the data set, using the rest of the observations for training a regression model, and using the model equation to generate a prediction for the observation that is kept aside and then calculate the residual error. This process should be repeated for all the observations. The same equation for R^2 can then be used to calculate the predicted R^2. The only difference is that the $\hat{Y}_i$ in Equation (2) is the predicted value of the removed observation in each model run.

2.2.4.4 **Mallows' Cp**

Mallow' Cp is statistical measure helps in choosing between multiple regression models (Upton & Cook, 2008). Mallow' Cp has a similar use as adjR^2 where Mallow' Cp helps to decide which independent variables that contribute to improve the regression model. It compares precision and bias of the full model that includes all the independent variables of interest to subset models that include a different subsets of these variables. This comparison gives Mallow' Cp the power to detect both overfitting (too many variables) and underfitting (too few variables) and thus it helps in balancing between overfitting and underfitting the models (Lesik, 2009). The Mallows' Cp can be calculated as follows:

$$\boldsymbol{Cp} = (\boldsymbol{n} - \boldsymbol{p})\frac{(\mathbf{1} - adj\boldsymbol{R^2}_{\boldsymbol{p}})}{(\mathbf{1} - adj\boldsymbol{R^2}_{\boldsymbol{k}})} + \mathbf{2}\boldsymbol{p} - \boldsymbol{n} \tag{4}$$

Where, n = sample size, $adj\boldsymbol{R^2}_{\boldsymbol{p}}$ = adjusted R-squared for the subset model contains p independent variables of k, $adj\boldsymbol{R^2}_{\boldsymbol{k}}$ = adjusted R-squared when the model contains all variables of interest (k) (Managa ,2018)

A small Mallows' Cp value that is less than the number of independent variables in the model plus one (P+1) indicates that the model is relatively unbiased and thus a large Mallows' Cp value indicates that the model has high bias.

2.2.4.5 **S-value**

S-value is known as the standard error of the regression or the standard error of estimate. The S-value like R^2 is also used as a measure of goodness-of-fit, but unlike R^2, it measures the precision of predictions generated by a regression model where its value represents the average distance between the observed variables and the regression line (Siegel, 2016). The S-value can be calculated as follows:

$$\boldsymbol{S} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{df}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}(\hat{\boldsymbol{Y}}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}})^{\mathbf{2}}} \tag{5}$$

Where, df = degree of freedom (sample size minus number of independent variables), $\boldsymbol{Y}_{\boldsymbol{i}}$ = observed value, $\hat{\boldsymbol{Y}}_{\boldsymbol{i}}$ = predicted value of $\boldsymbol{Y}_{\boldsymbol{i}}$

2.2.4.6 ***P*-value**

The probability value (*P*-value) is a number used to decide whether to accept or reject the null hypothesis. The null hypothesis means that there is no significance and statistical relationship between two observed variables being studied (Perdices, 2017). The *p*-value ranges between 0 and 1, where a *p*-value < 0.05 is considered a strong evidence against the null hypothesis and vice verse (Lovell, 2020). In this thesis, *p*-value will be calculated using a statistical test called the test of fixed effect terms (Kenward-Roger approximation) in a statistical program called Minitab to check whether the is an association between the independet varaibles (predictors) and the dependent variable (response) which will be used in the regression analysis. The *p*-value can also be manually calculated by calculating the z-value then using the resulted z-value to find the *p*-value from the z-score table. The z-value can be calculated as follows:

$$\boldsymbol{Z = \frac{\hat{p} - p0}{\sqrt{\frac{p0(1 - p0)}{n}}}} \qquad \textbf{(6)}$$

Where, n = sample size, $\boldsymbol{p0}$ = assumed population proportion in the null hypothesis, $\boldsymbol{\hat{p}}$ = sample proportion

2.2.4.7 **MAE**

Mean absolute error (MAE) is a measure that can be used to compare actual observations with their predictions where MAE represents the average absolute distance (error) between the observed value and the predicted one (Sammut & Webb, 2011). So, the lower MAE value indicates a better model that generates predictions close to the actual observations. In this thesis, MAE will be used to assess the prediction performance of the regression models. The MAE can be calculated as follows:

$$\boldsymbol{MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{Y}_i - Y_i\right|} \qquad \textbf{(7)}$$

Where, n = sample size, $\boldsymbol{Y_i}$ = observed value, $\boldsymbol{\hat{Y}_i}$ = predicted value of $\boldsymbol{Y_i}$

2.2.4.8 **RMSE**

The root mean square error (RMSE) as MAE is also used a measure of prediction error. Unlike MAE the differences between the observed values and predicted values are being squared in RMSE. Therefore, the RMSE is more sensitive than MAE to large differences (Willmott & Matsuura, 2005). In this thesis, RMSE will help to choose a regression model with lowest presence of large differences. The RMSE can be calculated as follows:

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}} \qquad \textbf{(8)}$$

Where, n = sample size, $\boldsymbol{Y_i}$ = observed value, $\boldsymbol{\hat{Y}_i}$ = predicted value of $\boldsymbol{Y_i}$

3 Case Study

This case study will focus on water leakage in one DMA in the water distribution system in the south of Bergen to investigate the impact of applying pressure management on the annual volume of the avoidable water leakage. Then the study will try to investigate the relationship between pressure and the life expectancy of the mains and services in the city of Bergen. For this purpose, data on pipes that had undergone breaks that were logged into Bergen's water utility's systems was collected.

3.1 Avoidable water leakage

The chosen DMA (Figure 4) is also a PMA with a PRV on the entry pipe and closed zone values with neighbour DMA's. It consists of approximately 8 km public pipes and 1.6 km private pipes include two pump stations. The main reason for choosing this particular PMA is that it has high leakage and high difference in altitude, representing a great part of the water network in Bergen.

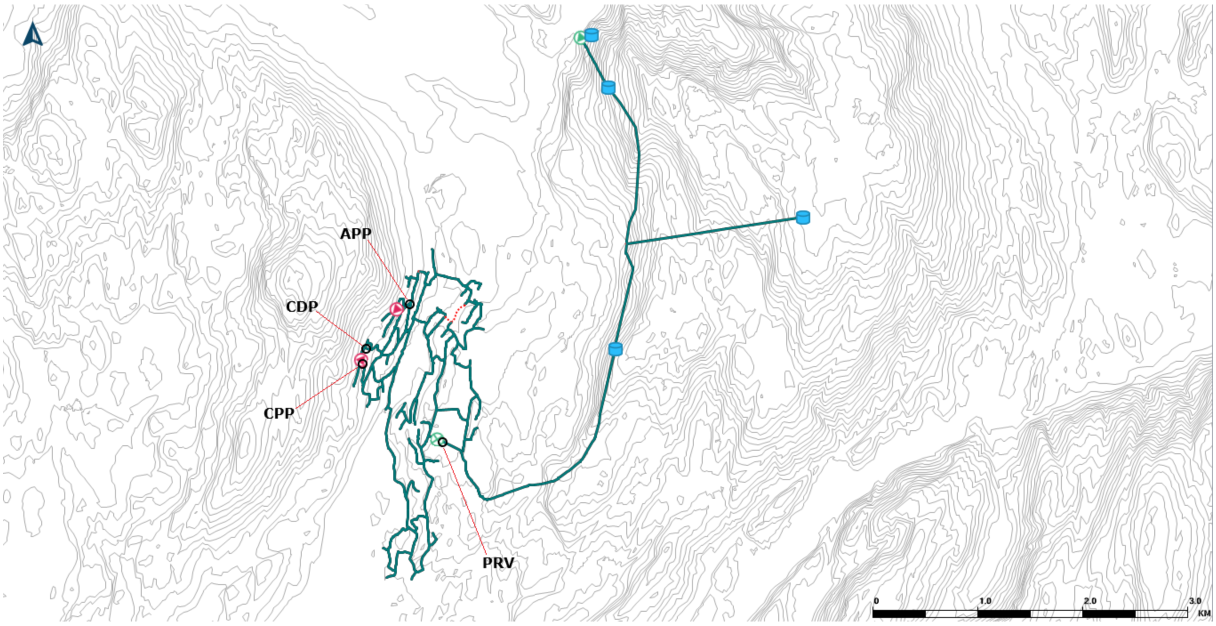


Figure 4 Representation of the chosen zone showing APP, CDP, CPP, PRV

3.1.1 Critical Pressure point (CPP)

The critical pressure point is a pump station located at a high altitude of 84.6 m AMSL, 35.8 meters higher than the DMA main entry. The pump station requires 18 m entry pressure for optimal operation and a minimum of 14 m to avoid a total shutdown.

3.1.2 Critical demand point (CDP)

The highest fire hydrant in the network is 73 m AMSL and the pipes that supply it have great roughness which makes this hydrant the most vulnerable point when it comes to water demand.

3.1.3 Average pressure point (APP)

Some pressure regulation methods such as FM use a point located at an average height and distance from the PMA PRV beside CPP when creating and calibrating the PRV control curve. At the same time, APP is also used to confirm the pressure readings from the CPP, prevent technical problems from interrupting pressure management, and to avoid mis-regulation in case of closed internal valves or open zone valves.

3.1.4 Firefighting demand

Bergen's water norm categories the firefighting demand into residential buildings and other types of buildings demand. The minimum requirement for residential buildings is set at 20 l/s while 50 l/s for the other buildings with a pressure of not less than 1.0 bar, where the pressure is lowest due to the firefighting demand. The DMA in this case study consists largely of residential buildings, so the firefighting demand is considered accordingly 20 l/s.

3.1.5 Flow and pressure data

A week (Monday - Sunday) of data from November 2021 is extracted for modelling, simulation, and analysis (Figure 5). This week has a repetitive flow pattern and stable pressure with an average of 55.8 m.

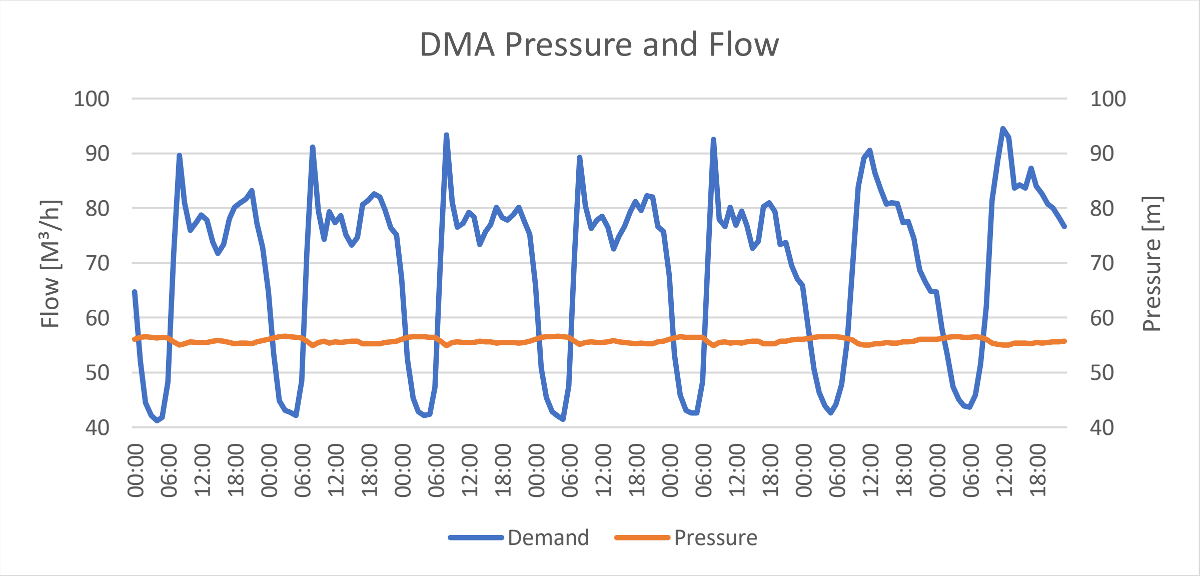


Figure 5 A week (Monday - Sunday) of DMA pressure and flow data from SCADA

3.2 Pipe-life expectancy

3.2.1 Data acquisition and parameters investigation

3.2.1.1 **Pipe material**

Bergen municipality's water distribution network consists of approximately 1020 km of public water mains and approximately 1500 km of private service pipes in operation. For the most part, the public water mains are from ductile iron (47%) and grey cast iron (23%) materials. While on the other hand, 71% of the private network consists of polyethylene materials.

As shown in Figure 6, the choice of pipe material types varied significantly over the years. By the mid-60s, grey cast iron made up over 90% of the public network. From 1965 to 1980 the utility transitioned away from grey iron to ductile iron and asbestos cement pipes. By 1979 the utility had stopped using asbestos pipes due to the material's health risk. The use of polyethylene pipes has seen an increase since the 1980s and constitutes today 10% of the entire public network.

It is noteworthy that 2% of the public network and 21 % of the private network lack information about pipe material. These pipes and the pipes with material that has low population are not included in the regression analysis.

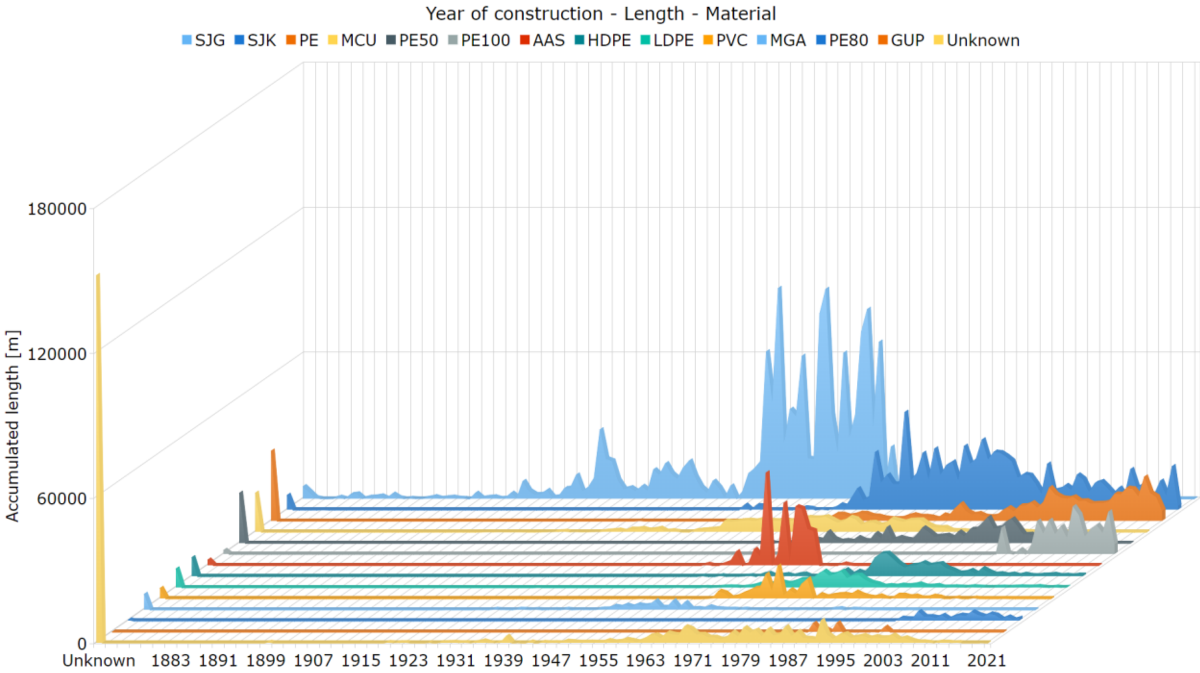


Figure 6 Area diagram of Pipe Length - material - year of construction

3.2.1.2 **Pipe burst**

The water utility in Bergen uses several data systems to register data over pipe bursts in its water network. Different amounts of information with different degrees of quality are stored in the operation and maintenance system, folders in one of the utility servers, the utility's economic system, and the water network's GIS map. It is also apparent that the data quality varies between the operators who submit it. In other words, there is a noticeable lack of following the utility burst registration routines.

Following are some of the reasons that may explain the differences in data quality between the different operators:

- Some of the operators are more concerned with getting the job done, i.e. repairing the bursts instead of updating the register with information about them.
- Others forget to digitalize the information they write on paper.
- Quite a few of the operators managed to report sufficient information about the bursts.
- Organizational challenges regarding the responsibility of registration.

- Lack of quality control and correction such as double registration, incorrect registration (incorrect pipe number, burst type, date of burst, location, etc.)
- Night -shift operators often leave the responsibility for burst registration to the day-shift operators. While the registration is supposed to be done within the shift time when the burst occurs or at least by those involved in the repair.

A summary of the registered bursts is obtained by merging the data from the utility systems, and a data quality control is performed. The summary shows 2,641 pipes, largely from grey cast iron (Figure 7), with up to 11 bursts (Table 2). Depending on the number of bursts and other factors such as the year of construction, importance, location, diameter, material, etc., corrective actions such as pipe removal, renewal or replacement can be decided when needed.

Table 2 Number of pipes grouped by number of burst

Number of burst	1	2	3	4	5	6	7	8	9	10	11
Number of pipes	2641	981	495	273	139	77	47	27	6	1	1

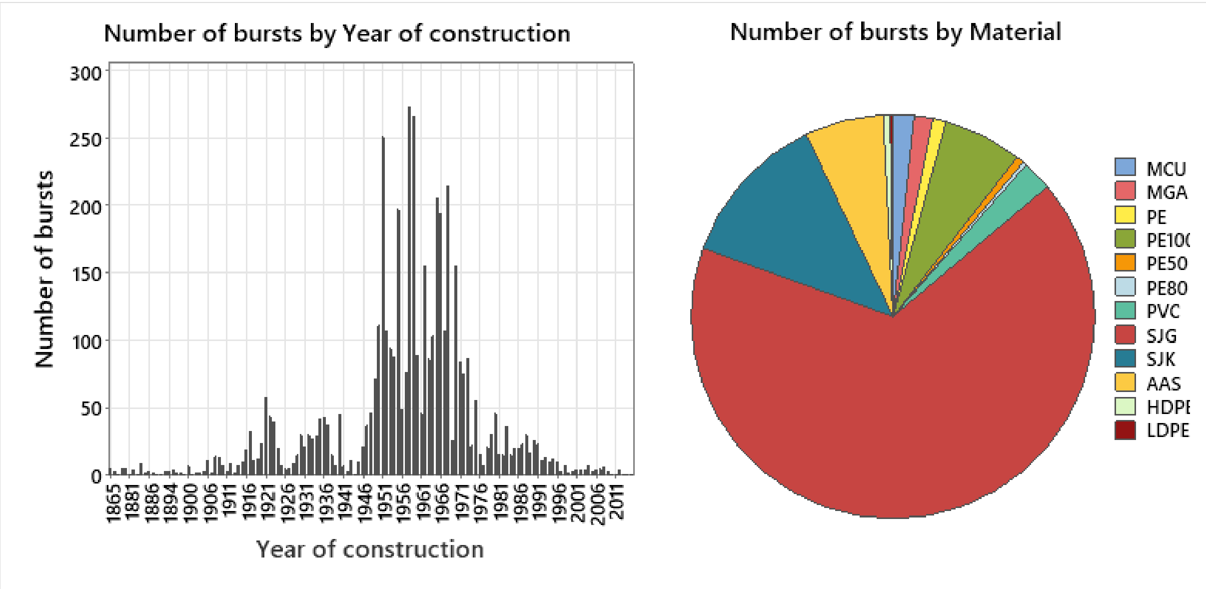


Figure 7 Number of bursts by year of construction and material

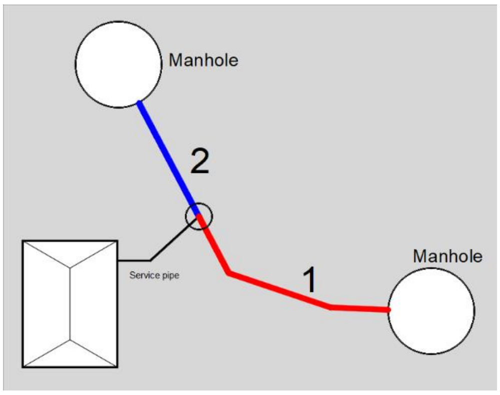
The oldest burst that can be found in the systems dates to 1974. At the same time, the oldest pipe dates to 1865 (Figure 7). This means there is no information about pipe bursts between 1865 and 1974, which presents a high uncertainty about the number of bursts, especially for pipes with a year of construction from this period.

3.2.1.3 **Pipe length**

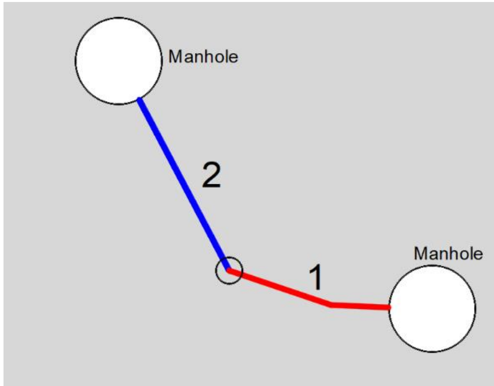
The recorded data on pipe lengths in the utility's systems show that pipe length is not always measured in the same way. These data show that the length may be the distance between:

- a manhole and a service pipe connection point (Figure8),
- two manholes (Figure 8b),

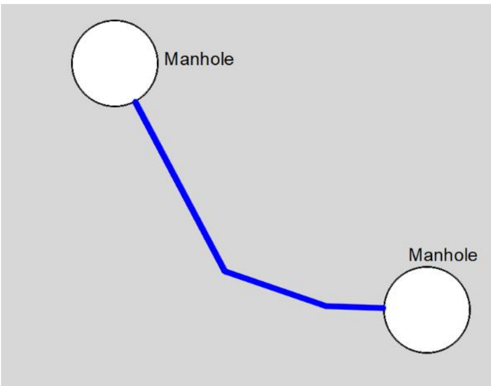
- a manhole and a virtual point or a point of replacement or renovation (Figure 8c) (Both pipes 1 and 2 have the same characteristics as the year of construction, diameter, and material. These virtual points are likely the result of an error that occurred while drawing the pipes.)
- or two virtual points or two points for replacement or renovation (Figure 8d, line 2). In this case, the data on bursts will be associated with a length shorter (the pipe of the new ones which gets the original pipe ID) than the actual length (original pipe, sum of 1, 2 and 3 lengths) and may be duplicated to all new pipes (1, 2 & 3)



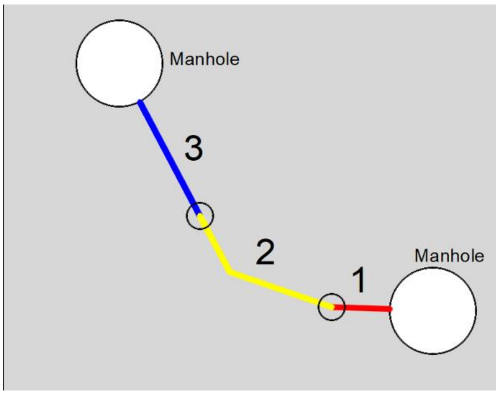
a) a manhole and a service pipe connection point



c) a manhole and a virtual point or a point of replacement or renovation



b) two manholes



d) two virtual points or two points for replacement or renovation

Figure 8 Reference points for measuring pipe length

To avoid the uncertainty of pipe lengths, pipe lengths in this study have been recalculated to either the length between two manholes or the length between a manhole and the point of replacement or renewal (in case the burst happened after replacement or renewal).

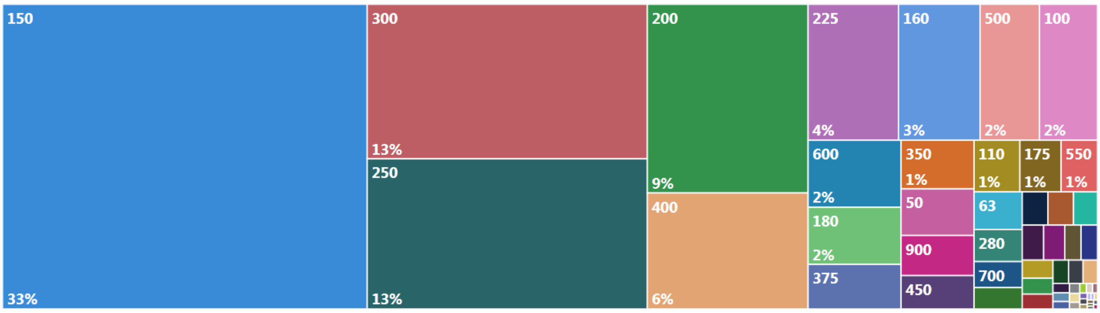
It is also worth mentioning that all pipe lengths are measured in two dimensions, i.e. calculated using the x and y coordinates of the pipe endpoints. Therefore, these 2D lengths do not consider the pipe slope, which in many cases results in pipe lengths much shorter than the actual lengths (3D length). Since Bergen is located on a terrain with a high difference in altitude, using 2D pipe lengths can cause significant errors in hydraulic, statistical, and other analyses.

3.2.1.4 **Pipe diameter**

The pipe diameter property used by the utility refers to either the inner, nominal, or outer diameter. Even though there are data fields for each type of diameter in the utility's systems, these are rarely used. Using the diameter parameter in a regression analysis without knowing the diameter type can cause a heterogeneity problem and impact the regression results.

Another problem that may affect the regression results is the different population sizes for each pipe diameter. From Figure 9, it is obvious which pipe diameters are used most in public and private networks. It is also clear that a few pipe dimensions represent a large majority of the total available dimensions. These overrepresented dimensions might overshadow whatever role the less represented dimensions play, which in turn will be reflected in the regression analysis.

It is also important to note that pipe diameter varies over time depending on the pipe material, hydraulic conditions, and water quality. Knowing the actual diameter of the pipe may significantly affect the results, but this is not an easy thing in water networks and requires costly and time-consuming fieldwork.



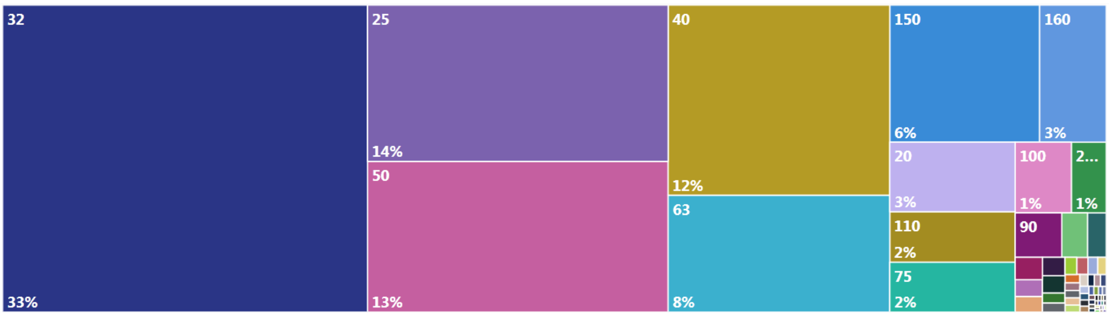


Figure 9 Percentage of installed pipes by diameter (mm) for water mains (a) and services (b)

3.2.1.5 **Pipe year of construction**

The year of construction is usually obtained from plumping reports stored in the municipality archive, giving this parameter a high degree of confidence. Still, the utility system lacks information about the year of construction for 1% of the public network and

16% of the private network. Figure 10 shows the year of construction for pipes in operation, where 90% of these pipes have a year of construction greater than 1955.

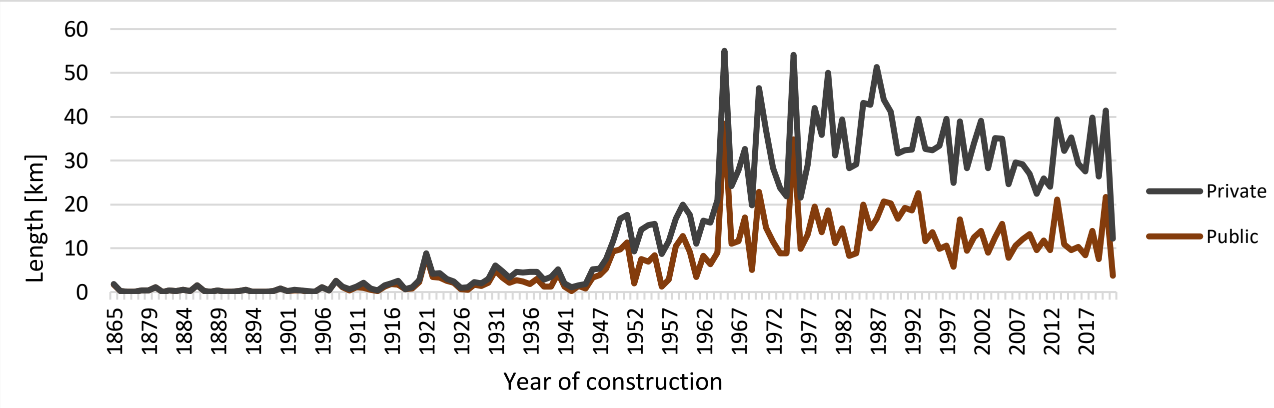


Figure 10 Installed pipes by year of construction

3.2.1.6 **Pipe piezometric head**

Piezometric pipe head is a parameter derived from maximum static zone pressure and pipe's minimum elevation. Sufficient information about the maximum static zone pressure is available in the utility systems, while the pipe's minimum elevation had to be obtained with the help of a data program called 'Saga GIS' by performing a B-spline interpolation between the pipes and a high-resolution DEM.

Since pipes in Bergen are positioned at many different elevations, simplification in deriving the pipe's piezometric head was necessary. Therefore, the depth for all pipes was assumed to be the same (2 meters). It is worth mentioning that neglecting pipe depth is not based on the assumption that depth is unimportant. On the contrary, pipe depth can be used as a parameter independent of piezometric pressure since depth can explain the impact of weather conditions on pipe failure.

3.2.1.7 **Pipe age at burst**

The pipe age at burst represents the dependent variable in the regression analysis. It is derived from the subtraction of the year of construction from the year of burst. The analysis will go through the individual observations of pipe age at burst to establish how they are related to pressure and the other parameters mentioned above to train a regression model that can generate predictions of pipe-life expectancy.

3.2.1.8 **Other important parameters**

The parameters mentioned above are not the only parameters that can describe the pipe's physical and hydraulic condition. Many other parameters may be no less important in describing these conditions, such as product standard, ring stiffness, pressure class, safety factor, reinforcement, internal protection, external protection, connection method, pipe depth, water temperature, water PH-value, and water quality etc. Unfortunately, there is no sufficient information about these parameters in the utility databases that can be used in a regression analysis.

4 Results and discussion

4.1 Avoidable water leakage

4.1.1 Minimum nightflow (MNF)

The minimum night flow components are calculated using IWA equations and shown in Table 3. It’s worth mentioning that:

- The flow data from the utility Scada system doesn’t show any exceptional customer night use. So this component is assumed to be zero.
- UBL is a pressure-dependent leakage, but it will be modelled as a constant consumption in this study. That's because UBL constitutes only a small part of MNF and is not the main interest of the study. In addition, costumer demands will be modelled as node demands in the hydraulic model, which will make it complicated to make only UBL pressure-dependent without affecting the other demands.
- Customer night use will be modelled as a pressure-independent mode of consumption. Its equations require information about the number of residents and addresses. Hopefully, the municipality has a population register database with high accuracy containing this information.
- Quite few costumers have flowmeters on their private pipes. Therefore, it is not possible to estimate the costumer night leakage. So its components are assumed to be zero.

Table 3 MNF components calculation

<table>
<tr><td rowspan="8">Night Leakage (NL)</td><td rowspan="7">Unavoidable Background Leakage (UBL)</td><td>Lm length of water mains [km]</td><td>19.98</td></tr>
<tr><td>Nc number of service connections</td><td>1633</td></tr>
<tr><td>lp average distance from curb stop to customer meter [m]</td><td>20</td></tr>
<tr><td>Lc total length of private pipe, Nc x lp [m]</td><td>32660</td></tr>
<tr><td>Average pressure [m]</td><td>55.8</td></tr>
<tr><td>N1-factor *</td><td>1</td></tr>
<tr><td>UBL Unavoidable Background Leakage [m³/h]</td><td>3.94</td></tr>
<tr><td>Avoidable Leakage (AL)</td><td colspan="2">MNF (Observed) - MNF (Theoretical)</td></tr>
<tr><td colspan="4"></td></tr>
</table>

Costumer Night Consumption (CNC)	**Costumer Night Leakage (CNL)**	**Inside Buildings [m³/h]**		**0**
		Outside Buildings [m³/h]		**0**
	Costumer Night Use (CNU)	**Assessed Night Use (ANU)**	NDP (Non-domestic properties)	**530**
			Non-domestic use [m³/h]	**4.24**
			CS average toilet cistern size [liter]	10
			DP (Number of domestic properties)	1549
			Number of residents (DP X R)	6694
			Domestic use [m³/h]	**4.02**
		ENU Exceptional Night Use [m³/h]		**0**
Avoidable Leakage (AL)	MNF (Theoretical) [m³/h]			12.19
	MNF (Observed) [m³/h]			41.20
	Estimated Avoidable Leakage [m³/h]			**29.01**

* The FAVAD N1 factor in the UBL equation is assumed to be 1; Lambert (2000) recommended a N1 of 1 for WDNs with mixed materials. The WDN in this case study consists of mixed material where the mains contain 89% metal and plastic 11%, and the services contain 66% metal and 34% plastic.

Using the flow data shown in Figure 5 and the results from Table 3. Customer consumption, UBL, and avoidable leakage can be illustrated as shown in Figure 11.

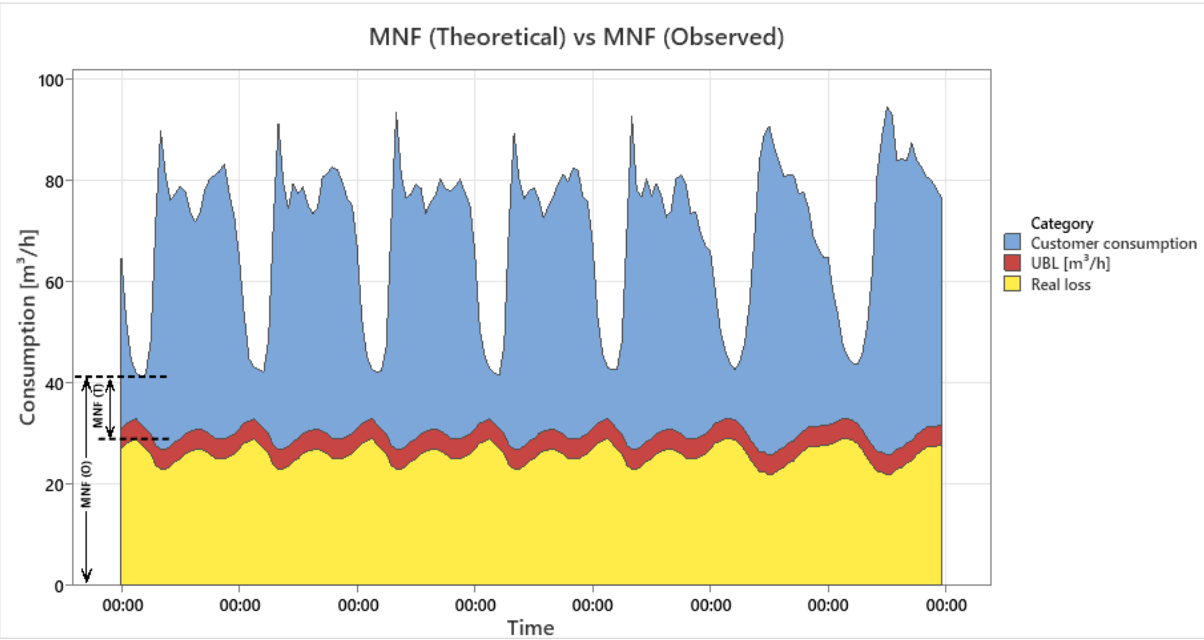


Figure 11 Demand components showing observed and theoretical MNF

4.1.2 Demand pattern and allocation

There are many hydraulic programs available in the market. These programs rely on the same hydraulic engine (EPANET), but they have different tools and functions and thus provide different possibilities. The main objective of this case study is to investigate the impact of pressure in an existing pressure-managed DMA in Bergen on its avoidable leakage. Therefore, a suitable program must provide comprehensive, high-resolution modelling that includes direct demand allocation (Figure 13) and accumulation from the addresses in DMA and offers different types of patterns ('date and time', 'year', 'day', 'hour'). Mike + 2021 meets these requirements and includes other advantageous features, so it was chosen.

The demand components differ in their pattern and must be modelled according to their nature. The customer consumption represented in blue in the Figure 11 varies throughout the day, is almost identical on working days and has higher flow peaks during weekends. Therefore, a corresponding hourly consumption pattern was created and implemented into the model (Figure 12). In contrast, UBL is considered constant and distributed equally to the service points. When it comes to avoidable leakage, over the years there has been different practices in modelling of this type of water loss. Some prefer to model it as constant demand for simplicity; others aim for higher accuracy and use emitters linked to the network points or demands attached to the pipes associated with their lengths. Emitters provide good accuracy, and their flow relies on the available pressure, making them preferable in this study. In Mike + emitters are defined with their exponent and flow coefficient. Mike + standard exponent of 0.5 is kept unchanged while the flow coefficient was found empirically to be 434e-5 l/s/m.

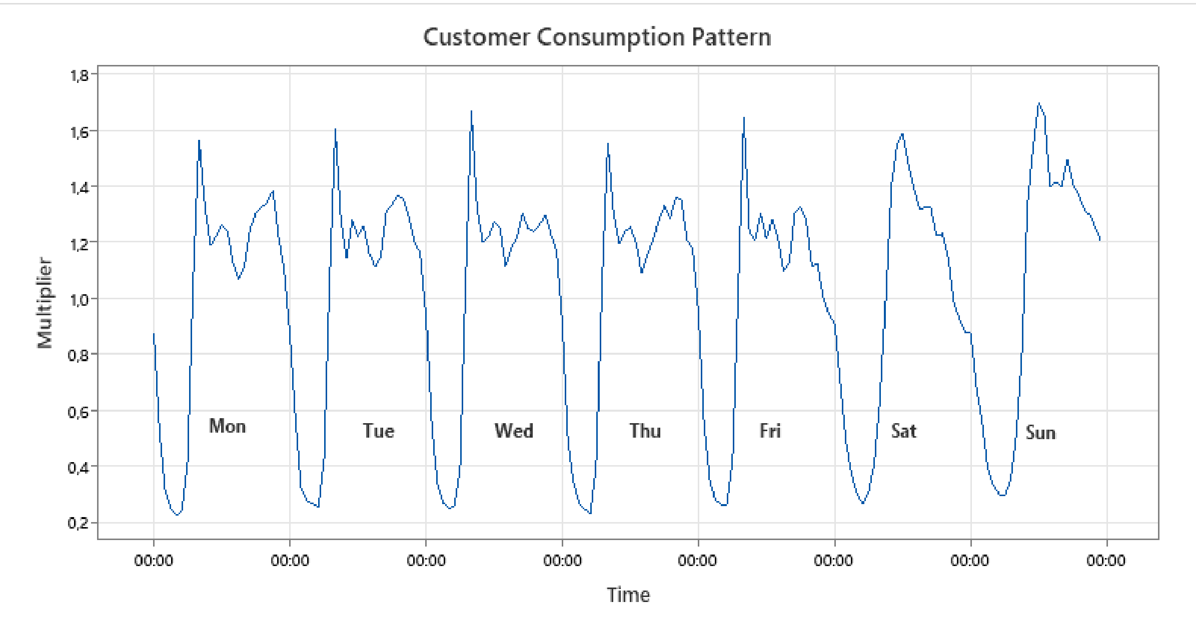


Figure 12 Customer consumption pattern

After setting a pattern for customer consumption, all that is needed is to know what the consumption is per address. A set of assumptions was tested to achieve average consumption in the model as the real average consumption from SCADA. The following assumptions have given the same average consumption: 130 l / pe / day for housing consumption and 70l / employee / day for public and private business consumption.

While the industrial consumption was taken from the flow meters (annually) to the industrial buildings and added directly to these.

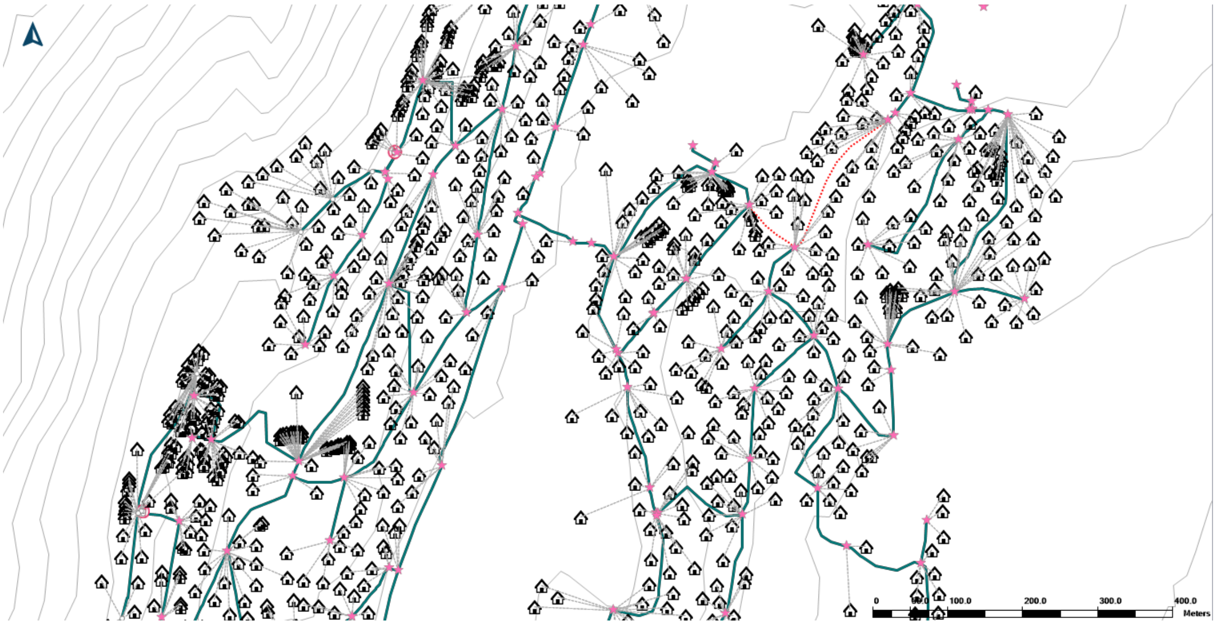


Figure 13 Representation of the chosen zone showing the consumer buildings and their connection to the WDN

Figure 14 shows the water demand (UBL, avoidable leakage and customer consumption) at 55.8 m (SCADA average pressure) compared to SCADA flow data. It appears that they overlap and that the model can be used further to investigate the relationship between avoidable leakage and pressure.

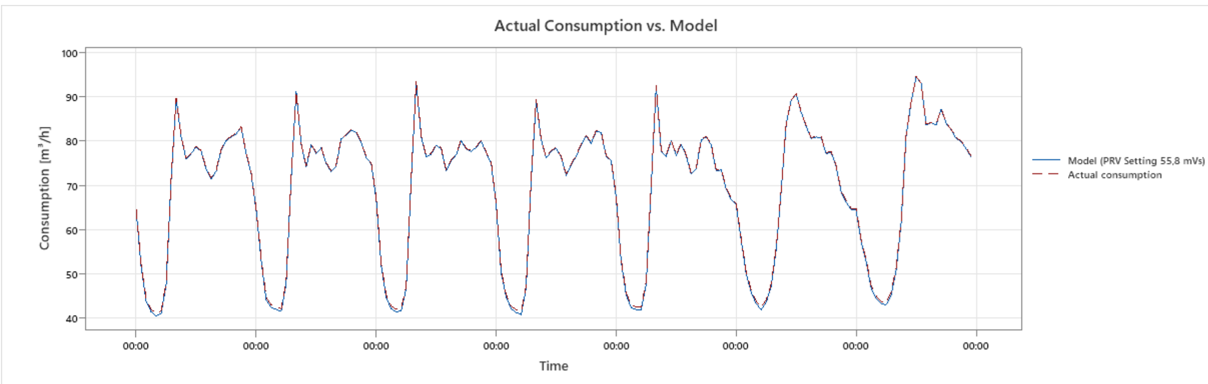


Figure 14 Actual DMA water input vs. model

4.1.3 Control curves

Time-based pressure regulation and flow-based pressure regulation, besides fixed outlet pressure, will be used in this case study. TM and FM demand a control curve that PRV can follow to meet the DMA hydraulic requirements. These curves are usually generated empirically or by using different machine learning techniques.

4.1.3.1 **Flow-modulation curve**

The flow control curve (FM) is empirically generated by using the 'curve fitting toolbox' in MATLAB and running the model using different control curves to reach a constant pressure of 18 m at the CPP (Equation 9). Another flow control curve (FM modified) was generated to simultaneously meet a constant pressure at CPP and a firefighting demand at the CDP (Equation 10). As shown in Figure 15, the PRV maintains a 76 m pressure after the flow requirement exceeds 45 l/s since 76 m is the maximum upstream pressure. The control curve 'FM modified' is included in this chapter only to show FM's flexibility and possibilities. This type of control curve is applicable in the event of significant differences in pipe roughness in the WDN.

$$\boldsymbol{PRV\ Setting}_{\boldsymbol{FM}} = \boldsymbol{h_x(t)} = \ \mathbf{480{,}2} * \boldsymbol{sin}(\mathbf{0.03383} * \boldsymbol{q_x(t)} - \mathbf{0.1469}) + \mathbf{455.4} * \boldsymbol{sin}(\mathbf{0.03876} * \boldsymbol{q_x(t)} + \mathbf{2.795}) + \mathbf{37} * \boldsymbol{sin}(\mathbf{0.05841} * \boldsymbol{q_x(t)} + \mathbf{5.154}) \quad \textbf{(9)}$$

$$\boldsymbol{PRV\ Setting}_{\boldsymbol{FM}_{\boldsymbol{Modified}}} = \boldsymbol{h_x(t)} = \ \mathbf{268.9} * \boldsymbol{sin}(\mathbf{0.03861} * \boldsymbol{q_x(t)} - \mathbf{0.1513}) + \mathbf{238.5} * \boldsymbol{sin}(\mathbf{0.05906} * \boldsymbol{q_x(t)} + \mathbf{2.535}) + \mathbf{49.03} * \boldsymbol{sin}(\mathbf{0.08751} * \boldsymbol{q_x(t)} + \mathbf{5.147}) \quad \textbf{(10)}$$

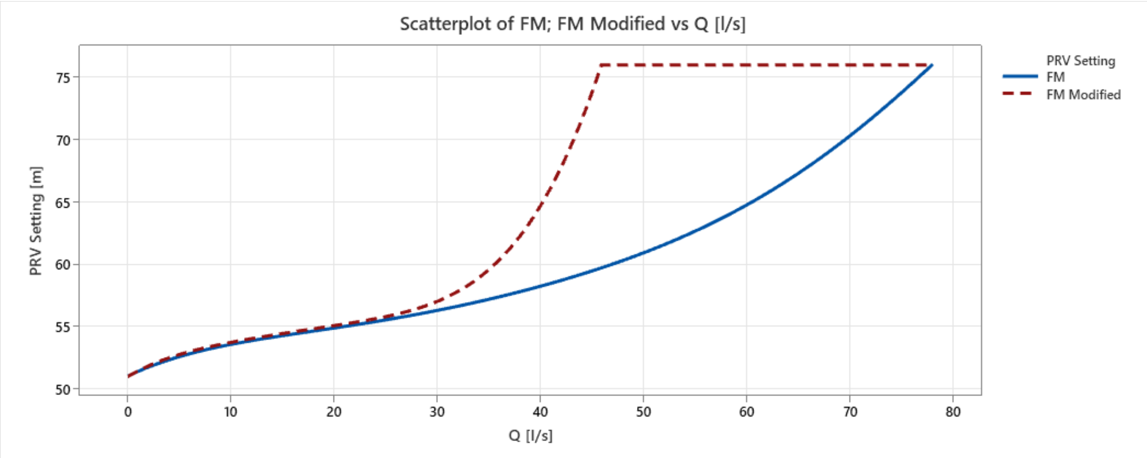


Figure 15 Flow modulation curves

4.1.3.2 **Time-modulation curve**

This control curve is computed also empirically by running the model with a different set of PRV outlet pressure per hour throughout the day to reach as low as possible a pressure at nighttime and daytime. The resulting curve is shown in Figure 16. This curve is not achievable with the available PRV valves in the market, where the pressure adjustment ratio (0.1 m) required by the curve is too low. Even though TM is not practical in this PMA, it will be investigated theoretically to estimate the leakage reduction potential.

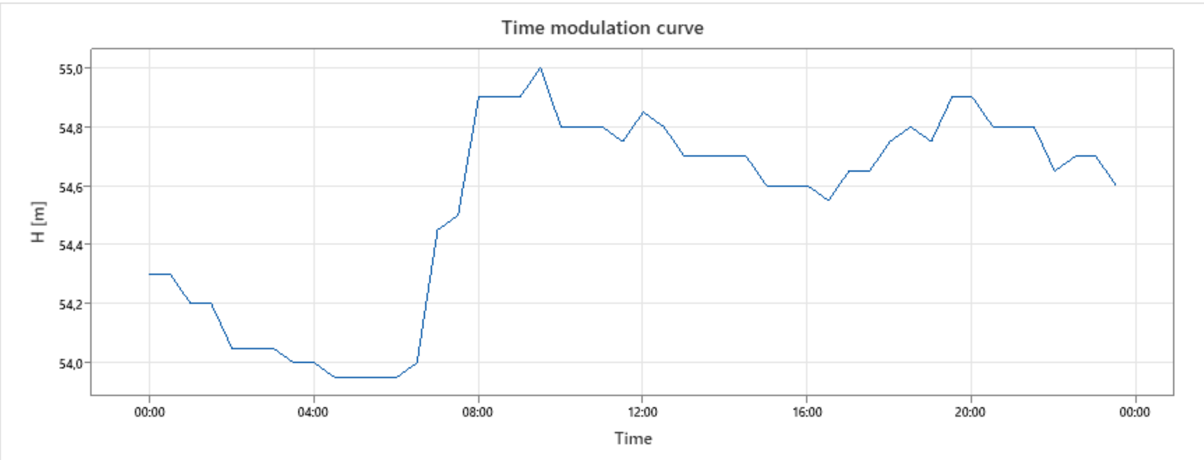


Figure 16 Time-modulation curve

4.1.4 The impact of pressure on water leakage

By running the model with various PRV settings, including FM and TM, CPP pressure and PRV flow curves are generated. From Figure 17, an increase in the PRV outlet pressure results in a corresponding increase to a large extent in CPP pressure due to the absence of high-pressure losses in the pipes between PRV and CPP, especially in normal consumption situations.

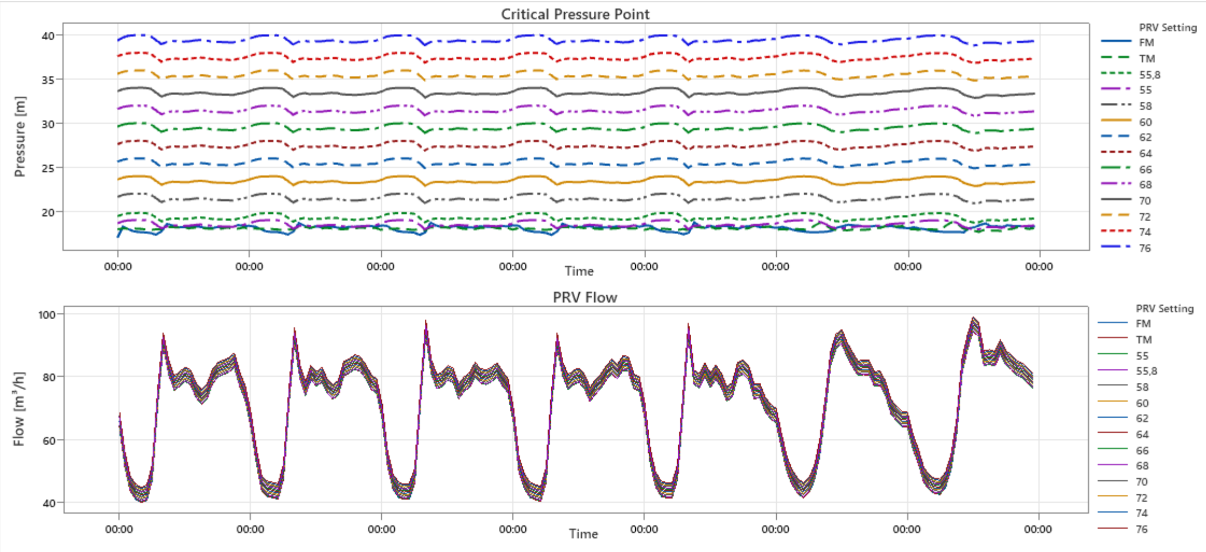


Figure 17 Pressure values at CPP and flow values at PRV for different PRV settings

An interesting but, at the same time, not surprising observation is that FM, TM, and PRV fixed outlet pressure offer different pressures between day and night. FM provides higher pressure during the day and lower at night, PRV fixed outlet pressure does the opposite, and TM provides almost a constant CPP pressure throughout the day (Figure 18).

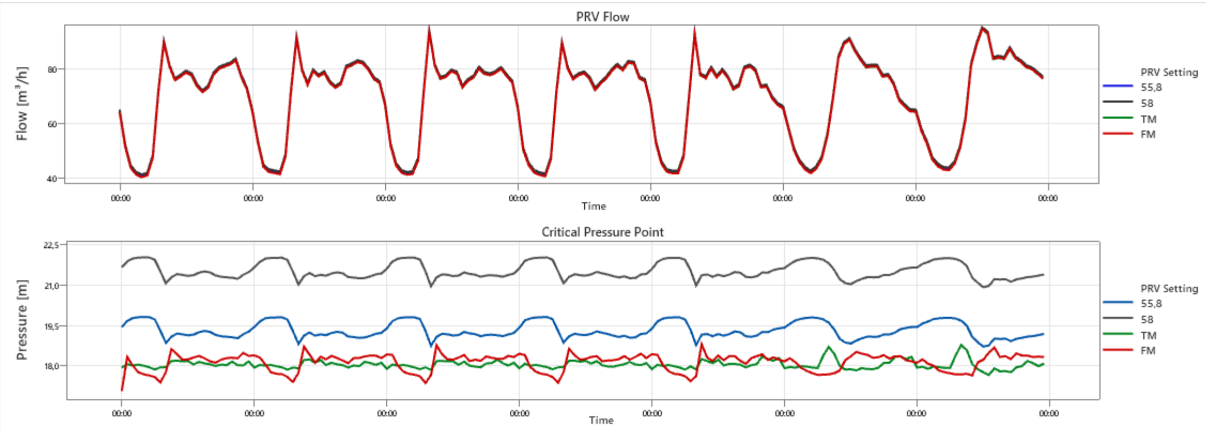


Figure 18 Pressure values at CPP and flow values at PRV for 55.8, 58, TM and FM PRV settings

The differences in PRV outlet pressure and between the day and night point to a difference in the avoidable leakage flow. Figure 19 shows hourly difference in consumption between the different valve settings. Both FM and TM give the most reduction in leakage compared to the rest of the PRV settings approximately 2,400 to 2,600 m³ annually (Figure 20). This corresponds to NOK 24,000 - 26,000 if the current average production and distribution price of NOK 10 / m³ is used as a basis. On the other hand, 76 m PRV setting (2 bar higher than current pressure) increases the leakage by approximately 38,000 m³ annually, which is 15% of the current annual leakage. In other words, the annual leakage increases at a rate of 8% for each one bar increase in pressure compared to the current annual leakage.

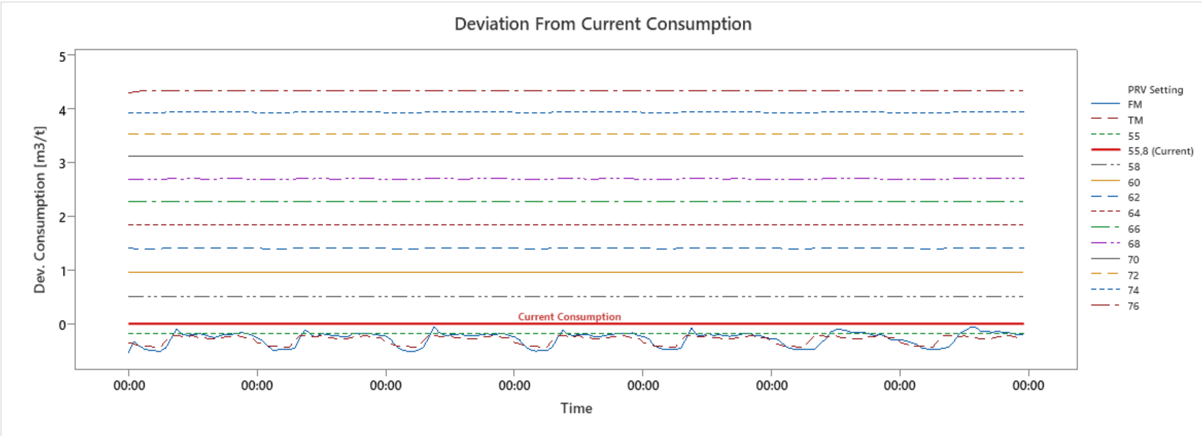


Figure 19 Deviation from current consumption at different PRV settings

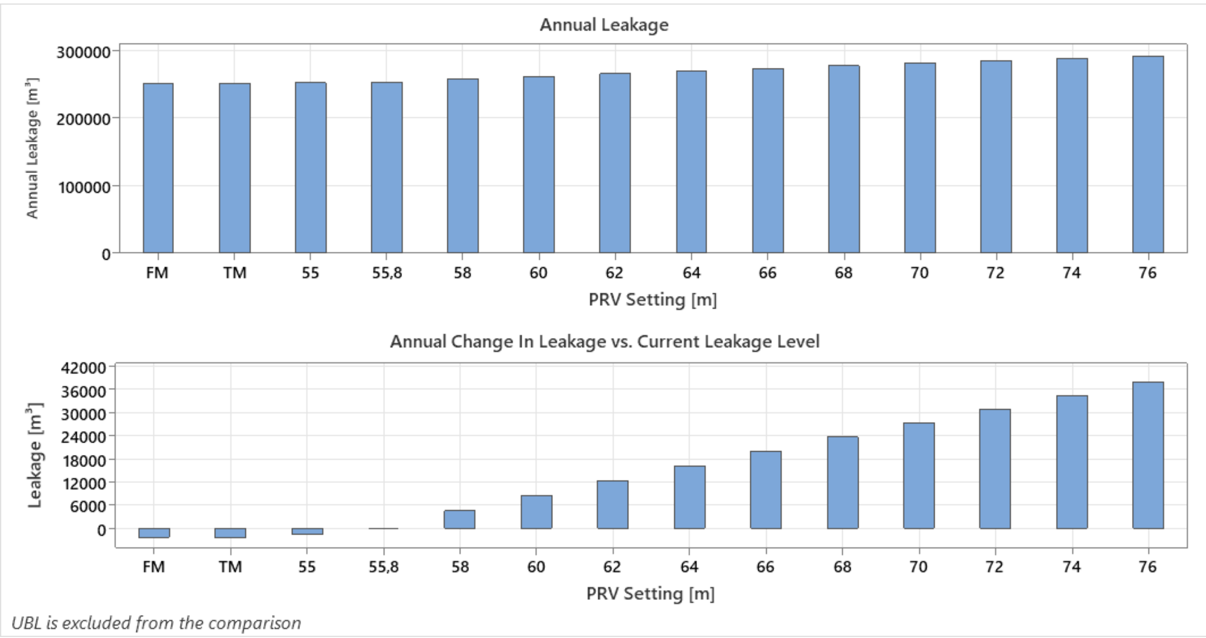


Figure 20 Annual leakage, Annual change in leakage level vs. current leakage level at different PRV settings

All the examined PRV settings managed to meet the CPP minimum pressure requirement in the normal consumption situation. But will these be able to do the same at higher demands at the network points connected to pipes with high roughness? The answer is no, not every setting. The only PRV settings that cover CPP's pressure requirements are always FM and 76 m. In case it is desired to use the other settings and at the same time ensure a minimum pressure in CPP, an extra PRV set point is usually used, which PRV switches to in case the demand is greater than what it can deliver with the current setting.

For any of the aforementioned pressure regulation methods to be considered suitable for the selected DMA, it must meet the requirements for firefighting in CDP as well as ensure a minimum pressure in CPP or that the latter must not be dramatically affected. To provide more flexibility when it comes to pressure regulation in emergencies, a pressure in CPP of no less than 14 m can be accepted. All methods have managed to ensure the desired flow (Figure 23) and pressure at CDP (Figure 21), but TM and 55 m failed to ensure a pressure of 14 m or above at CPP (Figure 22). Therefore, TM and 55 m are not considered suitable.

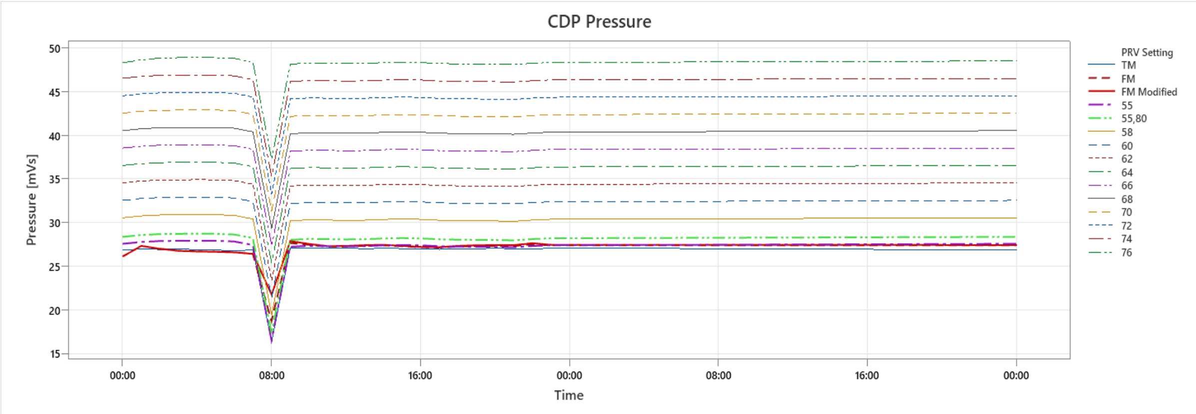


Figure 21 Pressure at CDP - firefighting demand at peak hour

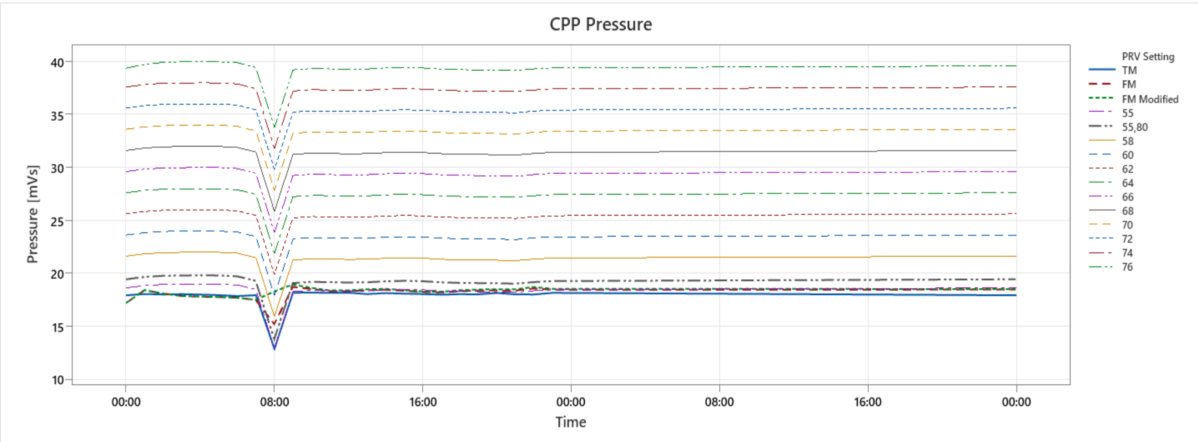


Figure 22 Pressure at CPP - firefighting demand at peak hour

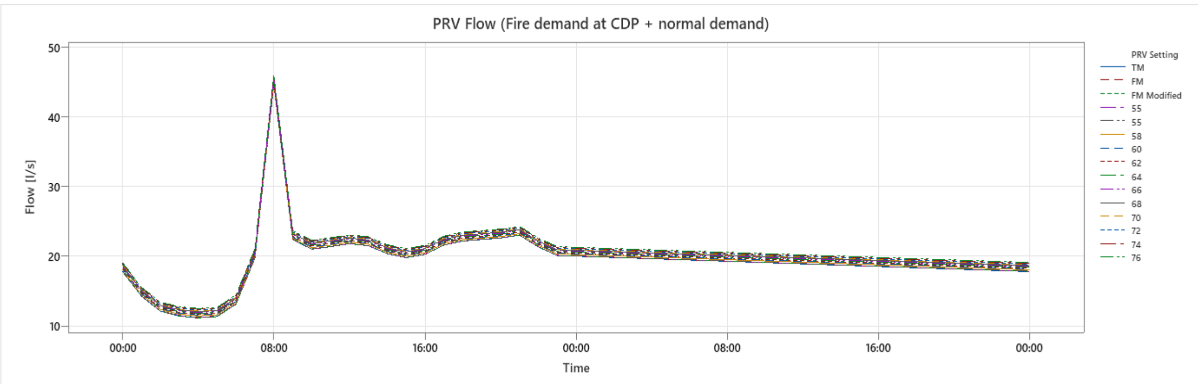


Figure 23 Flow at the DMA inlet - firefighting demand at peak hour

4.2 Pipe-life expectancy

4.2.1 Significance analysis

The results from the test of fixed effect terms (Kenward-Roger approximation) shown in Table 4 and Table 5 confirm that there are association between the independent variables and the dependent variable where all *P*-values are smaller than 0.05, i.e. a variation in the independent variable value causes a variation in the response value. So, all the independent variables are considered meaningful for the regression analysis, and the null hypothesis can be rejected.

Table 4 Variance Components

Source	Var	% of Total	SE Var	Z-Value	*P*-Value
Year	1471.861136	84.00%	187.139430	7.865051	< .001
Piezometric head	5.062326	0.29%	1.476365	3.428911	< .001
Burst number	124.067990	7.08%	59.772791	2.075660	.019
Diameter	50.279861	2.87%	18.945706	2.653892	.004
LENGTH	68.324977	3.90%	3.095608	22.071582	< .001
Error	32.549926	1.86%	1.063455	30.607700	< .001
Total	1752.146215				

Table 5 Tests of Fixed Effects

Term	DF Num	DF Den	F-Value	P-Value
Material	6.00	1759.76	4.13	< .001

4.2.2 Best subset analysis

The results from the best subset analysis are presented in Table 6. The values R^2, adjusted R^2, S, and Mallows' Cp estimate the performance of different combinations of the parameters. The higher R^2 and adjusted R^2 and the lower S and Mallows' Cp present a good model.

The results show that the year of construction is indispensable and combining it with one or several of the other parameters slightly increases the accuracy. Still, the best model remains a combination of all the parameters.

Table 6 Best subset analysis results

Model no.	R^2	R^2 adjusted	R^2 predicted	Mallows Cp	S	LENGTH	Year of construction	Diameter	Number of burst	Piezometric head
1	74.6	74.6	74.6	452.5	10.892		X			
2	1.2	1.2	1.1	15335.2	21.503	X				
3	74.7	74.7	74.7	444.5	10.883		X			X
4	2.1	2	2	15160.4	21.41				X	X
5	76.3	76.3	76.3	115.5	10.528		X		X	
6	75	75	75	384	10.818		X	X		
7	75	75	75	381.6	10.815		X	X		X
8	76.4	76.3	76.3	108.4	10.519		X		X	X
9	76.8	76.8	76.7	23.3	10.425	X	X		X	
10	76.5	76.5	76.5	73.9	10.481		X	X	X	
11	76.6	76.5	76.5	71.2	10.477		X	X	X	X
12	76.8	76.8	76.8	20.6	10.421	X	X		X	X
13	76.9	76.9	76.8	6.9	10.405	X	X	X	X	
14	76.9	76.9	76.8	6	10.403	X	X	X	X	X

4.2.3 The impact of pressure on pipe-life expectancy

Many programs can be used to build a regression model, such as Minitab and MATLAB. MATLAB has a regression learner toolbox that includes various regression models, which provides a great opportunity to find the most suitable model; therefore, it is used in the case study. The regression models used here are primarily optimizable models to include as many sub-models as possible in the analysis and to be able to choose the number of iterations and limit the training time. The number of iterations was set to 30 for all models, and the results are presented in Table 7. The higher R^2 and the lower RMSE and MAE represent a better model. However, this only applies to comparisons of models evaluated by the same validation methods. MATLAB's regression learner toolbox has three validation methods K-fold cross-validation, Holdout validation and Resubstitution validation. Holdout validation is usually used for large datasets. Resubstitution validation is used when overfitting is not a problem where all the data is used for training and testing the model. K-fold cross-validation is used when overfitting is a problem where a part of the data is left for testing the model. The dataset in this case study is not large, and overfitting is not desired, so K-fold cross-validation is most suitable here. However, an example of Resubstitution validation is included to show the difference between these methods. As a first look, the model evaluated by Resubstitution (OER model) appears to be the best but knowing that OER does not protect the model against overfitting, it will be excluded. From Table 7, both OGPR and OEC models show a good fit for the training and test data. OGPR scored best in model validation, and OEC scored best in model test.

Table 7 Regression Results

Regression response	Regression predictors[a]	Validation			Test			Model type[b]	Validation method[c]
		RMSE	R^2	MAE	RMSE	R^2	MAE		
Pipe age at burst	Y, N, M, P, D, L	8.95	0.83	6.88	5.80	0.93	4.43	OGPR	KFCV
		9.18	0.82	7.08	4.91	0.95	3.76	OEC	KFCV
		10.12	0.78	7.98	9.61	0.8	7.52	OSVM	KFCV
		10.17	0.78	8.18	9.28	0.82	7.40	OT	KFCV
		10.24	0.78	8.43	10.21	0.78	8.41	LR	KFCV
		2.77	0.98	2.00	2.77	0.98	2.00	OER	RV

[a]Y = Year of construction; P = Piezometric head; N = Number of burst; M = Pipe material; D = Pipe diameter; L = Pipe length.
[b]OER = Optimizable Ensemble, OEC = Optimizable Ensemble, OGPR = Optimizable Gaussian Process Regression, LR = Linear Regression, OT = Optimizable Tree, OSVM = Optimizable Support Vector Machines.
[c]RV = Resubstitution validation, KFCV = K-fold cross-validation (5-fold).

By plotting the predicted values from OGPR and OEC and the real data (Figure 24), it looks like OEC is more accurate in prediction. This is also confirmed statically in Table 8.

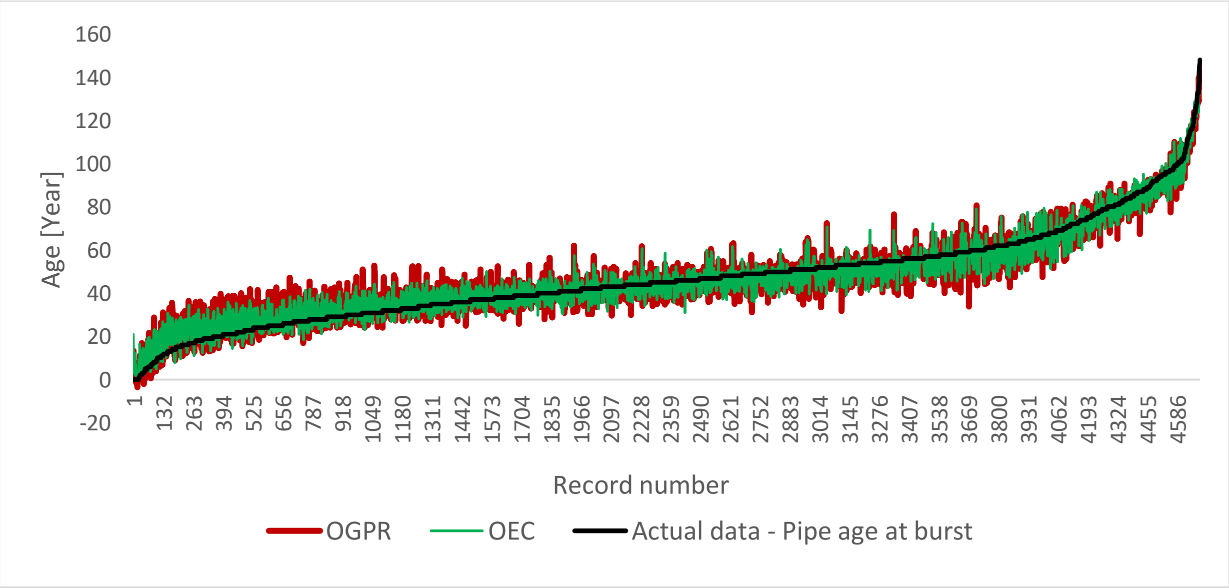


Figure 24 Comparison of predictions from OEC and OGPR vs. real observations of pipe age at burst

Table 8 Prediction deviations from real data

Deviation from real data	OGPR (% of total values)	OEC (% of total values)
0-5%	35.0	42.1
5-10%	25.6	23.9
10-15%	15.4	13.4
15-20%	7.9	7.8
20-25%	4.6	3.7
25-30%	2.6	2.0
Over 30%	8.9	7.1

Therefore, the OEC model is chosen to investigate the impact of maximum zonal pressure on pipe-life expectancy. The model is used to generate predictions of pipes age by reducing the piezometric head (by reducing the maximum zonal pressure) gradually by 2 m each run until a 10-m reduction is reached while keeping the other predictors unchanged. The results (Table 9) are grouped based on the piezometric head before reduction, and average change in pipe age is calculated for each group.

It appears the reduction in pressure resulted in increases in pipe age for all the groups, but especially in pipes with a high piezometric head. So, there is great correlation between the maximum zonal pressure and pipe-life expectancy. The latter increases with a decrease in maximum zonal pressure and vice versa.

Table 9 Average increase in pipe age [day] due to reduction in pressure [m] based on predictions from OEC model

Reduction in Piezometric head range [m]		Average increase in pipe age [day]	Accumulated average increase in pipe age [day]
Before reduction	After reduction		
30-39	20-29	518	518
40-49	30-39	641	1159
50-59	40-49	717	1877
60-69	50-59	853	2730
70-79	60-69	875	3604
80-89	70-79	1025	4629
90-99	80-89	1044	5673
100-109	90-99	1113	6787
110-119	100-109	1247	8034

Another way to investigate the effect of each predictor on pipe-life expectancy is by using the linear regression model. Linear regression coefficients are extracted, and its equation is generated (Table 10). The pressure coefficient has a negative value which means an increase in pressure will shorten pipe life while a decrease will do the opposite. However, the coefficient (0.0167) is insignificant, and pressure only slightly affects pipe life. In other words, for each bar change in pressure, the pipe burst date will be changed by approximately 62 days. This change is too small compared to the results from OEC. It can be explained by the fact that the relationship between the pressure and the pipe-life expectancy is not linear. The impact of pressure change on the pipe-life expectancy changes due to interactions between the pressure and the other independent predictors.

Table 10 Pipe material cofficient and Linear regression Equation

	MCU	MGA	PE	PVC	SJG	SJK
Pipe material coefficient	12.974	0.5614	-0.0054	5.5041	0.1458	3.7057

Response	Regression equation
Pipe age at burst (Life pipe expectancy) =	1978.59 + Pipe material coefficient – 0.98801 Year of construction – 0.01666 Piezometric head + 2.350 Burst number – 0.00443 Diameter – 0.01264 Length

Conclusion

This paper attempted to determine the impact of maximum zonal pressure on the life expectancy of pipes and the impact of applying different pressure management practices on avoidable leakage in water distribution systems.

Based on the regression analysis of the historical data of the pipes and their physical properties, it can be concluded that the maximum pressure has a significant influence on the pipe's life expectancy. The results indicate that even a small reduction of maximal pressure can potentially increase pipe life.

The magnitude of the reduction needed depends on the pressure piezometric head before pressure management is applied. A one bar reduction in pressure can increase the life of a high-piezometric-head pipe by more than three years, but only by one year for low-piezometric-head pipes.

This increase due to one bar reduction may sound minimal compared to the total pipe life. Still, given the cumulative effect on all pipes in the region where pressure management is applied, the benefit is significant. Further, the results from this thesis can be used in pipe replacement and renovation prioritization algorithms where it will lead to a better timing of applying these corrective measures which again the water utilities benefit from economically, reputationally, environmentally and operationally.

It can also be concluded that this thesis fills part of the research gap regarding the pressure-pipe life relationship. Future research should investigate other forms of pressure than maximum pressure such as average zonal pressure on pipe-life expectancy. At the same time, I recommend the collection of a bigger data set for pipe information like getting data from several municipalities or countries. A bigger data set will give a better opportunity of excluding corrupted and low-quality data without the risk of misrepresentation of the predictors in the analysis and at the same time it will lead to a bigger population for the pipes of relatively new materials or standards. Future research should also investigate the pressure impact on pipe-life expectancy in light of other factors such as pipe depth and water temperature which are not included in thesis due to lack of data.

The results from the simulations of the hydraulic model in the case study show that pressure impacts avoidable leakage. This confirms the findings in most of the prior research. These results estimate that annual water leakage increases by 8% for each one bar increase in pressure. It's worth mentioning that the results are related to the particular DMA which is used for the case study. Still, the methodology can be replicated in its entirety to investigate the pressure-avoidable water leakage for other DMAs.

References

AbdelMeguid, H. (2011). "Pressure, leakage and energy management inwater distribution systems." Ph.D. thesis, De Monfort Univ., Leicester,U.K

Abdelmeguid, H., & Ulanicki, B. (2010). *PRESSURE AND LEAKAGE MANAGEMENT IN WATER DISTRIBUTION SYSTEMS VIA FLOW MODULATION PRVS*.

Berardi, L., Giustolisi, O., Kapelan, Z., & Savic, D. A. (2008). Development of pipe deterioration models for water distribution systems using EPR. *Journal of Hydroinformatics*, *10*(2), 113–126. https://doi.org/10.2166/hydro.2008.012

Berardi, L., Laucelli, D., Ugarelli, R., & Giustolisi, O. (2015). Leakage management: Planning remote real time controlled pressure reduction in Oppegård municipality. *Procedia Engineering*, *119*(1), 72–81. https://doi.org/10.1016/j.proeng.2015.08.855

Bergen Vann. (2015-2020). *Årsmelding*. Retrieved from https://www.bergenvann.com/ansatte/arsmeldinger_/

Bergen Vann. (2021). *Leakage summary*. Internal report: unpublished.

Fantozzi, M., & Lambert, A (2010). Legitimate night use component of minimum night flows initiative. IWA Water Loss Conference 2010, São Paulo, Brazil

Farrow, J., Jesson, D., Mulheron, M., Nensi, T., & Smith, P. (2017). *Achieving Zero Leakage By 2050: The Basic Mechanisms Of Bursts And Leakage*. London: UK Water Industry Research Limited.

Folkman, S. (2018). Water Main Break Rates In the USA and Canada: A Comprehensive Study. *Mechanical and Aerospace Engineering Faculty Publications.* Paper 174. https://digitalcommons.usu.edu/mae_facpub/174

Fontana, N., Giugni, M., Glielmo, L., Marini, G., & Zollo, R. (2018). Real-Time Control of Pressure for Leakage Reduction in Water Distribution Network: Field Experiments. *Journal of Water Resources Planning and Management*, *144*(3), 04017096. https://doi.org/10.1061/(asce)wr.1943-5452.0000887

Ghorbanian, V., Guo, Y., & Karney, B. (2016). Field Data–Based Methodology for Estimating the Expected Pipe Break Rates of Water Distribution Systems. *Journal of Water Resources Planning and Management*, *142*(10), 04016040. https://doi.org/10.1061/(asce)wr.1943-5452.0000686

Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons. *Statistical Science*, *35*(4). doi: 10.1214/19-sts733

Hoțupan, A., Mare, R., & Hădărean, A. (2019). Water Loss Reduction in Water Distribution Networks. Case Study. *Journal of Applied Engineering Sciences*, *9*(1), 73–80. https://doi.org/10.2478/jaes-2019-0009

Jara-Arriagada, C., & Stoianov, I. (2021). Pipe breaks and estimating the impact of pressure control in water supply networks. *Reliability Engineering and System Safety*, *210*. https://doi.org/10.1016/j.ress.2021.107525

Jolly, M. D., Lothes, A. D., Sebastian Bryson, L., & Ormsbee, L. (2014). Research Database of Water Distribution System Models. *Journal of Water Resources Planning and Management*, *140*(4), 410–416. https://doi.org/10.1061/(asce)wr.1943-5452.0000352

Lambert A (2000): What do we know about Pressure: Leakage Relationships in Distribution Systems? IWA Conference on System Approach to Leakage Control and Water Distribution Systems Management, Brno ISBN 80-7204-197-5

Lambert, A. (2009). Ten years' experience in using the UARL formula to calculate infrastructure leakage index. In *Proceedings of IWA Waterloss 2009 Conference*. Cape Town, South Africa.

Lambert, A., & Hirner, W. (2000). *Losses from Water Supply Systems: Astandard Terminology and Recommended Performance Measures*. IWA.

Lesik, S. (2009). *Applied statistical inference with MINITAB* (pp. 290-291). CRC Press.

Lewis-Beck, M. S., Bryman, A., & Liao, T. F. (2004). R-squared. In *The SAGE encyclopedia of social science research methods* (Vol. 1, pp. 984-984). SAGE Publications, Inc., https://dx.doi.org/10.4135/9781412950589.n877

Liemberger, R., & Wyatt, A. (2018). Quantifying the global non-revenue water problem. *Water Supply*, *19*(3), 831-837. https://doi.org/10.2166/ws.2018.129

Lovell, D. (2020). Null hypothesis significance testing and effect sizes: can we 'effect' everything … or … anything?. *Current Opinion In Pharmacology*, *51*, 68-77. doi: 10.1016/j.coph.2019.12.001

Maggs, I. (2007). DEMAND MANAGEMENT: A SIMPLIFIED OPERATIONAL METHOD FOR TARGETING RESOURCES TO WATER LOSS. NAUGURAL WIOA NSW WATER INDUSTRY ENGINEERS & OPERATORS CONFERENCE. The Water Industry Operators Association of Australia (WIOA).

May, J.. (1991) Pressure Dependent Leakage. Water Supply, Vol. 15, No. 1, Mumbai, pp. 45-50, 1997. IWSA

Moksony, F. (1999). Small is beautiful. The use and interpretation of R2 in social research. *Szociológiai Szemle, Special issue*, 130-138.

Moslehi, I., & Jalili_Ghazizadeh, M. (2020). Pressure-Pipe Breaks Relationship in Water Distribution Networks: A Statistical Analysis. *Water Resources Management*, *34*(9), 2851–2868. https://doi.org/10.1007/s11269-020-02587-4

Perdices, M. (2017). Null Hypothesis Significance Testing, p-values, Effects Sizes and Confidence Intervals. *Brain Impairment, 19*(1), 70-80. doi:10.1017/BrImp.2017.28

Rezaei, H. (2017). *Impact of pressure fluctuations on pipe failures in water distribution networks* (Doctoral dissertation, Imperial College London). Imperial College London. https://doi.org/10.25560/73983

Robles-Velasco, A., Cortés, P., Muñuzuri, J., & Onieva, L. (2020). Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliability Engineering and System Safety*, *196*. https://doi.org/10.1016/j.ress.2019.106754

Sammut, C., & Webb, G. (2011). Mean Absolute Error. *Encyclopedia Of Machine Learning*, 652-652. doi: 10.1007/978-0-387-30164-8_525

Thornton, J., and Lambert, A. (2007). "Pressure management extends infrastructure life and reduces unnecessary energy costs." *Proc., Water Loss 2007*, IWA Publishing, London.

UNESCO. (2021). *UNITED NATIONS WORLD WATER DEVELOPMENT REPORT 2021: valuing water.* UNITED NATIONS EDUCATIONA.

Upton, G., & Cook, I. (2008). *A dictionary of statistics* (2nd ed.). Oxford: Oxford Univ. Press.

Van Nes, N., & Cramer, J. (2006). Product lifetime optimization: a challenging strategy towards more sustainable consumption patterns. *Journal Of Cleaner Production*, *14*(15-16), 1307-1318. doi: 10.1016/j.jclepro.2005.04.006

Vann og avløpsetaten. (2020). *Hovedplan for vannforsyning 2019–2028*. Retrieved from https://www.bergen.kommune.no/api/rest/filer/V51677

Vicente, D. J., Garrote, L., Sánchez, R., & Santillán, D. (2016). Pressure Management in Water Distribution Systems: Current Status, Proposals, and Future Trends. *Journal of Water Resources Planning and Management*, *142*(2), 04015061. https://doi.org/10.1061/(asce)wr.1943-5452.0000589

Wang, Y., Zayed, T., & Moselhi, O. (2009). Prediction Models for Annual Break Rates of Water Mains. *Journal of Performance of Constructed Facilities*, *23*(1), 47–54. https://doi.org/10.1061/(asce)0887-3828(2009)23:1(47)

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79-82. doi: 10.3354/cr030079

World Bank. 2017. *World Development Indicators*. Washington, D.C.: World Bank.

Appendices

Appendix 1: Scatter plot of predicted response (OEC model) versus predictor 'Length'

Appendix 2: Scatter plot of predicted response (OEC model) versus predictor 'Year'

Appendix 3: Scatter plot of predicted response (OEC model) versus predictor 'Piezometric head'

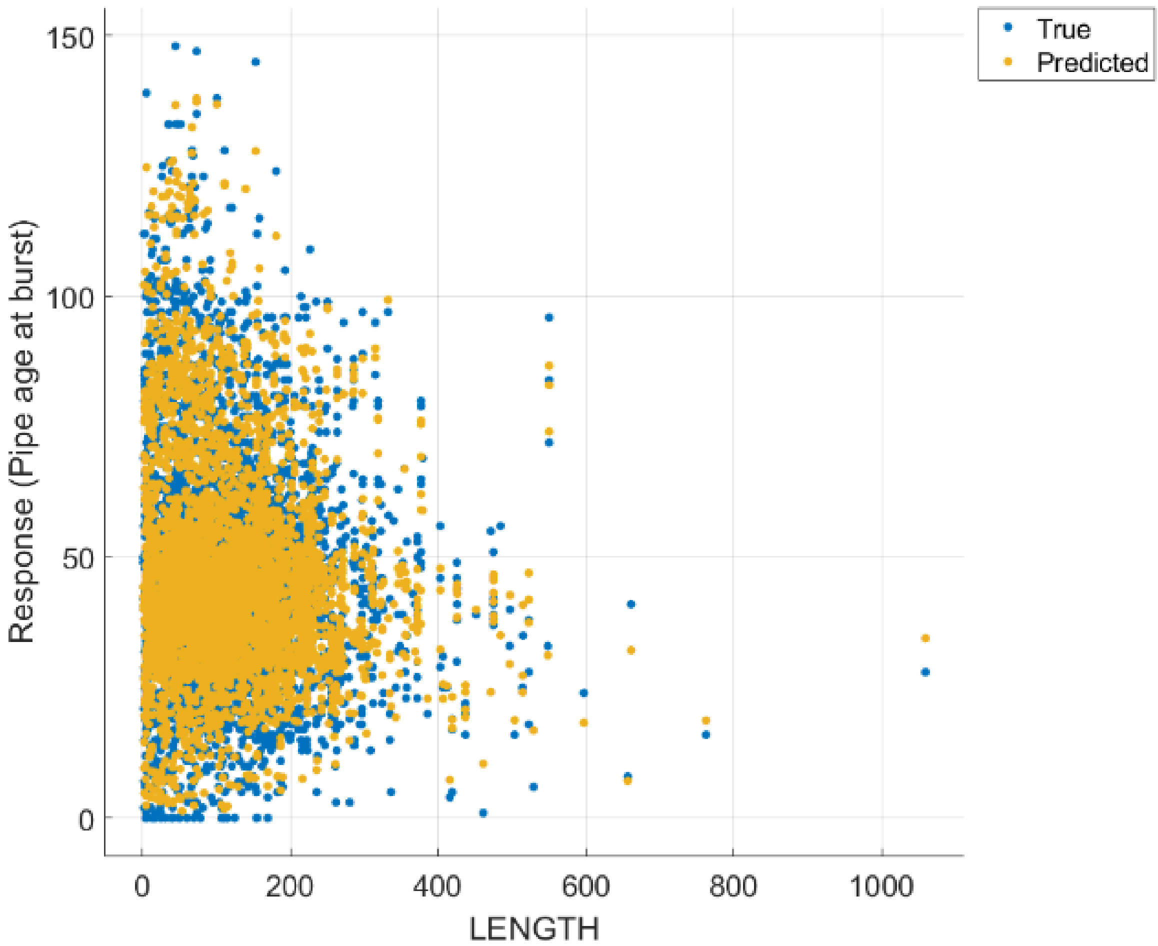
Appendix 4: Scatter plot of predicted response (OEC model) versus predictor 'Diameter'

Appendix 5: Box plot of predicted response (OEC model) versus predictor 'Material'

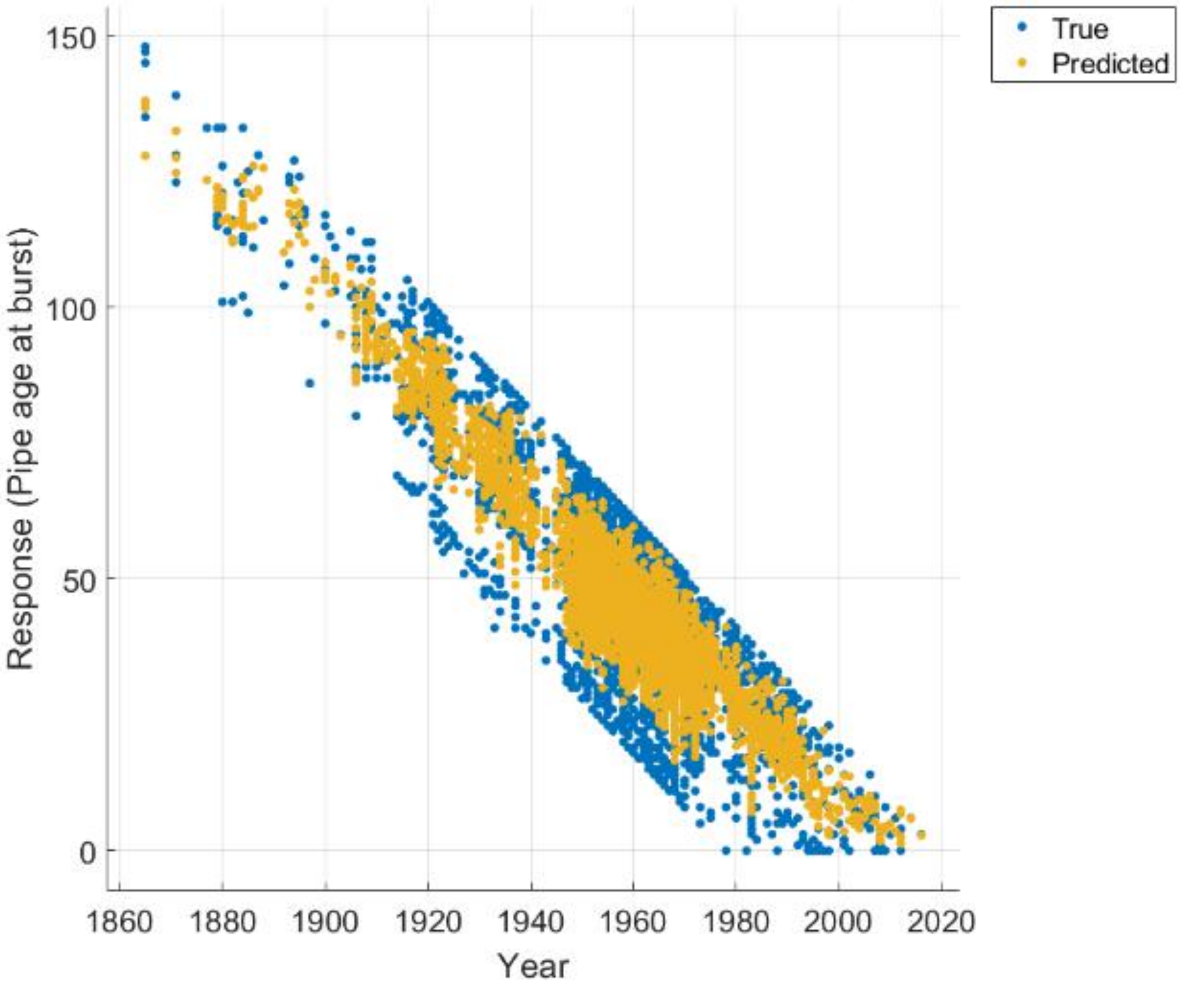
Appendix 6: Box plot of predicted response (OEC model) versus predictor 'Number of burst'

Appendix 7: Predicted response (OEC model) versus true values for the response 'pipe age at burst'

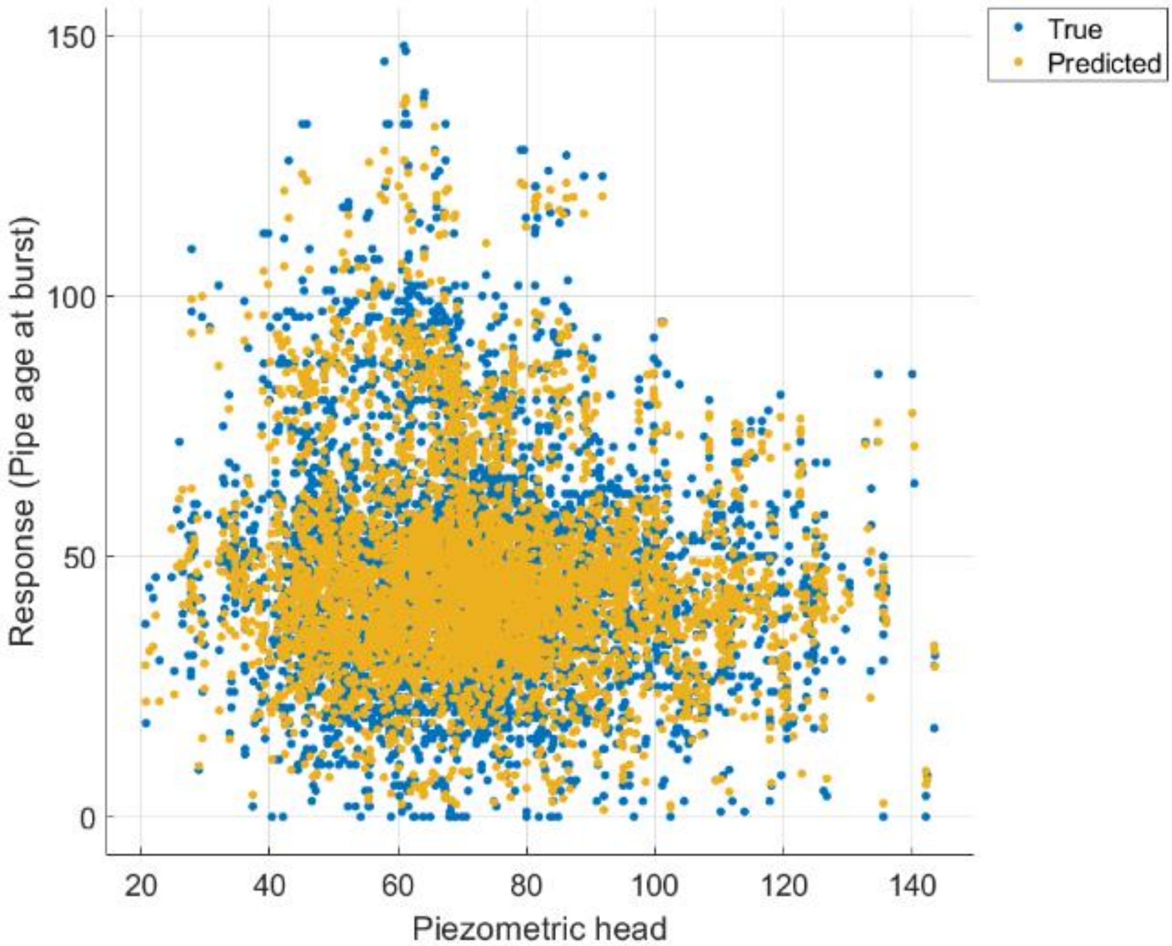
Appendix 1: Scatter plot of predicted response (OEC) versus predictor 'Length'



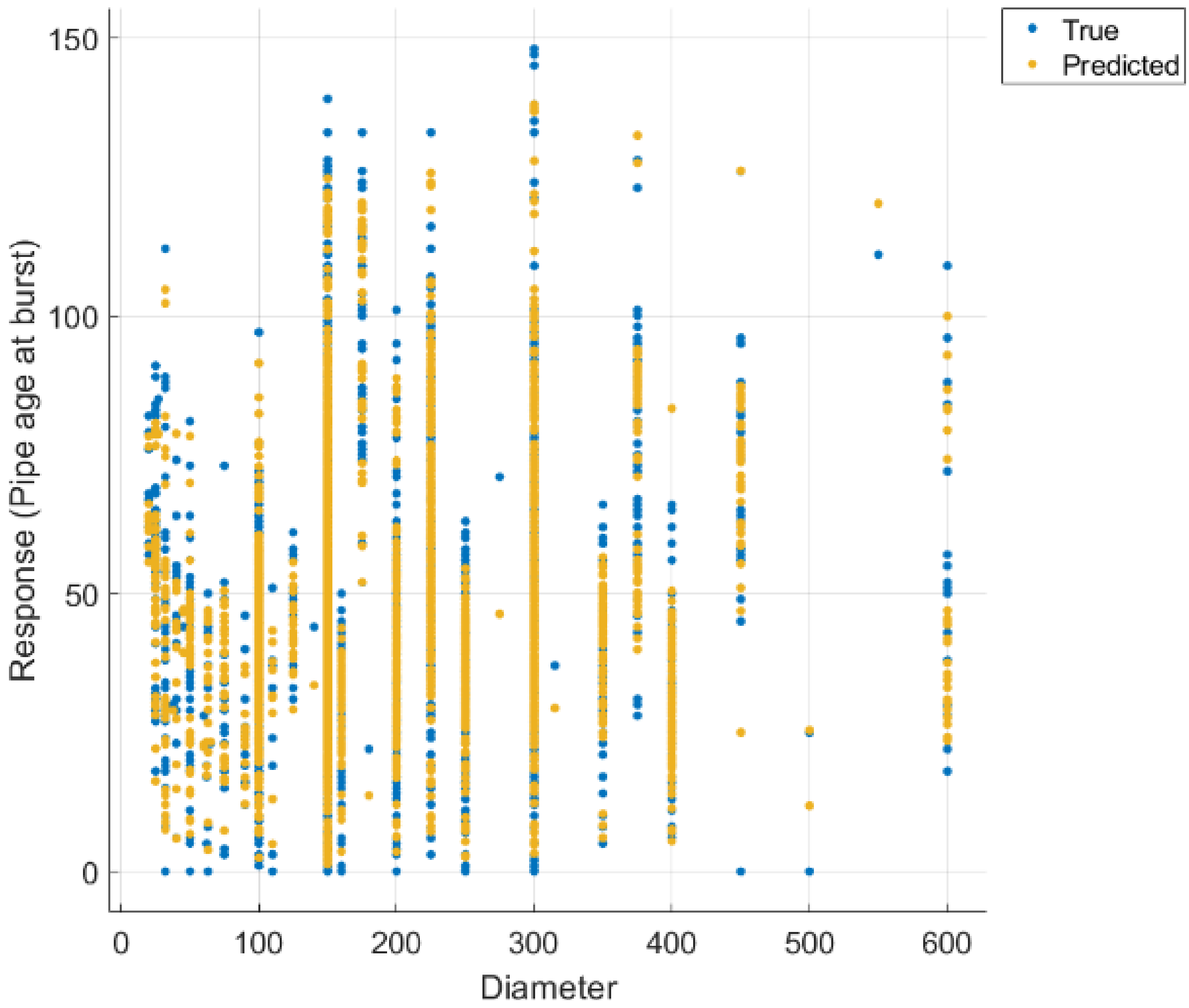
Appendix 2: Scatter plot of predicted response (OEC model) versus predictor 'Year'



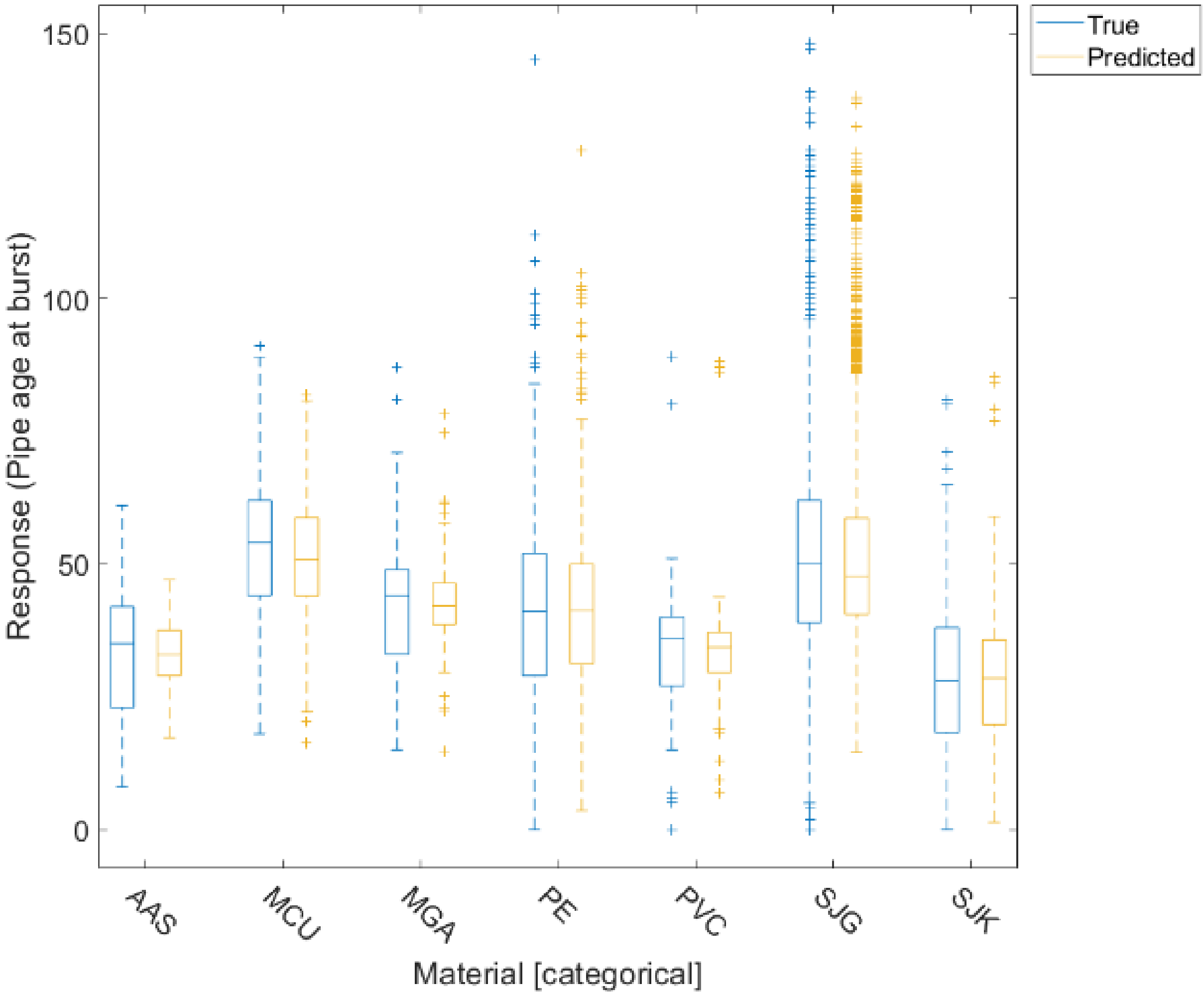
Appendix 3: Scatter plot of predicted response (OEC model) versus predictor 'Piezometric head'



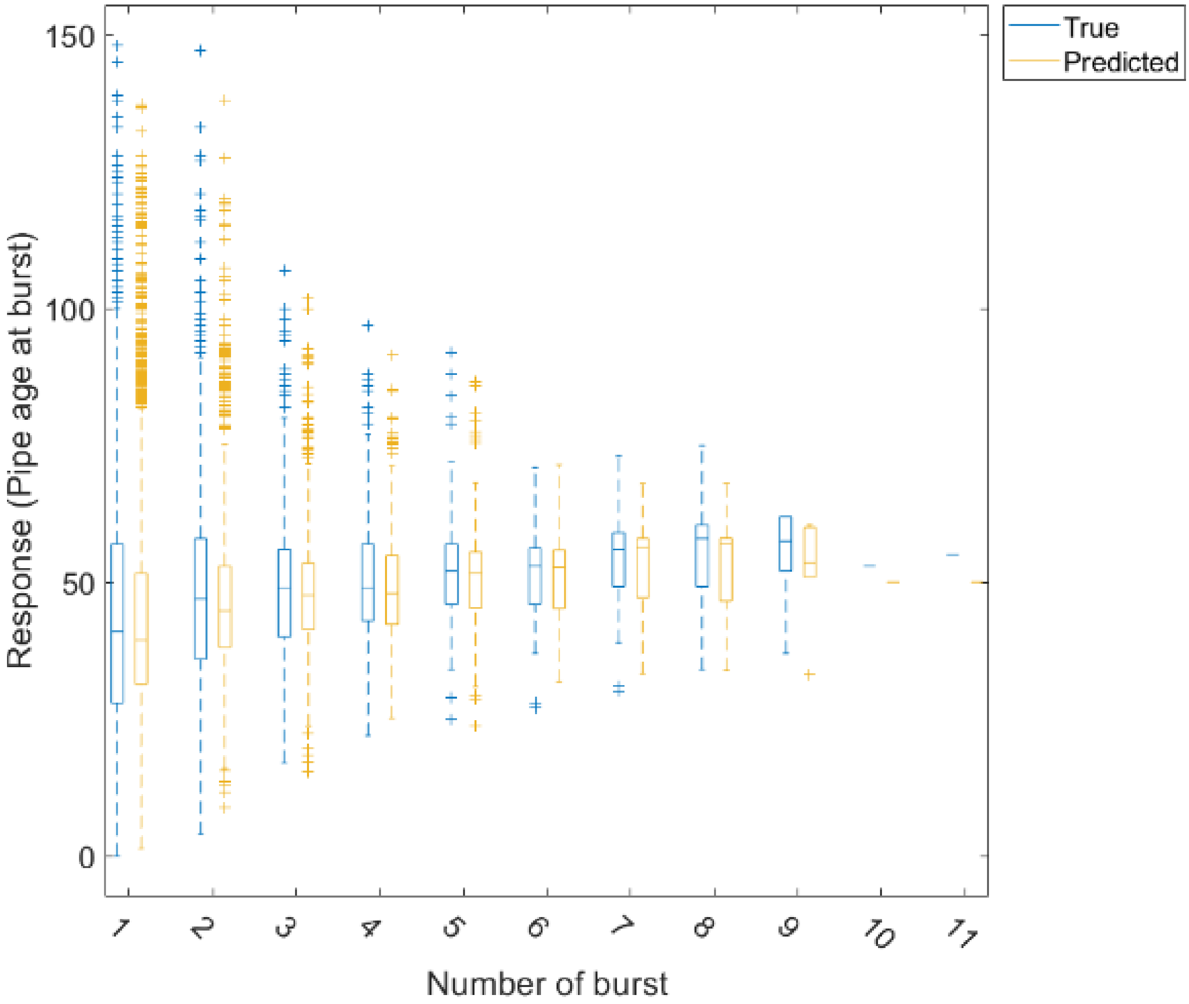
Appendix 4: Scatter plot of predicted response (OEC model) versus predictor 'Diameter'



Appendix 5: Box plot of predicted response (OEC model) versus predictor 'Material'



Appendix 6: Box plot of predicted response (OEC model) versus predictor 'Number of burst'



Appendix 7: Predicted response (OEC model) versus true values for the response 'pipe age at burst'